S

Scanning Electron Microscope

Robert Schmitt
Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen, Germany

Synonyms

Scanning electron microscopy

Definition

A scanning electron microscope (SEM) is an instrument for imaging topography and for obtaining material information of conductive specimen using a focused beam of high-energy electrons. The electron beam is deflected in a magnetic field and performs a scanning movement in a raster pattern to capture the specimens' surface. For imaging purposes interaction phenomena of the electron beam with the specimen like emission of secondary electrons (SE) or backscattered electrons (BSE) are detected and converted to grey values. A frequency analysis of X-rays reveals information about the present material. Nonconductive surfaces have to be covered with a conductive layer.

Theory and Application

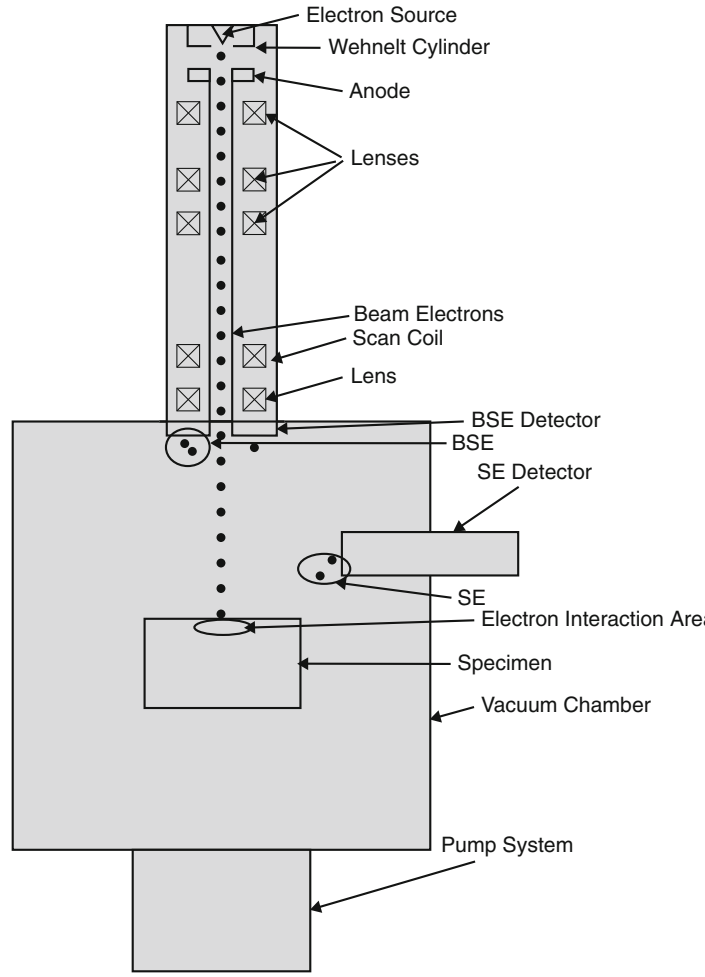
Historical Background

The invention of the SEM principle cannot be pinpointed to only one contributor in history. However, it was the German scientist Max Knoll who built the first "scanning microscope" in 1935 (Bogner et al. 2007; McMullan 1993). Manfred von Ardenne laid a further foundation for the SEM as well as for the transmission electron microscope (TEM). He described the theoretical principles in a 1938 published paper (von Ardenne 1938). The first true SEM with a resolution of 50 nm and magnification of 8000× was a contribution from the US-American scientists Zworykin, Hillier, and Sniyder in 1942 who still could not convince their fellow scientists of its usefulness (Zworykin et al. 1942). Charles Oatley, in his two-decade long process of researching the SEM, finally achieved the acceptance of the SEM as "one of the most powerful and productive methods of microscopy yet invented" (Oatley 1982; Smith 1997).

Structure and Basic Principle of the SEM

The SEM can be used to obtain information about the surface with high depth of focus or material composition of a specimen. The SEM uses a focused electron beam to scan the surface of a specimen in a raster pattern (Hafner 2007). The electron beam is focused by electromagnetic

The International Academy for Production Engineering et al. (eds.), *CIRP Encyclopedia of Production Engineering*,
https://doi.org/10.1007/978-3-662-53120-4



Scanning Electron Microscope, Fig. 1 The basic structure of an SEM

lenses on a small spot on the surface. Spot diameters of about 1 nm are possible (FEI Company 2016). The interaction products (e.g., secondary electrons, backscattered electrons, X-rays, cathodoluminescence) of the beam electrons and the specimen are detected.

Figure 1 shows the basic structure of an SEM. The SEM consists of an electron source, an accumulation of lenses (electromagnetic coils), detectors, and a vacuum system (chamber and pump system). The image is recorded/displayed on a computer system and/or a screen.

For surface imaging, the most common operation modes use signals of backscattered electrons and secondary electrons (Egerton 2005). While the electron beam scans the surface line by line, the detectors collect and count the secondary and backscattered electrons. The resulting signals for each point are converted to grey values. Scanning movement and displaying device are synchronized. As a result a grey value image of the topography is obtained, similar to a black and white photography or a black and white television image. The more electrons are detected, the brighter the image point will be. SEMs feature a magnification in the range of 100–1,000,000× (Hansen et al. 2006). The magnification is determined by

$$\text{Magnification} = \text{width on screen} / \text{scan length on specimen} \quad (1)$$

For material analysis a frequency analysis of emitted X-rays is evaluated. Since every element

has its characteristic X-ray spectrum, the identification of the characteristic signal gives information about the chemical composition. Using an adequate standard, a quantitative evaluation is possible. The X-rays can also be used to create a map of the specimen which represents the elemental distribution.

The specimen for SEM has to fulfill certain requirements. It needs to withhold vacuum and electron bombardment, it has to be electrically conductive, and furthermore, it needs to be clean and dry. A special version of an SEM is the environmental SEM (ESEM) which uses a different vacuum system and other types of detectors. The ESEM can be operated in higher, more natural pressures.

The Electron Source

The electron source consists of a cathode (electron source in Fig. 1), an anode, and a Wehnelt cylinder.

The cathode emits electrons that are accelerated towards the anode with a defined voltage, typically 50–30,000 V. There are basically two types of electrodes: thermionic cathodes (tungsten or LaB_6 (lanthanum hexaboride)) and field emission cathodes. The Wehnelt cylinder controls the current density and brightness of the electron beam. Brightness is defined as current per unit area normal to the given direction, per unit solid angle, and a criterion for beam quality.

The voltage defines the energy of the electron beam and determines the de Broglie wavelength λ:

$$\lambda = \frac{h}{p} = \frac{c \cdot h}{\sqrt{2m_0 \cdot eU}} \qquad (2)$$

where h is the Planck constant and p the relativistic impulse of the electron, e the electron charge, U the voltage, m_0 the electrons' rest mass, and c relativistic correction factor. The wavelengths of electrons are significantly smaller compared to wavelengths of visible light. Therefore, higher resolution is possible. The resolution of SEM is limited by the focal spot size of the electron beam and the imaging signals coming from a large interaction volume of the specimen. Common SEMs feature a minimal lateral resolution of about 2 nm (Hansen et al. 2006).

Lenses

Electromagnetic lenses focus and direct the electron beam in the SEM. The aim is to minimize the size of the focal spot at the specimens' surface. Condenser lenses bundle the electrons whereas the objective lens focuses the beam onto the specimen. The scan movement of the beam is controlled by a set of scan coils.

Detectors

SEM can be equipped with various detector systems like scintillator detectors (e.g., Everhart-Thornley detector) and solid state detectors. The standard setup contains a backscattered electron detector and a secondary electron detector. Since BSEs are deflected out of the specimen, the detector is mounted at the exit point of the electron beam. The collector system for SE and BSE are shown in Fig. 1.

Vacuum System

The vacuum system consists of the vacuum chamber and the pump system. The vacuum is created by a rough pump and a turbo molecular pump. The vacuum prevents interactions of the beam electrons with gas atoms and avoids sparkovers which could destroy the electron source and the detectors.

Specimen: Electron Beam Interactions

The electron beam interactions with the specimen produce various signals which are evaluated for the imaging process. These interactions can be distinguished between elastic and inelastic scattering of beam electrons in the specimen. In the scattering processes, the electrons interact with the coulomb potential of the atoms. In elastic scattering the total kinetic energy is conserved. In contrast, in inelastic scattering the total kinetic energy is not conserved. The most important interactions are explained below (Reimer 1998; Goldstein 2003).

Secondary Electrons (SE)

The most common imaging mode uses low-energy secondary electrons as a result of an inelastic scattering process. Due to their low energy, the SE originates from a small interaction volume only few nanometers below the surface. Therefore, the SE signal delivers the best resolution. The main mechanisms that generate secondary electrons are:

1. Interactions of electrons from the incident beam with specimen atoms
2. Interactions of high-energy backscattered electrons with specimen atoms
3. Produced by high-energy BSEs which strike the pole pieces and other solid objects within the specimen chamber

The SE is attracted to the detector using a positively biased collector grid placed on one side of the specimen. The topographic contrast depends mainly on the incident angle and the border effect (Fig. 2).

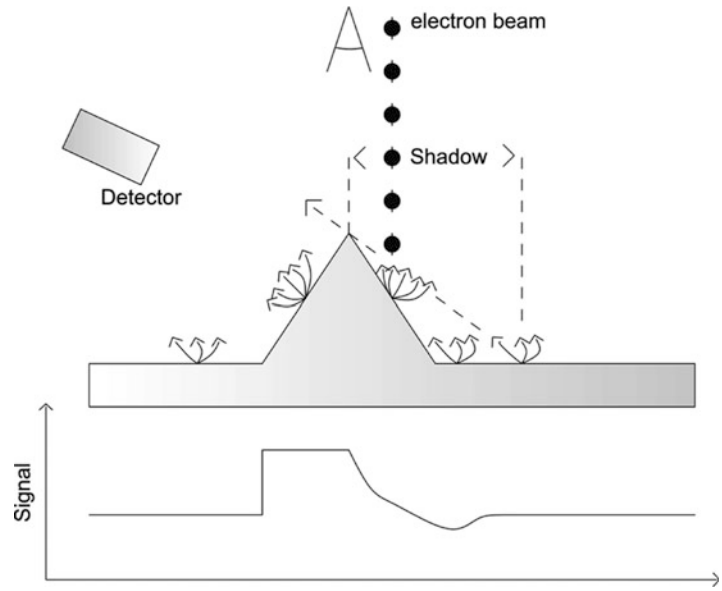
Backscattered Electrons (BSE)

Backscattered (high-energy) electrons are elastically scattered beam electrons. The electrons are deflected back out of the specimen interaction volume.

The probability of backscattering depends on the atomic number. In areas with a higher mean atomic number, the surface looks brighter in the image. This is called material contrast.

X-ray Emission

X-rays are generated by high energetic electrons impinging on a solid specimen. In general, the characteristic radiation is analyzed to determine the chemical composition of the specimen. Beam electrons can knock out bound electrons from inner shells when their kinetic energy is higher than the binding energy. The vacancy is subsequently filled by an electron of a higher shell. The energy difference is emitted as characteristic X-ray radiation. X-rays, though less common than, e.g., BSE and SE, can also be used for imaging.



Scanning Electron Microscope, Fig. 2 SEM imaging of a specimen: illustration of the border effect and the incident angle influencing the imaging. The *arrows* represent SEs. At edges more SEs are emitted. The more electrons are emitted, the brighter the image appears. Since the detector is located on one side of the specimen, edges can produce a shadow

Other phenomena that can be exploited are, e.g., emission of photons (light) and absorption of electrons (Amelinckx et al. 1997). If the specimen is thin enough and therefore efficiently penetrable by electrons (transparent), it provides an excellent sample for the transmission electron microscope (TEM).

Key Applications

- Image morphology of samples (e.g., view bulk material, coatings, sectioned material, foils, even grids prepared for transmission electron microscopy).
- Image compositional and some bonding differences (through contrast and using backscattered electrons).
- Image molecular probes: metals and fluorescent probes.
- Undertake micro- and nano-lithography: remove material from samples; cut pieces out or remove progressive slices from samples (e.g., using a focused ion beam).
- Heat or cool samples while viewing them (while possible in many instruments, it is generally done only in ESEM or during cryo-scanning electron microscopy).
- View frozen material (in an SEM with a cryostage).
- Generate X-rays from samples for microanalysis (EDS; WDS).
- Study optoelectronic behavior of semiconductors using cathodoluminescence.
- View/map grain orientation/crystallographic orientation and study related information like heterogeneity and microstrain in flat samples (electron backscattered diffraction).
- Electron diffraction using electron backscattered diffraction. The geometry may be different to a transmission electron microscope but the physics of Bragg diffraction is the same.

(Key applications: http://www.ammrf.org.au/myscope/sem/background/practical/. Accessed 20 Jan 2016).

Cross-References

▶ Metrology

References

Amelinckx S, van Dyck D, van Landuyt J, van Tendeloo G (1997) Electron microscopy: principles and fundamentals. VHC, Weinheim

Bogner A, Jouneau PH, Thollet G, Basset D, Gauthier C (2007) A history of scanning electron microscopy developments: towards "wet-STEM" imaging. Micron 38:390–401

Egerton RF (2005) Physical principles of electron microscopy: an introduction to TEM, SEM, and AEM. Springer, New York

FEI Company (2016) An introduction to electron microscopy. Retrieved from http://www.fei.com/introduction-to-electron-microscopy/. Accessed 20 Jan 2016

Goldstein J (2003) Scanning electron microscopy and x-ray microanalysis. Kluwer Academic/Plenum, New York

Hafner B (2007) Scanning electron microscopy primer. Retrieved from http://www.charfac.umn.edu/instruments/sem_primer.pdf. Accessed 20 Jan 2016

Hansen HN, Carneiro K, Haitjema H, De Chiffre L (2006) Dimensional micro and nano metrology. CIRP Ann Manuf Technol 55(2):721–743

McMullan D (1993) Scanning electron microscopy 1928–1965. Retrieved from http://www-g.eng.cam.ac.uk/125/achievements/mcmullan/mcm.htm. Accessed 20 Jan 2016

Oatley CW (1982) The early history of the scanning electron microscope. J Appl Phys 53(2):R1–R13

Reimer L (1998) Scanning electron microscopy: physics of image formation and microanalysis. Springer, Berlin/Heidelberg

Smith KCA (1997) Charles Oatley: a pioneer of the SCM. Retrieved from http://www2.eng.cam.ac.uk/~bcb/cwo1.htm. Accessed 20 Jan 2016

von Ardenne M (1938) Das Elektronen-Rastermikroskop [the scanning electron microscope]. Z Phys 109(9–10):553–572

Zworykin VA, Hillier J, Snyder RL (1942) A scanning electron microscope. ASTM Bull 117:15–23

Scanning Electron Microscopy

▶ Scanning Electron Microscope

Scanning Force Microscopy

▶ Atomic Force Microscopy

Scanning Tunneling Microscope

Fengzhou Fang[1,2] and Bingfeng Ju[3]
[1]The State Key Laboratory of Precision Measuring Technology and Instruments, Centre of MicroNano Manufacturing Technology, Tianjin University, Tianjin, China
[2]Centre of MicroNano Manufacturing Technology (MNMT-Dublin), University College Dublin, Dublin, Ireland
[3]Department of Mechanical Engineering, Zhejiang University, Hangzhou, China

Synonyms

Scanning tunneling microscopy

Definition

Scanning tunneling microscope (STM) is an instrument for imaging conductive solid surfaces with an atomic resolution based on the concept of quantum tunneling.

Theory and Application

Introduction

STM was originally developed in 1981 by Gerd Binnig and Heinrich Rohrer (Binnig and Rohrer 1982), who were awarded the 1986 Nobel Prize in Physics for this great invention. Over the years, the STM has been proved to be an extremely versatile and powerful technique for many disciplines in material science (Yao and Wang 2004), precision engineering (Weckenmann and Hoffmann 2007; Hansen et al. 2006), physics, biology, and so on (Gao 2010).

The STM can be used not only in ultrahigh vacuum but also in ambient of air, water, liquid, or gas and at temperatures ranging from near-zero Kelvin to a few hundred degrees Celsius. Apart from surface topograph imaging, since the quantum tunneling also depends on the chemical nature of sample and tip, the STM also serves for characterization of electronic properties of solid samples, atomic manipulation, and nanostructure fabrication.

Physics Principle of Tunneling

Tunneling phenomena have been studied for long time and can be well understood in terms of quantum theory. Considering an one-dimensional vacuum barrier between two electrodes (the sample and the tip) and assuming their work functions to be the same and thus the barrier height to be F, if a bias voltage of V is applied between the two electrodes with a barrier width d, according to quantum theory under first-order perturbation (Hansen et al. 2006), the tunneling current is

$$I = \frac{2\pi e}{\eta}\sum_{\mu,v} f\left(E_{\mu}\right)\left[1 - f(Ev + eV)\right] \times |M_{\mu v}|^2\,\delta\left(E_{\mu} - Ev\right)$$

where $f(E)$ is the Fermi function, M μ v is the tunneling matrix element between states ψ μ and ψ v of the respective electrodes, and E μ and E v are the energies of ψ μ and ψ v. The tunneling matrix element $M_{\mu}v$ can be expressed as

$$M_{\mu v} = \frac{\eta^2}{2m}\int \mathrm{dS}\cdot\left(\psi_{\mu}{}^{*}\nabla\psi_{v} - \psi_{v}{}^{*}\nabla\psi_{\mu}\right)$$

where the integral is over all the surfaces surrounding the barrier region. To estimate the magnitude of $M\mu$ v, the wave function of the sample ψ v can be expanded in the generalized plane-wave forma:

$$\psi_{v} = \Omega_{s}{}^{-1/2}\sum_{G} a_{G}\exp\left[\left(k^2 + |\mathrm{k}_{G}|^2\right)^{1/2} z\right] \times \exp(i\mathrm{k}_{\mathrm{G}}\cdot x)$$

where Ω_{s} is the volume of the sample, $k = \mathrm{h}^{-}1 (2\mathrm{m}\text{ø})^{1/2}$ is the decay rate, ø is the work function, $kG = \mathrm{k}_{//} + G,k_{//}$ is the surface component of Bloch vector, and G is the surface reciprocal vector.

The above formulas mean that the tunneling current depends on the tunneling gap distance d very sensitively. In the typical case, the tunneling current would change one order, while the gap

distance changes only 1 Å. This accounts for extremely high vertical resolution of 0.1 Å of STM.

Operation Principle of STM

The basic operation principle of STM is illustrated in Fig. 1. A sharp conductive probe tip is fixed on the top of a piezodrive (*Pz*) to control the height of the tip above a surface. When the tip is brought close enough to the sample surface, electron can tunnel through the vacuum barrier between tip and sample surface. Applying a bias voltage on the sample, a tunneling current can be measured through the tip, which is extremely sensitive to the distance between the tip and the surface. Another two piezodrives (*Px* and *Py*) are used to scan the tip in two lateral dimensions. There are two basic operation modes to acquire surface images (Whitehouse 1994). Figure 2 shows the schematic of the two operation modes in STM.

Constant Current Mode

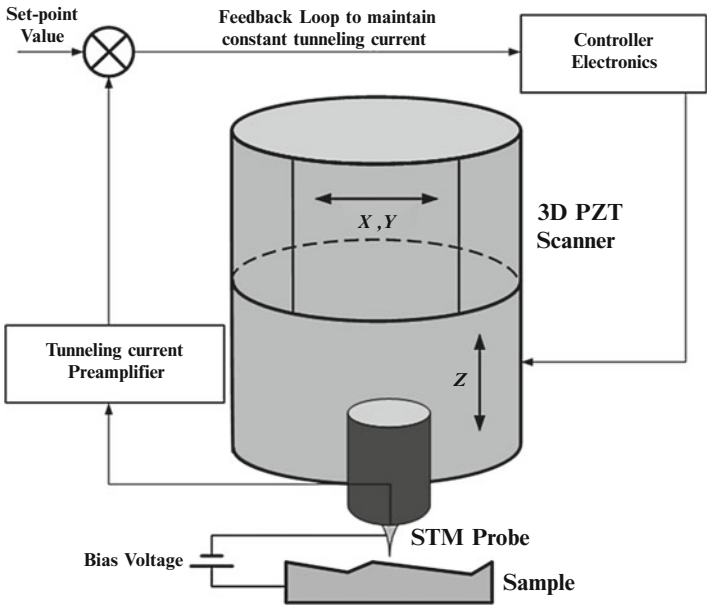
A feedback controller is employed to adjust the height of the tip to keep the tunneling current constant. During the tip scanning on the surface, the height of the tip (the voltage supplied to *Pz* piezodrive) is recorded as an STM image, which represent the topograph of the surface. This operation mode of STM is called constant current mode. Constant current mode is mostly used in STM topograph imaging. It is safe to use the mode on rough surfaces since the distance between the tip and sample is adjusted by the feedback circuit.

Constant Height Mode

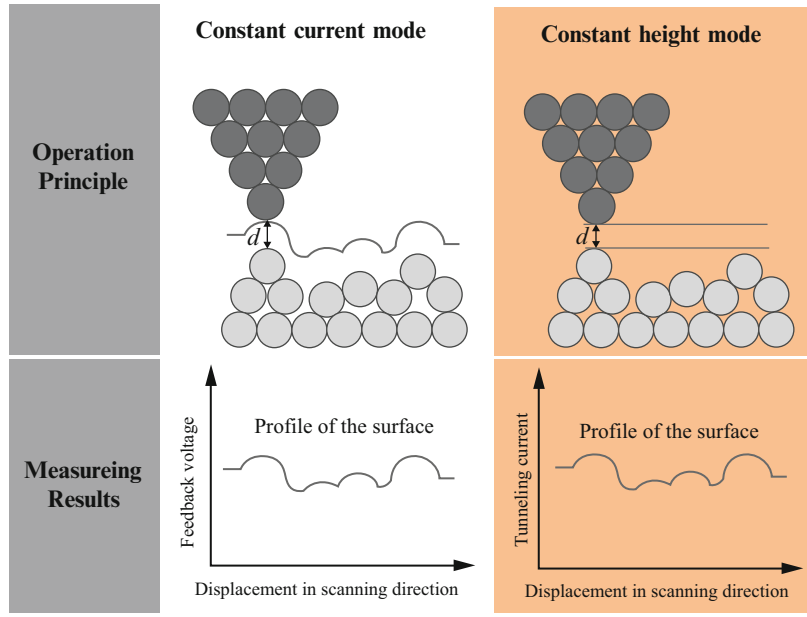
For a smooth surface, it is also possible to keep the tip height constant above the surface. The variation of the tunneling current reflects the small atomic corrugation of the surface. The constant height mode has no fundamental difference to the constant current mode. However, the tip could be crashed if the surface corrugation is big. On the other hand, the STM can scan very fast in this mode for research of surface dynamic processes.

Key Applications

- Characterization of nanostructures with atomic resolution
- The reconstruction illustration of materials deposited on the surface of gold
- Morphological characterization of carbon nanotubes
- Analysis of nanoparticles
- Determination of the structure in thin layers of organic semiconductors
- Characterization of ceramic layers obtain from the method of CVD/PVD



Scanning Tunneling Microscope, Fig. 1 Basic principal of STM system



Scanning Tunneling Microscope, Fig. 2 Basic operation mode in STM: constant current mode and constant height mode (Reproduced from "Scanning tunneling microscopy," PK Hansma and J Tersoff,. J Appl Phys 61(2) 1987:R1–R23, with the permission of AIP Publishing)

- Characterization of nanostructure conductors and semiconductors by using STM
- Study of surface atomic dynamic processes
- Study of electronic properties of surfaces
- STM-based atomic manipulation
- STM-based fabrication

Cross-References

► Metrology

References

Binnig G, Rohrer H (1982) Scanning tunneling microscopy. Helv Phys Acta 55:726–735
Gao W (2010) Precision nanometrology: sensors and measuring systems for nanomanufacturing. Springer, London
Hansen HN, Carneiro K, Haitjema H, De Chiffre L (2006) Dimensional micro and nano metrology. CIRP Ann Manuf Technol 55(2):721–743
Hansma PK, Tersoff J (1987) Scanning tunneling microscopy. J Appl Phys 61(2):R1–R23
Weckenmann A, Hoffmann J (2007) Long range 3D scanning tunneling microscopy. CIRP Ann Manuf Technol 56(1):525–528
Whitehouse DJ (1994) Handbook of surface metrology. Institute of Physics Publishing, Bristol
Yao N, Wang ZL (eds) (2004) Handbook of microscopy for nanotechnology. Tsinghua University Press/Dordrecht, Beijing

Scanning Tunneling Microscopy

► Scanning Tunneling Microscope

Scatterometry

Matteo Calaon[1], Morten Hannibal Madsen[2] and Richard Leach[3]
[1]Department of Mechanical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
[2]Topsil GlobalWafers A/S, Frederikssund, Denmark
[3]Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, Nottingham, UK

Synonyms

Reflectometry, Ellipsometry

Definitions

Scatterometry is a technique for measuring periodic structures on a surface with dimensions from a few nanometers to tens of micrometers. Scatterometry is a fast, nondestructive technique and is, therefore, suitable for in-line metrology.

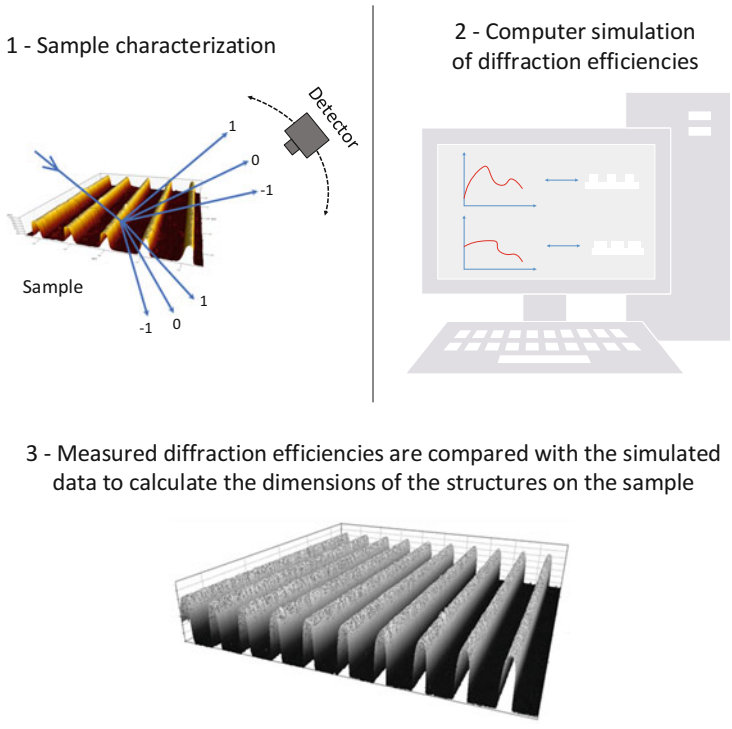
Theory and Applications

The basic working principle of scatterometry is to use the information in the interference of light interacting with periodic structures, for example, a diffraction grating, to characterize a surface. The intensities of the resulting diffraction orders are used as a unique fingerprint for a given surface. Scatterometry is often applied where imaging techniques cannot be used due to a lack of resolution and can be considered a super-resolution technique.

The Principal Workflow

The principal workflow (Fig. 1) when performing scatterometry measurements uses an inverse modeling approach, with three steps: (1) the experimental measurement in which the light from an optical source is incident on the sample to be measured, reflects from the surface, and the resultant diffraction pattern is detected, (2) modeling of the diffraction pattern that would result if reflecting the light from an ideal grating, and (3) comparison of the experimental and simulated diffraction patterns. Note that some scatterometers operate in transmission through the sample, but only reflection mode systems will be discussed in the rest of the document.

The experimental quantities to be measured in scatterometry are the intensities of the diffracted light as a function of the angle of incidence, the reflection angles of the diffracted orders, and/or the wavelength of the diffracted light. Quantities of interest are represented by the diffraction efficiencies calculated as the ratio of the incident intensity and scattered intensity for a given order.



Scatterometry,
Fig. 1 Principal workflow while performing scatterometry measurements. (1) Experimental measurements, (2) modeling diffraction efficiency, (3) comparison of the data from the first two steps

Simulation of the diffraction efficiencies is typically performed using software for solving Maxwell's equations, such as rigorous coupled-wave analysis (RCWA) (Moharam and Gaylord 1981) or finite element methods (FEMs) (Humphries 1997). A degree of a priori information about the sample being measured is needed in order to simulate the diffraction efficiencies. This includes the material properties of the periodic structures and substrate and an estimate of the dimensions of the structure. Such a priori information is almost always available, for example, when scatterometry is used for quality control of fabricated structures.

The last step is to compare the measured diffraction efficiencies to the simulated ones. Performing many simulations of structures with small dimensional variations can produce a database of diffraction efficiencies. The database approach greatly reduces the time for the comparison. Instead of performing the lengthy calculations described in step (2), a simple database lookup process is used. On a normal desktop computer, this can be accomplished in a few milliseconds.

Applications

Continuous advancement in the semiconductor industry and development of large volume production technologies (i.e., injection molding, microinjection molding, roll-to-roll manufacturing) have introduced new fabrication methods to produce micro- and nanoscale periodic structures enabling different functionalities. To characterize nanostructured surfaces, scanning electron microscopy (SEM) or atomic force microscopy (AFM) is generally used to ensure the required lateral and vertical resolutions. However, AFM and SEM are far too slow to use for in-line characterization, unless highly costly parallel realizations are employed. As such, due to its potential for high-speed measurement, scatterometry is a competitive technique for in-line applications (Calaon et al. 2015). With scatterometry, it is possible to characterize relatively large areas (up to square centimeters) with nanometer precision for the geometry of periodic structures.

The ability to only measure periodic structures may seem limiting. However, many structures, especially in the semiconductor industry, are periodic by design. Furthermore, periodic structures for the purpose of characterization can often be placed near the real structures of interest, for example, in the dice lines on a wafer. Instead of measuring on the real structure, measurements can be performed on the test structure and correlated to the real design. This method is also known from the printing industry, for example, test fields at the edge of a newspaper.

Semiconductor Industry

Scatterometry is used for in-line dimensional measurements of critical geometries during the fabrication of silicon components, for example, measuring the etch depth before continuing to the next process step. The increasing use of multiple patterning steps extended the initial adoption of scatterometry to monitor etch processes on integrated circuit masks to complex features, where as many as ten parameters are included in the model (e.g., line-width, height, sidewall angles, roughness, and pitch measurements). In some applications, the use of multi-patterned features (e.g., antireflective coatings) can prevent ultraviolet light from penetrating into deeper layers. Recent developments have focused on applying scatterometry to overlay metrology, but challenges in multilayer measurements need to be resolved (Hsu et al. 2015).

The semiconductor industry is currently adopting scatterometry for line shape metrology, providing statistically valid average values for large numbers of increasingly smaller features. Recent technological advances have demonstrated scatterometry to be a viable solution for fast nondestructive in-line process control and monitoring for extreme ultraviolet lithography in electronic circuit fabrication (Li et al. 2013).

Instruments are commercially available for the semiconductor industry, but they are highly expensive and only target large production facilities.

Polymer Industry

With the development in nano-texturing of surfaces with materials other than semiconductors, scatterometry has found applications in fields from microfluidic channels (Calaon et al. 2015)

to roll-to-roll applications (Petrik et al. 2014). Scatterometry has recently been used for in-line optical critical dimension quantification of injection molded nanostructures on transparent polymer surfaces (Madsen et al. 2017). Quantification of replication fidelity between master geometries and polymer structures can be used as direct feedback for continuous quality control of injection molded production of nanostructured products. Scatterometers can characterize the critical geometries of different nanostructures with an accuracy of a few nanometers in less than a second for single acquisitions.

Scatterometers

Scatterometers can be divided into two main categories: (1) those that scan the incoming and/or outgoing angles, referred to as "angular scatterometers," and (2) those that scan a range of wavelengths, referred to as "spectroscopic scatterometers."

The simplest scatterometer comprises a static setup with the incoming beam and detector kept at fixed angles, limiting the intensity measurement acquisition to a single diffraction order (Kleinknecht and Meier 1978). Due to its simplicity, this setup is suitable for in-process characterization.

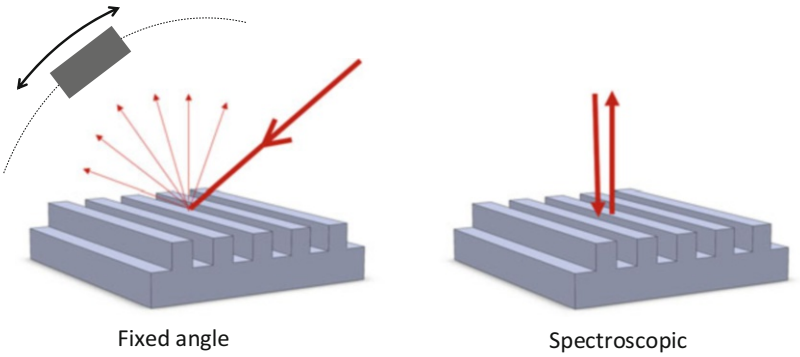
In angular scatterometers, a fixed incidence angle and a moving detector in the same plane as the incoming beam are used to measure the angle-dependent scattering intensities (Fig. 2 left). Such a setup can also measure the pitch of a periodic lattice. However, for pitch measurement, a Littrow configuration (Chernoff et al. 2008) is superior. This configuration enables overlapping of the diffracted light with the incident light. Concurrent sample rotation allows precise measurement of the angles between different diffraction orders from which the pitch is calculated.

Spectroscopic scatterometry scans the wavelength of the incoming light, filtered at the detector side with, for example, a spectrometer. Specular reflection is studied, instead of the first (and higher) orders of the diffracted light. In the simple configuration (Fig. 2 right), both the light source and the detector are positioned normal to the sample. The setup of the normal incidence configuration, without any moving parts, provides simple alignment and good vibration isolation properties. The possibility of imaging specific areas of interest has been demonstrated by building a scatterometer into a conventional optical microscope (Madsen et al. 2015).

In imaging scatterometry, a set of multispectral images acquired at different wavelengths is used to determine the diffraction efficiencies of different structures at single pixels in the images. Reference images (e.g., a dark image and an image of a blank sample) are acquired for each wavelength and used to characterize the diffraction efficiency of a single pixel. The diffraction efficiency from a single pixel is then used for reconstruction of the grating shape parameters in that local area.

Modeling

For modeling of grating structures, RCWA is the most widespread technique to calculate diffraction efficiencies. The introduction of RCWA enables scatterometers to determine the dimensions of gratings. RCWA decomposes periodic structures into smaller boxes characterized by representative parameters that describe the grating, such as pitch,



Scatterometry, Fig. 2 Schema of different scatterometry setups. (Left) A fixed angle setup where only the detector moves. (Right) Spectroscopic system with no moving parts

height, width, sidewall angles, and, for more advanced models, the roundness of the top and bottom corners of the grating. The computer model solves the electromagnetic field of the different boxes composing each single object of the periodic structure. Consequently, the boundaries of the different rectangular boxes are matched to find the electromagnetic field outside the grating structure.

FEM can also be applied for simulation of diffraction efficiencies. With FEM, more complex structures can be handled with the support of a CAD model in the design phase and a meshing procedure for the determination of Maxwell's equations (Jin 2002). However, the drawback of this method is the need for much longer computational times, which can only be reduced by the use of supercomputers.

Comparison of Experimental and Numerical Data

Diffraction efficiencies are compared with the simulated data and the best match yields the dimensions of the structures on the sample (Fig. 1). Direct optimization or library search approaches are the main approaches for the comparison.

Direct optimization is generally used for simple geometries, and all the free parameters (e.g., height and width) are optimized using, for example, a differential evolution algorithm (Storn and Price 1997). The optimization identifies the best fit using either scalar diffraction theory, RCWA, or FEM between selected free parameters of the structure under analysis and reference measurements.

For the library search approach, all values are computed for different elements in a pre-generated database. The size of the generated database containing scattering intensity profiles for different structures is limited by the time available to generate it. The subsequent comparison is based on a very fast algorithm, such as chi-squared minimization, where χ^2 can be defined as

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\eta_i - f(\boldsymbol{\Omega}_i,\boldsymbol{\alpha})}{\delta\eta_i}\right)^2, \qquad (1)$$

where η_i is the ith measured diffraction efficiency, N is the number of measured diffraction efficiencies, $\delta\eta$ are the experimental uncertainties, and $f(\Omega_i, \alpha)$ are the modeled diffraction intensities for a given set of parameters.

Typically, the minimization only takes a few milliseconds, and direct feedback using the library search approach is a valuable tool when used on a production line. The approach will always provide a best fit within the produced library database. The validity of the generated "true value" can be subject to errors coming from values outside the pre-generated database and faulty data calculated using incorrect optical material properties.

Uncertainty Analysis

The uncertainty of scatterometry measurements has influence factors resulting from both the modeling and the experimental steps. For scatterometry instruments, the most important influence factors are (Madsen and Hansen 2016):

- Type A: Detector noise, light fluctuations
- Type B: Long-term stability of the light source, polarization of the light, incident angle variations, wavelength variations of the detected light, orientation of the sample, light losses in the system, incomplete collection of the scattered light

Sources of random error can be reduced by increasing the integration time or averaging the dataset, and their quantification and computation are well addressed by least-squares approaches. Type B influence factors require instrument calibration to identify and quantify their magnitude.

The uncertainties resulting from the model are more challenging to handle, and a full uncertainty analysis for the combined uncertainties has yet to be presented in the literature. For the advanced scatterometry technique known as "Mueller polarimetry," where changes in light polarization are measured, traceable measurements have recently been demonstrated (Hansen et al. 2017). Here, a general least-squares analysis approach was used to estimate the combined instrument and model uncertainties.

Cross-References

► Metrology

References

Calaon M, Madsen MH, Weirich J, Hansen HN, Tosello G, Hansen PE, Garnaes J, Tang PT (2015) Replication fidelity assessment of large area sub-μm structured polymer surfaces using scatterometry. Surf Topogr Metrol Prop 3:045005

Chernoff DA, Buhr E, Burkhead DL, Diener A (2008) Picometer-scale accuracy in pitch metrology by optical diffraction and atomic force microscopy. Proc SPIE 6922:69223J–692211

Hansen PE, Madsen MH, Lehtolahti J, Nielsen L (2017) Traceable Muller polarimetry and scatterometry for shape reconstruction of grating structures. Appl Surf Sci 421 B:471–479

Hsu SC, Yuan CP, Chena C, Yu CC, Hsing H, Wu R, Kuo KTL, Amir N (2015) Scatterometry or imaging overlay a comparative study. Proc SPIE 9424:942409

Humphries S (1997) Field solutions on computers. CRC Press, New York

Jin J (2002) The finite element method in electromagnetics. IEEE Press, Wiley, New Jersey

Kleinknecht HP, Meier H (1978) Optical monitoring of the etching of SiO2 and Si3N4 on Si by the use of grating test patterns. J Electrochem Soc 125:798–803

Li J, Kritsun O, Dasari P, Volkman C, Wallow T, Hu J (2013) Evaluating scatterometry 3D capabilities for EUV. Proc SPIE 8681:86810S

Madsen MH, Hansen PE (2016) Scatterometry- fast and robust measurements of nano-textured surfaces. Surf Topogr Metrol Prop 4:023003

Madsen MH, Hansen PE, Zalkovskij M, Karamehmedović M, Garnæs J (2015) Fast characterization of moving samples with nano-textured surfaces. Optica 2:301–306

Madsen JS, Thamdrup LH, Czolkos I, Hansen PE, Johansson A, Garnaes J, Nygard J, Madsen MH (2017) In-line characterization of nanostructured mass-produced polymer components using scatterometry. J Micromech Microeng 27:085004

Moharam MG, Gaylord TK (1981) Rigorous coupled-wave analysis of planar-grating diffraction. J Opt Soc Am 71:811–818

Petrik P, Kumar N, Juhasz G, Major C, Fodor B, Agocs E, Lohner T, Pereira SF, Urbach HP, Fried M (2014) Optical characterization of macro-, micro and nano-structures using polarized light. J Phys Conf Ser 558:012008

Storn R, Price K (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 11:341–359

Scheduling

Giuseppe Stecca
IASI - CNR, Institute for Systems Analysis and Computer Science, National Research Council, Rome, Italy

Synonyms

Executive planning; Machine scheduling; Programming; Sequencing

Definition

Scheduling deals with the allocation of resources to tasks over given time periods, and its goal is to optimize one or more objectives (Pinedo 2008).

Theory and Application

Role and History of Scheduling

The role of scheduling is the assignment of resources to working jobs over time. Scheduling is an operational decision-making process affecting company and organization performance and its ability to add value and to respect contracts. The application of scheduling is wide, starting from manufacturing and production systems to information processing environments as well as transportation and distribution systems. Typical scheduling problems are the sequencing of batches in continuous and discrete manufacturing environments with the aim of minimization of the total time spent on setups and/or the maximization of throughput while meeting the due dates, assigning gate in airports, scheduling of tasks in computing processing units, scheduling of project activities in a team, healthcare, and timetabling.

Scheduling theory and application was pioneered by Henry Gantt at the beginning of the twentieth century; he developed the famous Gantt charts during the First World War. Gantt's findings were further developed by the researches of Johnson (1954), Jackson (1956), and Smith (1956).

Other important references are in Conway et al. (1967), Lawer (1973), Błażewicz et al. (1996), and Lee et al. (1997). Further, recent important references are Pinedo (2008), and Potts and Strusevich (2009).

Scheduling Problems

Elements of a Scheduling Problem

Usually, a scheduling problem is defined by the following characteristics:

- A finite number n of jobs that need to be processed
- A finite number m of machines

Each job j to be processed can be detailed by using different characteristics of features:

- The release date r_j defined as the time when the job is ready for processing
- The due date d_j, defined as the completion date required for the job
- The weight w_j defined as the relative importance of the job in respect to the others
- The setup time s_j defined as the time needed (e.g., for tooling) before processing the job j

For each machine i and for each job j, p_{ij} defines the processing time.

Additional elements of a scheduling problem may be:

- Precedence constraints among jobs (*prec*).
- The presence of sequence-dependent job setup times (SDJST): when immediately scheduling job k after scheduling job j on machine i, a s_{ijk} setup time is required.
- Preemption (*prmp*): the job execution on a machine may be interrupted and continue in a later time.
- Blocking constraints (*block*): blocking may arise in the presence of limited buffers among machines arranged in series. The upstream machine will be blocked until the downstream buffer is not full. The most common modeled situation foresees zero buffer size.
- No-wait constraints (*nwt*): jobs are not allowed to wait, during execution, between machines. This is a normal situation in continuous manufacturing environments.
- Breakdown (*brkdwn*): machines are not continuously available mostly because of scheduled maintenance or shifts.
- Batch processing (*batch*): the machine can process a number b of jobs simultaneously, and the duration of the process is equal to the longest job processed in the batch.
- Common operations (*cos*): jobs share operations; once the shared operation i is completed, it is completed for all the jobs; that require i.

With respect to machines and layout, the scheduling problems can be classified in:

- Single-machine problems: the simplest machine environment with a single machine.
- Flow-shop problems: when the m machines are arranged in series. Under this layout, the job is constrained to visit the job in the same sequence from the first machine to the last machine (permutation flow shop). More general hypothesis allows overtake of jobs (general flow shop). In flexible flow shop problems machines are arranged in stages. Jobs are constrained to visit stages in sequence, but each stage has a set of identical machines, and in some cases, transportation system may allow overtake (Baffo et al. 2013).
- Job-shop problems: machines are arranged in a job-shop layout and each job has its own route among machines.
- Open-shop problems: The machines are multipurpose and the result of the scheduling is the determination of the route of the jobs among the machine together with the sequencing of the job processing.

Notation

The most commonly used notation for theoretical scheduling is the so-called three-field notation introduced by Graham et al. (1979). The notation foresees three fields called $\alpha|\beta|\gamma$. The first field defines the scheduling environment: α equal to F stands for flow shop, J for job shop, P for parallel machines,

and *O* for open shop. A specific number may be inserted to indicate the number of machines. The second field specifies job characteristics indicating, for example, the preemption, ready times, and additional resources. The third field specifies the performance index (objective). Some examples can be the following: $1|s_{jk}|C_{\max}$ denotes a single-machine problem with sequence-dependent setup times and minimization of maximum completion time (*makespan*); $Jm||C_{\max}$ denotes a job-shop problem with *m* machines *and makespan* minimization.

Objectives

The objective of the scheduling problem is to minimize a performance indicator computed on the schedule. A schedule is the specification of a feasible sequence of starting (and waiting) times of operations for each job in each machine.

For a single job *j*, the most frequent performance indicators are:

- C_j: the completion time of job *j* intended as the date when the job *j* exits the system
- F_j: the flow time of job *j,* i.e., the total time spent in the shop by the job *j,* $F_j = C_j - r_j$
- L_j: the lateness of the job *j*, $L_j = C_j - d_j$
- T_j: the tardiness of the job *j*, $T_j = \max\{0, L_j\}$
- E_j: the earliness of the job *j*, $E_j = \max\{0, -L_j\}$
- U_j: Number of tardy jobs. The job is late: if $(C_j > d_j)$ $U_j = 1$; else $U = 0$

Most common objectives of scheduling problems are the minimization of:

- *Makespan,* defined as $C_{\max} = \max_{(j \in J)}\{C_j\}$
- Maximum lateness $L_{\max}$
- Maximum tardiness
- Weighted sum of completion times
- Weighted sum of jobs tardiness
- Weighted sum of late jobs
- Weighted sum of lateness – earliness

Types of Scheduling and Scheduling Problems

Dynamic and Stochastic Scheduling

Dynamic scheduling is considered when the list of jobs to schedule is partially or totally unknown at the beginning of the schedule. In this case, the three-field notation is expanded in a four-field notation. The first field defines the distribution function of arrivals for jobs. A typical dynamic scheduling environment is the task processing problem in computation and communication systems. Stochastic scheduling, in a more general meaning, is defined when a scheduling problem has a random feature. One common random feature considered may be the processing time of jobs.

Decentralized Scheduling

In decentralized or distributed scheduling, there are local schedulers responsible for local or shared resources who must coordinate in order to compose a global, optimized schedule of given jobs. The discipline of distributed scheduling is considered not only in manufacturing science (Shen 2002) but also in distributed artificial intelligence, in multiagent systems, and in economic theory. One of the most famous applications of distributed scheduling is the contract-net protocol (CNP) proposed by Smith (1980), where each resource is modeled as a local scheduler agent. The CNP coordinates task allocation, providing dynamic allocation and natural load balancing. A way to connect the theoretical classical scheduling framework with decentralized scheduling is proposed by research which investigates the multiagent scheduling problem in a formal way (Agnetis et al. 2004). Duffie and Prabhu (1994) introduced real-time scheduling for decentralized, heterarchical manufacturing systems. Recent research trends consider scheduling in distributed artificial intelligence and swarm robotics.

Integrated Scheduling Problems

Together with decentralized scheduling, a current trend in deterministic scheduling is the so-called integrated scheduling problem. Complex real manufacturing and logistics environments impose the integrated analysis of planning and scheduling problems. Another direction of research is actually the integration along the supply chain of the scheduling problem, considering in the scheduling problems inventory, distribution, and routing issues.

Complexity and Solution Approaches

Most of the scheduling problems proposed are NP-hard in the strong sense and difficult to solve even for small instances. A list of complexity results for scheduling problems can be found on Brucker and Knust (2009). Exact approaches for solving scheduling problems are indeed very rare. Heuristics and metaheuristics are very often used. Relaxation techniques such as column generation, Lagrangian relaxation, or branch and cut can be used when the mixed integer linear programming (MILP) model of the scheduling problem is formulated. For dynamic and stochastic scheduling, dispatching rules may be designed. Some classes of problems (e.g., problems with sequence-dependent setup times) can be treated with graph theory.

Cross-References

- ▶ Artificial Intelligence
- ▶ Distributed Manufacturing
- ▶ Optimization in Manufacturing
- ▶ Planning
- ▶ Production Planning
- ▶ Statistical Process Control

References

Agnetis A, Mirchandani PB, Pacciarelli D, Pacifici A (2004) Scheduling problems with two competing agents. Oper Res 52(2):229–242

Baffo I, Confessore G, Stecca G (2013) A decentralized model for flow shop production with flexible transportation system. J Manuf Syst 32:68–77

Błażewicz J, Ecker KH, Pesch E, Schmidt G, Weglarz J (1996) Scheduling computer and manufacturing processes. Springer Science & Business Media, Berlin

Brucker P, Knust S (2009) Complexity results for scheduling problems. http://www.informatik.uni-osnabrueck.de/knust/class/. Accessed 12 Mar 2016

Conway RW, Maxwell WL, Miller LW (1967) Theory of scheduling. Courier Corporation, New York

Duffie NA, Prabhu VV (1994) Real-time distributed scheduling of heterarchical manufacturing systems. J Manuf Syst 13(2):4–107

Graham RL, Lawler EL, Lenstra JK, Rinnooy Kan AHG (1979) Optimization and approximation in deterministic sequencing and scheduling: a survey. Ann Discret Math 5:287–327

Jackson JR (1956) An extension of johnson's results on job lot scheduling. Nav Res Logist Q 3:201–203

Johnson SM (1954) Optimal two and three-stage production schedules with setup times included. Nav Res Logist Q 1:61–67

Lawler EL (1973) Optimal sequencing of a single machine subject to precedence constraints. Manag Sci 19:544–546

Lee C-Y, Lei L, Pinedo M (1997) Current trends in deterministic scheduling. Ann Oper Res 70(1):1–41

Pinedo M (2008) Scheduling: theory, algorithms and systems, 3rd edn. Springer, New York

Potts CN, Strusevich VA (2009) Fifty years of scheduling: a survey of milestones. J Oper Res Soc 60:41–68

Shen W (2002) Distributed manufacturing scheduling using intelligent agents. IEEE Intell Syst 17(1):88–94

Smith WE (1956) Various optimizers for single-stage production. Nav Res Logist 3:59–66

Smith RG (1980) The contract net protocol: high-level communication and control in a distributed problem solver. IEEE Trans Comput C-29(12):1104–1113

Schema

- ▶ System

Science of Measurement

- ▶ Metrology

Second-Hand Use

- ▶ Reuse

Section Bending

- ▶ Bending (Tubes, Profiles)

Self-Excited Vibration

► Vibration

Self-Excited Vibrations in Metal Cutting

► Chatter

Self-Propelled Rotary Tool

Hossam A. Kishawy
Department of Automotive, Mechanical and Manufacturing Engineering, Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, ON, Canada

Synonyms

Cutting tool; Machining; Rotary tools; SPRT

Definition

Self-propelled rotary tool (SPRT) is a general term which is usually entitled to a family of round cutting tools in the form of circular inserts that spin around their axis during machining operation (Shaw et al. 1952; Venuvinod and Rubenstein 1983; Armarego et al. 1991, 1993, 1994a, b). The SPRTs offer a superior performance over the conventionally used cutting tools where the tool rotates continuously which provides a fresh part of the cutting edge into the cutting area. The insert spinning around its center provides a way for carrying the fluid to the tool point as in the case of a journal bearing. This rotation allows the tool to be cooled down; hence, it significantly reduces the adverse effects of temperature on the tool life as well as the workpiece surface quality. In addition, employment of self-propelled rotary tools results in a higher material removal rate (MRR) in machining difficult to cut materials such as hardened steels and titanium alloys.

Theory and Application

Introduction

Machining is a general term commonly used to define the process of material removal from a part to achieve a desired shape by means of a cutting tool. In machining operations, a wedge shape cutting tool is pushed against and moved over the workpiece surface to remove the unwanted materials and generate the part with desired shape and geometry. Regardless of the process, all of the machining operations are typically conducted in two distinct stages, namely, roughing and finishing. In some cases, a third stage is also added to the abovementioned categories which is called semi-finishing.

The main objective of the roughing operations, as the first stage in machining, is to remove the bulk of material from the workpiece as quick as possible. The result of this stage is a part with shape close to the desired form with no dimensional or surface accuracy. Finishing operations are then implemented to remove the remaining undesired materials and accomplish the job by achieving the final dimension with the desired tolerances and surface finish.

Almost all of the traditional machining operations share the same fundamental principles to remove the unwanted materials. In conventional machining operations, when the cutting tool or insert enters the cutting zone, it remains engaged with the workpiece for the entire pass. During this period of time, a certain part of the cutting edge is constantly in touch with the workpiece and plastically deforms the workpiece material to form chips. Plastic deformation absorbs energy and generates heat; as a result, cutting tool experiences

high mechanical and thermal stresses when it is engaged with the workpiece.

Heat is an influential factor yet complex to study that influences the successful accomplishment or failure of a machining operation. Accurate prediction of temperature distribution and its intensity during machining is not straightforward as the behaviors of workpiece material and temperature are interrelated parameters. The characteristics of workpiece material alter with temperature, while, in the same manner, variation of temperature during machining is also affected by the material characteristics. From one viewpoint, the elevated temperature in the cutting zone can be considered as a desirable factor to some extent as it locally softens the workpiece material and facilitates the ease of machining; however, constant exposure to high temperature drastically reduces the tool life and dramatically increases the machining costs. Moreover, keep machining with the blunt tool not only further intensifies the temperature and repeatedly damages the tool but also deteriorates the surface quality of the workpiece. The temperature-induced problems become even more challenging when it comes to machining difficult to cut materials due to their inherent characteristics such as high hardness (higher than 45 HRC up to 70 HRC, e.g., hardened steels) or very low thermal conductivity and modulus of elasticity (e.g., titanium). The amount of temperature generated during machining of these materials is extremely high; hence, only relatively expensive cutting tools such as ceramics, polycrystalline cubic boron nitrides (PCBN) (Kishawy and Wilcox 2003), and polycrystalline diamonds (PCD) that are capable of sustaining elevated temperatures can be employed in machining such materials.

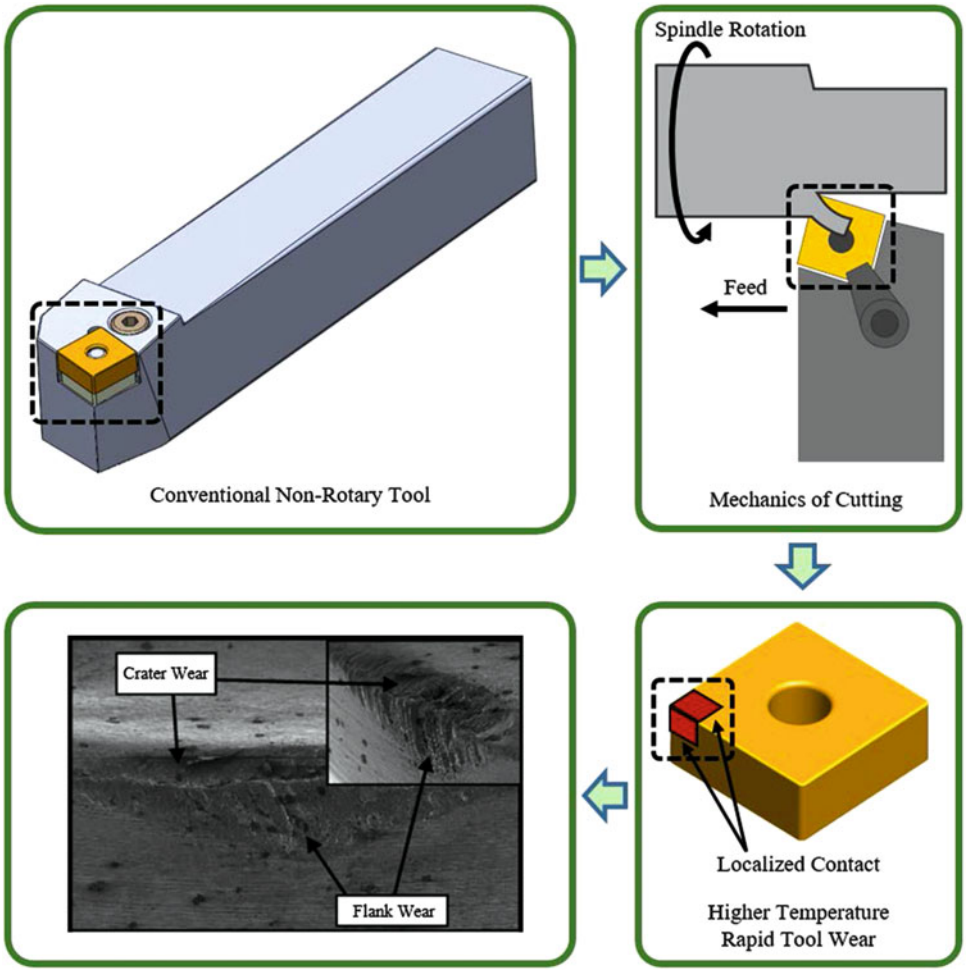
One of the most commonly used approaches to eliminate the adverse effects of high temperature on the tool life and integrity of workpiece is employing coolant. However, application of coolant is not fully advantageous and has its own drawbacks. These drawbacks include but are not limited to corrosion if the coolant is not properly selected, environmental side effects if the standard disposal procedures are not appropriately followed, as well as negative effects on the operator's health. As a result, finding a cheap yet effective way to minimize the temperature-related issue was always a never ending job in machining industry.

A more effective yet simple solution to reduce the temperature-related issues and protect the cutting tool is to replace the active portion of the tool or in the other words by constantly providing a fresh cutting edge to the cutting zone (Kishawy et al. 2004). This may look unrealistic at the first glance, while it is the fundamental idea behind the introduction of a new generation of cutting tools entitled rotary tools.

What Are Rotary Tools?

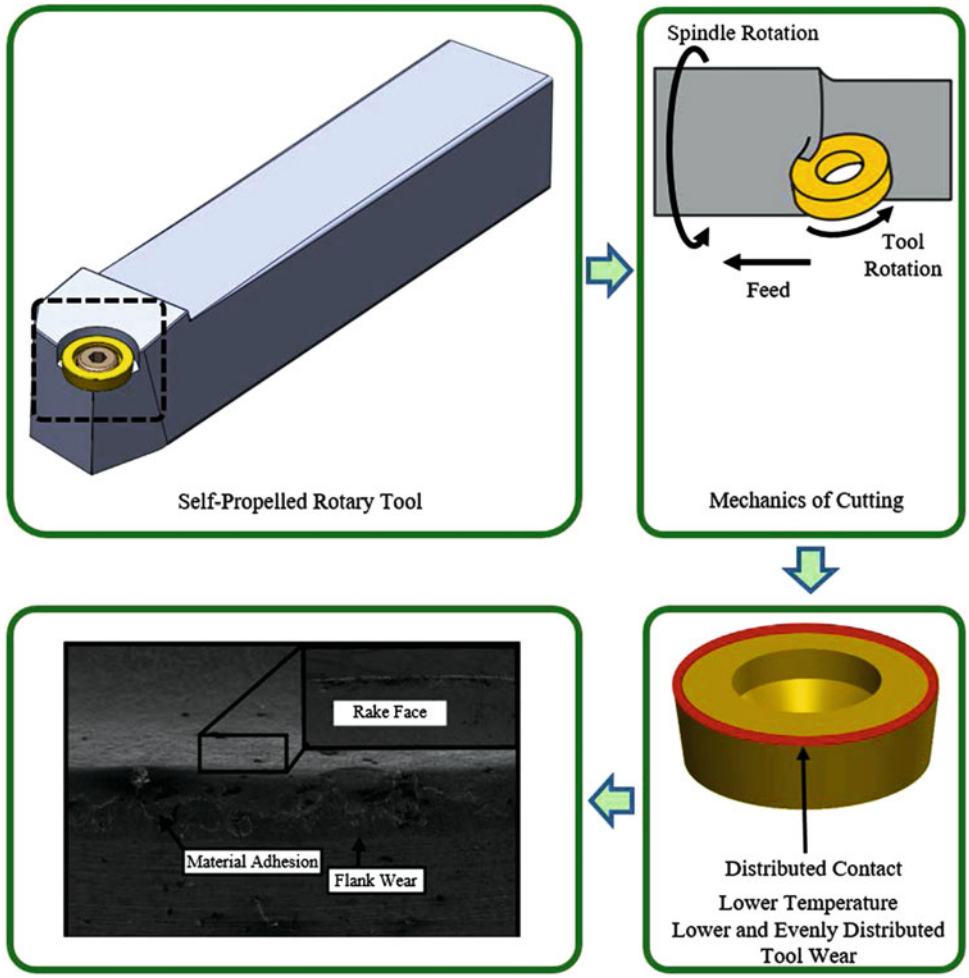
In order to describe the rotary tools and their effectiveness in machining technology, the major difference between rotary tools and their stationary counterparts must be primarily described. This major difference is the type of relative motion between the tool and workpiece. In conventional machining operations such as milling, tuning, and drilling tools, the tool motion is usually confined to the main cutting and feed motions. It may come into question that the milling tools are rotary tool spinning about their axis. A careful attention to this case reveals the fact that the rotation of milling tool provides cutting motion same as the rotation of workpiece in turning; thus, once a unique point on the cutting tool engages the workpiece, it remains in cut until the tool leaves the workpiece. This mechanism of cutting constantly exposes a certain portion of the cutting edge to an elevated temperature which in most of the cases gradually, if not rapidly, deteriorates the tool and makes it blunt. The dominant mechanisms of wear in such cases are crater wear on the rake face as a result of elevated temperature due to constant rubbing of the chip against the same area of the tool and also flank wear due to contact ploughing between the flank face of the tool and newly machined surface (Kishawy and Wilcox 2003, Kishawy et al. 2004) (see Fig. 1).

In rotary tools, another degree of freedom will be added to the system. This degree of freedom is the relative motion between the tool itself and tool holder in terms of rotation of the cutting



Self-Propelled Rotary Tool, Fig. 1 Machining with non-rotary tools and its dominant wear mechanisms

insert around its axis. Rotary tools are typically circular inserts capable of rotating freely about their axis. There are two types of rotary tools, namely, driven and self-propelled. In self-propelled rotary tools (SPRT), the insert is driven as a result of engaging the tool with the workpiece in the absence of any external actuator. In this case the inclination angle and the friction between the chip and the rake face of the tool are the main driving source of the insert. In case of the driven rotary tools, the insert is forced to rotate by means of an external source such as motor or actuator. Due to their inherent functional characteristic, in one full rotation of the rotary tools, each portion of the cutting edge performs cutting for a short period of time and then disengages from the cutting zone which provides a cooling period which ends when this portion of the insert reaches the workpiece again. The rotation dissipates the accumulated temperature to the surrounding environment and allows the tool face to be cooled down during the disengagement period. This exactly resembles a tool with several cutting edges over which the tool wear can be uniformly distributed. This configuration significantly reduces the tool wear and increases the material removal rate (Kishawy et al. 2011) (see Fig. 2).



Self-Propelled Rotary Tool, Fig. 2 Machining with self-propelled rotary tools and dominant wear mechanisms

References

Armarego E, Smith A, Karri V (1991) Mechanics of cutting model for simulated oblique rotary tool cutting processes. J Mater Process Technol 28(1):3–14

Armarego E, Karri V, Smith A (1993) Computer-aided predictive models for fundamental rotary tool cutting processes. CIRP Ann Manuf Technol 42(1):49–54

Armarego E, Karri V, Smith A (1994a) Fundamental studies of driven and self-propelled rotary tool cutting processes – I. Theoretical investigation. Int J Mach Tools Manuf 34(6):785–801

Armarego E, Karri V, Smith A (1994b) Fundamental studies of driven and self-propelled rotary tool cutting processes – II. Experimental investigation. Int J Mach Tools Manuf 34(6):803–815

Kishawy H, Wilcox J (2003) Tool wear and chip formation during hard turning with self-propelled rotary tools. Int J Mach Tools Manuf 43(4):433–439

Kishawy H, Becze C, McIntosh D (2004) Tool performance and attainable surface quality during the machining of aerospace alloys using self-propelled rotary tools. J Mater Process Technol 152(3):266–271

Kishawy H, Pang L, Balazinski M (2011) Modeling of tool wear during hard turning with self-propelled rotary tools. Int J Mech Sci 53(11):1015–1021

Shaw M, Smith P, Cook N (1952) The rotary cutting tool. Trans ASME 74(Aug):1065–1076

Venuvinod P, Rubenstein C (1983) The principle of equivalent obliquity and its application to rotary cutting. CIRP Ann Manuf Technol 32(1):53–58

Self-Vibratory Drilling

Henri Paris
Université Grenoble Alpes, GSCOP Laboratoire des Sciences pour la Conception, l'Optimisation et la Production de Grenoble, Grenoble, France

Synonyms

Deep drilling; Drilling; Vibratory drilling

Definition

The self-vibratory drilling is an innovative technology that allows increasing the productivity, without coolant. The poor removal of chips in deep drilling of small diameter is often the cause of tool breakage and poor quality surface. The self-vibratory drilling enables the chip to be split, thanks to the axial vibrations of the drill self-maintained by the cutting energy. Therefore chips are evacuated easily. A specific tool holder with a variable axial stiffness was developed.

Theory and Application

Introduction

The drilling is a machining operation to obtain a hole in a workpiece with a twist drill. This operation is difficult because the cutting edge is embedded in the material. The fragmentation of the chip is difficult. The drill geometry, including the orientation of the drill tip and propeller of the flute, is supposedly designed to evacuate directly the chip tape or force it to undergo a second deflection by winding on itself which can lead to the breakup. The majority of the holes are shallow on industrial parts (5 × diameter in most cases), and the evacuation of the chip is easy.

Deep drilling may be defined as a hole whose depth is greater than five times the diameter. These holes were primarily designed to bring a fluid at a specific location (lubrication, injection channel, etc.). Deep drilling almost exclusively uses monobloc twist drills, which limits the range of application for the machining of small diameter holes, statistically lower than 13 mm. These operations are greatly facilitated by the use of lubrication through the center of the spindle with pressure ~ 40 bars optionally available on almost all modern machine tools.

The drilling of deep holes of small diameters required the use of retreat cycles and successful lubrication to evacuate the chip. The main limit for a higher productivity in drilling operation was directly connected to the poor chip evacuation. It cannot be a reference solution to industrialize parts.

Dry machining processes are much better for environment than classical machining processes using coolant. There are still machining processes where coolant is part of physical process, not only for thermal matters. The deep drilling is a difficult operation because the drill flute must evacuate the chips formed at the cutting edge. A new technology was developed to be a dry technology. The splitting of the chip in deep drilling can be obtained by vibrating the drill axially.

Self-Vibratory Drilling Principle

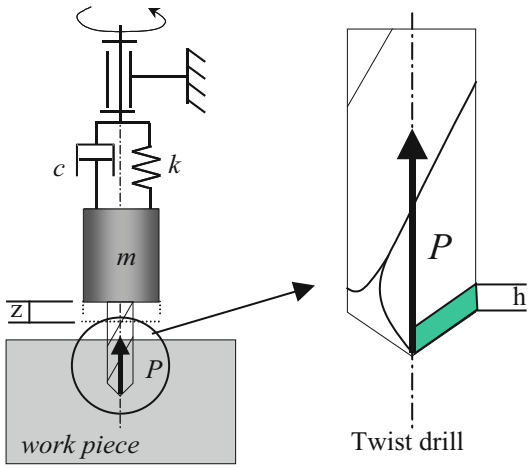
The approach consists in using the cutting energy to create the axial vibrations necessary to split the chip. The objective here is to choose the cutting conditions in order to obtain drill axial chatter. The axial vibrations at low frequency have an amplitude bigger than the feed rate allowing to split the chip which is easily evacuated.

The principle adopted is to insert a spring between the spindle and the twist drill to get a consistent stiffness with the phenomenon of chatter leading to the fragmentation of the chip. A sliding connection between the spindle and the twist drill is also required to maintain, on the one hand, good coaxiality between the axis of the spindle and the drill axis and, on the other hand, to allow the axial movement from vibrations.

During self-vibratory drilling, the axial vibrations that cause the splitting of the chip are obtained by the instability of the cut (Paris et al. 2005). The chip is then naturally evacuated without coolant.

Elements constituting the machining system (machine tool, work piece, fixture, and cutting tool) have stiffness higher than the axial stiffness of the spring between the spindle and the drill. Furthermore, the spring has a very big rigidity in twisting and a low rigidity in compression. From these hypotheses, the vibrations in twisting and the dynamic behavior of the machine tool, fixture, and workpiece can be neglected. The dynamic behavior of the machining system was modeled as a second-order system characterized by a mass (m), stiffness (k), and damping (c) (Fig. 1). (h) is the instantaneous thickness of the chip and (P) is the cutting thrust force.

The surface generated by a lip of the drill is the result of the axial feed of the drill (combination of axial vibratory and feed rate) and of the spindle speed. The following lip then machines this surface. This regeneration of the chip thickness is modeled via the block diagram presented in Fig. 2.



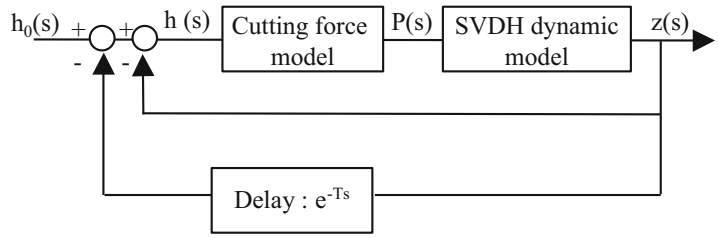
Self-Vibratory Drilling, Fig. 1 Dynamic model of self-vibratory drilling

The cutting forces in drilling are modeled by a thrust force and a torque around the drill, with the hypothesis that the sharpening of the tool is perfect and that the lips of the drill work in the same way, the lateral force being on every lip counterbalance.

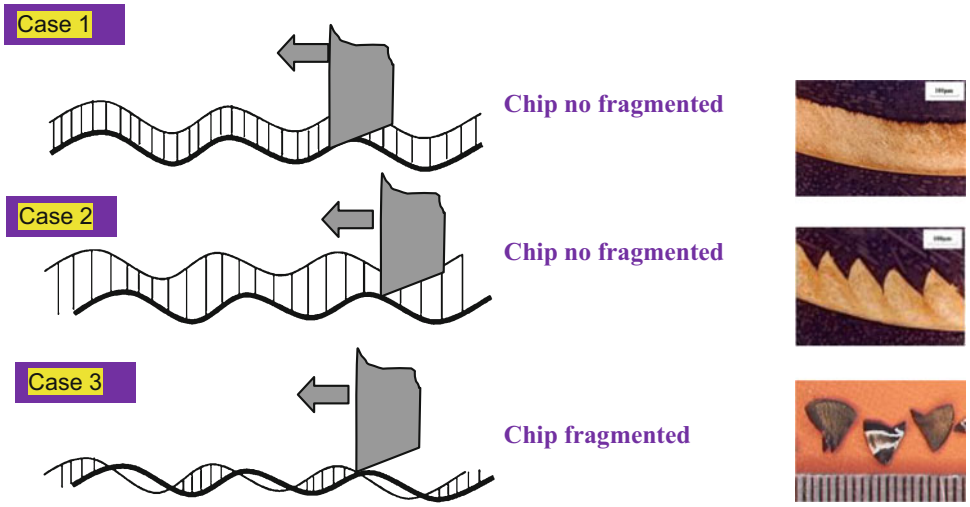
The machined shape is a wavy surface owing to vibrations of the machining system (Fig. 3). This surface is then machined by the following lip of the twist drill. Three situations can arise:

- Case 1: the shape of the machined surface is perfectly superimposed to that generated by the passage of the previous lip. In this case, the thickness of the chip is always the same for each lip passing. The system is stable and the chip is not fragmented.
- Case 2: the shape of the machined surface does not overlap perfectly to that generated by the passage of the previous edge. The thickness of the chip is not always the same in each pass of the lip. The system is unstable and the chip is not fragmented.
- Case 3: the shape of the machined surface is offset by a value close to a half wave. The chip thickness varies with each pass of the cutting edge. The chip is fragmented since the cutting edge is outside the material. The system is highly unstable and we have the presence of the chatter phenomenon. Figure 4 shows the geometry of the bottom of the hole when the chip is fragmented.

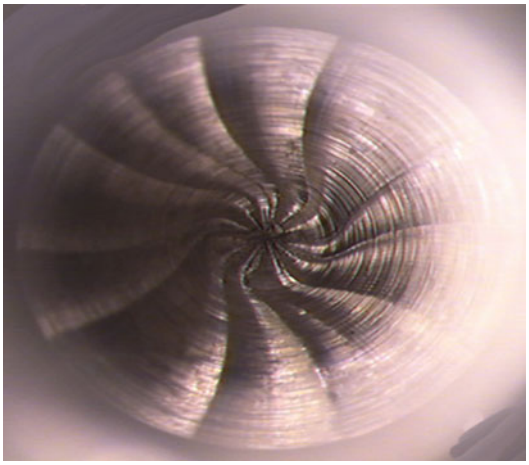
The chatter prediction and linear stability analysis (Altintas 2000) of the machining system modeled allow identification of the stable and unstable domains according to the cutting parameters of depth of cut and the spindle speed. As the width of cut is imposed by the drill diameter, it is



Self-Vibratory Drilling, Fig. 2 Block diagram of chatter dynamics



Self-Vibratory Drilling, Fig. 3 Influence of the phase shift between two lips passing on splitting chip



Self-Vibratory Drilling, Fig. 4 Geometry of the bottom of the hole in case 3

more interesting to draw these stability lobes according to three variables: the stiffness of the spring, the spindle speed, and the mass of the drill holder (Fig. 5). The zone outside the limits corresponds to conventional drilling because it does not provide sufficient energy for the chatter. The zone inside the limits covers the vibratory field of instability, which is an interest in the context of self-vibratory drilling.

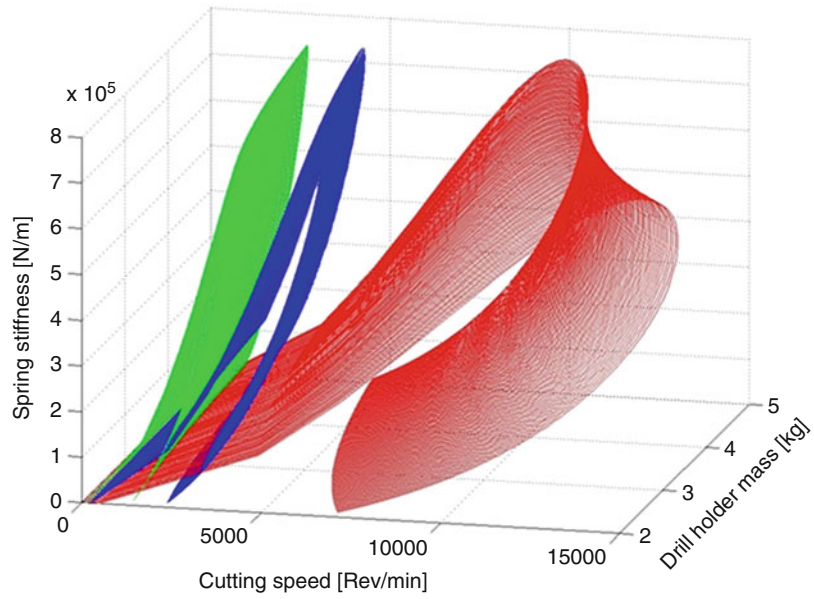
The aim is to use the natural axial chatter to fragment the chips. The challenge is to keep them stabilized at a suitable frequency and magnitude for a good quality of cutting. So, the cutting parameters are chosen to be in the unstable domain. All the values (k, m, and cutting speed N) belonging to the unstable domain have no interest. Indeed, the amplitude of the vibrations must be compatible with the feed rate in order to have a chip fragmentation and with the flank angle to avoid plowing effect, which tends to dampen the vibrations.

Self-Vibratory Drilling Head

An innovative tool holder for drilling operations to obtain a fragmented chip using cutting conditions compatible with good productivity and life of the drill, while ensuring a good quality of the deep hole (diameter, straightness, roughness), was developed in G-SCOP laboratory (Fig. 6).

It is composed of three main elements; the drill holder called sliding body, with a mass m, fixes the twist drill and enables axial vibrations. The vibrating device body connects the device to the spindle and the leaf spring, with a stiffness (k), is a dynamic cutting fixture enabling a magnitude and an adapted frequency. The cutting parameters are chosen to be in the area of unstable cutting.

The self-vibrating drilling head (SVDH) generates chatter for adequate cutting parameters by using the small stiffness of the spring located



Self-Vibratory Drilling, Fig. 5 Stability lobes according to the stiffness of the spring, spindle speed, and drill holder mass for a drill diameter 5 mm

Self-Vibratory Drilling, Fig. 6 Components of SVDH and SVDH with a twist drill

between the body and the drill holder. This low-rigidity part creates conditions for controlled axial regenerative vibrations.

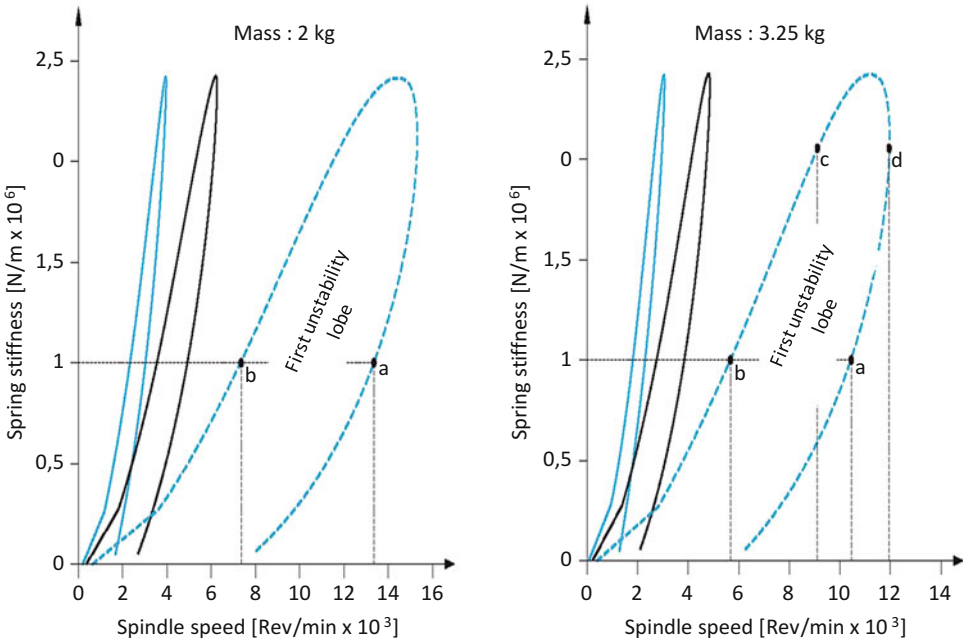
The objective here is to choose the cutting conditions in order to obtain drill axial chatter. The axial vibrations at low frequency have amplitude bigger than the feed rate allowing splitting the chip which is easily evacuated.

Adjustment Parameters of the Self-Vibratory Drilling Head

The adjustment parameters of the SVDH are the stiffness of the spring, the mass of the drill holder, and the spindle speed. It is indeed difficult to modify the other parameters: the diameter of the drill is imposed by the machining operation. To facilitate the choice of the parameters, it is interesting to determine the stability zones according to the stiffness of the spring, the mass of the drill order, and the spindle speed (Fig. 4).

All the values belonging to unstable domains are of no interest. Indeed, the amplitude of the vibrations must be compatible with the flank angle to avoid the plowing effect, which tends to dampen the vibrations. Smith (Smith and Tlusty 1997) showed the interest to place it in the unstable domains of the first lobes (high rotation speed) to avoid the damping. On one hand, the presence of plowing leads to the drill degradation and on the other hand to the attenuation of the vibrations, and its consequence is a ribbon chip. Furthermore, this amplitude must be higher than the feed by a lip to obtain short chips.

The study of the influence of the speed of rotation is done by keeping the other two parameters mass and stiffness constant. Figure 7 shows the stability lobes for a drill diameter of 5 mm into a steel to 300 HB. The mass of the drill holder is 3.25 kg (Fig. 7 right). By positioning stiffness at 1000 N/mm, it is possible to study the influence of the spindle speed. The spindle speed range is more important when one is situated in the first lobe, where the spindle speed is between 5800 (b) and



Self-Vibratory Drilling, Fig. 7 Stability lobes according to the stiffness of the spring, spindle speed for a drill diameter 5 mm

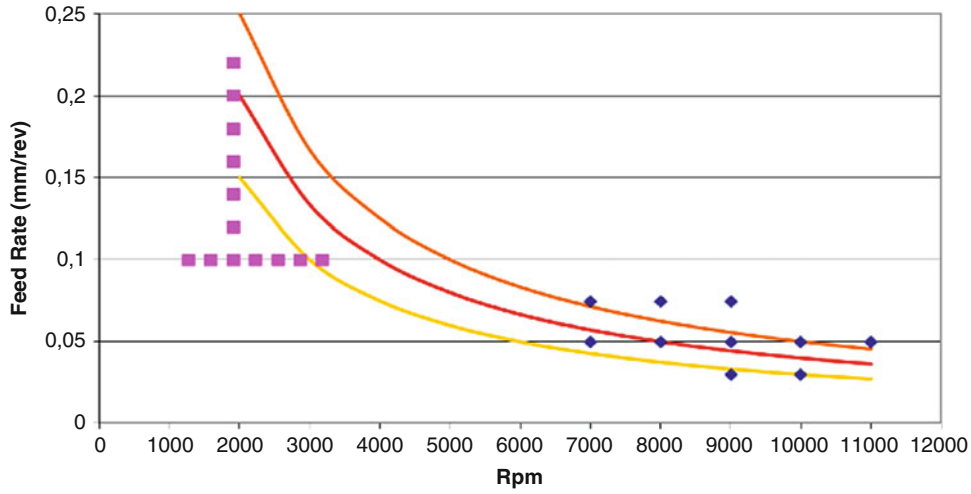
10,300 (a) rev/min, between 2700 and 3850 rev/min for the second lobe, and between 1800 and 2400 rev/min for the third lobe. Beyond the third lobe, it is very difficult to obtain cutting edge jumps out of the material. For the first lobe, the chatter frequency is close to 1.7 vibrations per revolution. This means that the cut is interrupted 1.7 times for a revolution of the twist drill. Since the twist drill has two lips, then the chips are fragmented into 3.4 pieces per revolution. They are very small and easily evacuated by the flute of the drill. The chatter frequency is 3.5 for the second lobe and 5.5 for the third lobe.

The influence of the mass is observed maintaining stiffness and spindle speed. Figure 7 shows the lobes of instability for two different masses: a mass of 3.25 kg and a mass of 2 kg. For stiffness of 1000 N/mm and for a mobile mass of 3.25 kg, the position of the first instability lobe is for cutting speeds of 5800 rev/min (b) and 10,300 rev/min (a). For the same stiffness and lower mass (2 kg), the unstable zone is between 7500 rev/min (b) and 13500 rev/min (a). The increase in mass translated all of instability lobes toward the low spindle speeds. For a high spindle speed, use low mass. In contrast, for a low spindle speed, choose a large mass.

The influence of the stiffness is observed in keeping a constant mass. The influence of the stiffness of the spring is visible in Fig. 7 right. The stiffness has an effect on the width of the unstable cutting regions. With a stiffness of 1000 N/mm and a mass of 3.25 kg, it is possible to obtain an unstable cut for spindle speeds between 5800 rev/min (a) and 10,300 rev/min (b) to level of the first lobe. With a stiffness of 2000 N/mm and a mass of 3.25 kg, it is possible to obtain an unstable cut between 9000 rev/min (c) and 12,000 rev/min (d). Finally, with an excessive stiffness (greater than 2500 N/mm), whatever the spindle speed, it is not possible to obtain an unstable cut.

Productivity

We now compare productivity between the two of obtaining deep-hole techniques: a long drilling with high pressure lubrication and vibratory drilling without lubrication. The operating conditions are different between a traditional drilling and a



Self-Vibratory Drilling, Fig. 8 Operating range and iso-productivity curves

self-vibratory drilling. Indeed, the temperature at the cutting edge is lower in self-vibratory drilling, and it is possible to increase the cutting speed. However, the chip thickness must remain compatible with the mechanical characteristics of the drill. As we have a variable chip thickness in self-vibratory drilling and cutting rates close to 50%, it is not possible to choice a big feed rate. A feed rate of 0.15 mm/rev in self-vibratory drilling leads to a maximum chip thickness corresponding to an advance of 0.3 mm/rev in traditional drilling.

It can also be seen on Fig. 8 that the productivity of self-vibratory drilling exceeds the productivity of traditional drilling: the pink dots correspond to the cutting conditions recommended by the manufacturer of the twist drill during traditional drilling, and the blue dots correspond to the conditions used in vibratory drilling in order to obtain a good split of the chip. The three curves of iso-productivity, which represent three different material removal rates, show that the performance of self-vibratory drilling exceeds those achieved in traditional drilling with a twist drill. The energy consumption of the drilling operation is substantially identical in both cases.

Dry Machining

The vibratory drilling is compatible with dry drilling. This new technique allows extending the dry machining of delicate operations confined. This technological leap can meet the new environmental standards.

The use of coolant in traditional drilling generates at least 90% of the environmental impacts during the drilling operation (Paris and Museau 2012). Self-vibratory drilling is equivalent to the classical technology of deep drilling without any coolant in terms of environmental impacts.

Lifespan of Twist Drills and Quality of the Deep Hole

The purpose of the wear tests was to identify the lifetime of a carbide drill of diameter 5 mm on the drilling of crankshaft lubrication holes. The holes diameter is 5 mm and the depth is 100 mm. The crankshaft is made of steel 35MnV7.

The results of the wear tests are very encouraging. To the last hole, the vibration of the drill is preserved. The wear of the tool does not seem to have any influence on the fragmentation of the chip.

During our experimental studies, several results on the lifespan of a twist drill have been obtained. In each case, the lifespan of a twist drill, defined by the length of drilled holes, exceeds 40 m during self-vibratory drilling.

The three parameters on hole quality are as follows: the drilling diameter, the straightness, and the surface roughness of the hole at the beginning and end of the hole.

The diameter of the hole is constant at 4.98 mm during the first half of the life of the drill and then gradually decreases to 4.96 mm at the end of life of the drill.

The straightness was around a mean deviation of 150 microns and remains permanently below 300 microns.

Roughness measured at the end of hole gradually decreases with the length of cumulative drilling. The quality of the surface seems to be improving with the wear of the tool, which is relatively unusual in machining. Industrially, this presents a huge advantage for the vibratory drilling. Indeed, if the quality of the surface of the hole is good during the first holes, it will remain so throughout the use of the tool.

Cross-References

- ▶ Chatter
- ▶ Chatter Prediction
- ▶ Drilling

References

Altintas Y (2000) Manufacturing automation: metal cutting, mechanics, machine tool vibrations, and CNC design. Cambridge University Press, Cambridge, UK

Paris H, Museau M (2012) Contribution to the environment performance of the dry-vibratory drilling technology. CIRP Ann Manuf Technol 61(1):47–50

Paris H, Tichkiewitch S, Peigne G (2005) Modelling the vibratory drilling to foresee cutting parameters. CIRP Ann Manuf Technol 54(1):418–421

Smith S, Tlusty J (1997) Current trends in high-speed machining. J Manuf Sci Eng 119(4B):664–666

Semiductile Mode Machining

- ▶ Ultraprecision Grinding

Semisolid Metal Forming

- ▶ Thixoforming

Sensing Element

- ▶ Sensor (Assembly)

Sensor (Assembly)

Ming C. Leu
Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, Missouri, USA

Synonyms

Detector; Measurement device; Measuring probe; Sensing element; Transducer

Definition

A sensor is any functional unit that records a desired information from a process and provides it for subsequent processing (CIRP Dictionary of Production Engineering 2004).

A sensor is typically a device that transforms signals from the mechanical, thermal, radiant, chemical, or magnetic domain to the electrical domain, but it could measure electric values as well. A single sensor may be based on cross-effects between different signal domains to achieve the signal transformation to the electrical domain to be read by a human or an electronic instrument. These cross-effects are shown in Table 1, where the input signal domains are on the left-hand side, and the output signal domains are at the top (Meijer 2008).

Sensors can also be divided into passive and active sensors. Passive sensors achieve their output energy from an input signal such as a force or pressure; active sensors achieve it from an internal source such as a chemical balance (Meijer 2008).

With the MEMS (micro-electro-mechanical systems) technology and information technology developed in the past decades, sensors also include the "integrated smart sensor," which

integrates a sensor, an analog interface circuit, an analog-to-digital converter (ADC), and a bus interface in one chip, and, hence, can process signals and provide decisions. Furthermore, multiple "smart sensors" can communicate with each other and with remote users through the internet to construct a "sensor network," which can sense the environment with a group of sensors and deliver the information remotely.

Theory and Application

Sensor Types and Applications

Many assembly tasks are performed manually, but high-volume assembly tasks are often performed by special-purpose machines that form an automated assembly line or a semiautomated assembly line. With a multisensor assembly station, the assembly task can be accomplished with high accuracy, repeatability, and efficiency.

Assembly tasks can be considered to consist of two main functions: (1) sorting parts and (2) mating parts in the final assembly (Soloman 2010). To sort parts, their parameters, type, and position/orientation have to be sensed and sent to mechanical devices to perform containerization, feeding, fixturing, and gripping. During part mating, the present state of the parts to be manipulated needs to be monitored by sensors and compared with a priori information, and then modifications of the parts' positions can be performed to mate them finally in the required assembly pattern.

Sensors commonly used for assembly are listed in Table 2 and discussed in the following. The table depicts that most sensors used in assembly are based on cross-effects between different domains. A comprehensive review of sensor technology in assembly systems was given by Santochi and Dini (1998). There were two recent review articles on sensors for assembly used on robotic hands (Almassri et al. 2015; Saudabayev and Varol 2015).

Photoelectric Sensors

A photoelectric sensor consists of a source and a detector, as shown in Fig. 1 (Soloman 2010).

Sensor (Assembly), Table 1 Physical sensor effects

In/Out	Radiant	Mechanical	Thermal	Electrical	Magnetic	Chemical
Radiant	Photo luminance	Radiant pressure	Radiant heating	Photo conduction	Photomagnetic	Photo chemical
Mechanical	Photoelastic effect	Conservation of moment	Friction heat	Piezo-electricity	Magnetostriction	Pressure-induced explosion
Thermal	Incandescence	Thermal expansion	Heat conduction	Seebeck effect	Curie-Weiss law	Endotherm reaction
Electrical	Inject luminance	Piezoelectric effect	Peltier effect	PN junction effect	Ampere's law	Electrolysis
Magnetic	Faraday effect	Magnetostriction	Ettingshausen effect	Hall effect	Magnetic induction	
Chemical	Chemo luminance	Explosion reaction	Exothermal reaction	Volta effect		Chemical reaction

Sensor (Assembly), Table 2 Sensors commonly used in assembly

Sensor/Domain	Radiant	Magnetic	Mechanical	Electrical
Photoelectric	√			√
Vision	√			√
RFID	√			√
Inductive		√		√
Magnetic		√		√
Force-torque			√	√

The source can be a light-emitting diode (LED) that emits a beam of light either in the infrared or visible light spectrum, or a laser. The detector can detect the light presence or absence with a photodiode. An amplifier is used in the detector so that it can respond to the light emitted by the source without being affected by the ambient light.

Based on the target, three detection methods are applied by photoelectric sensors: through-beam detection, reflex detection, and diffuse detection (Soloman 2010).

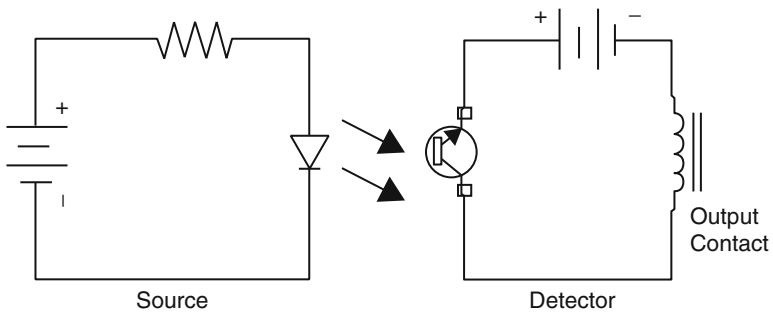
The through-beam detection method requires that the source and detector are aligned accurately to make the greatest possible amount of light travel from the source to the detector (Fig. 2). If an object passes between the source and detector, the light will be broken and a signal of object detection will be generated. This method is suitable for detecting opaque or reflective objects. It provides the longest range among three methods and a high power at a shorter range, which make it suitable in a contaminant environment.

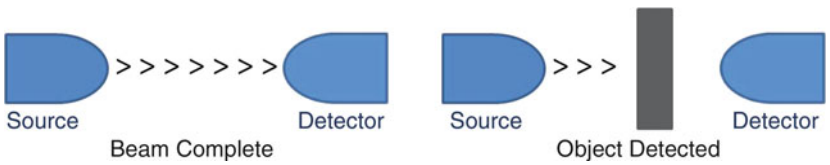
The reflex detection method requires the source and the detector in one unit, as shown in Fig. 3. The light from the source is transmitted to a retroreflector and returned to the detector. The object is detected when it passes between the sensor and the reflector and breaks the light. This method is flexible and easy to install, but the detected object must be less reflective than the retroreflector.

In the diffuse method, the source and the detector are in one unit, as shown in Fig. 4. When an object is present before the sensor, the light from the source is diffused from the object's surface at all angles. When the detector receives enough reflected light, the object is detected. If the source light is laser and the detector is a position sensor detector (PSD), the ability to detect small parts, high switching speeds, or longer sensing distances can be achieved (Fig. 5).

One type of photoelectric sensor is the photoelectric color sensor, which uses a true color technology to provide accurate information. A halogen light source is used to illuminate the target. The reflected light is transmitted over a fiber-optic bundle and received by a detector with RGB optical filters, which analyze the reflected light into red, green, and blue (RGB) color contents. An analog signal is generated based on RGB color contents and then amplified and converted to digital data, which is then processed by the microprocessor to meet specific requirements by using different algorithms. A photoelectric color sensor can provide high-speed operation (1-ms response time) for most manufacturing environments and even hazardous environments with a fiber-optic cable. It is simple

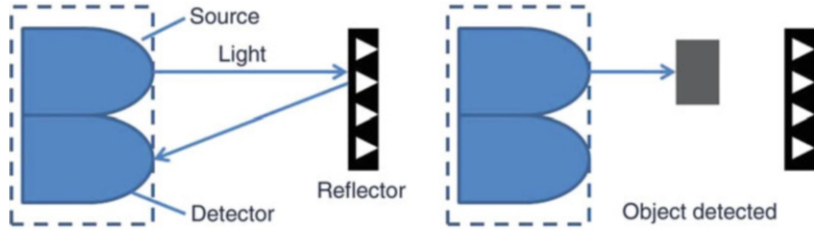
Sensor (Assembly), Fig. 1 Photoelectric sensor

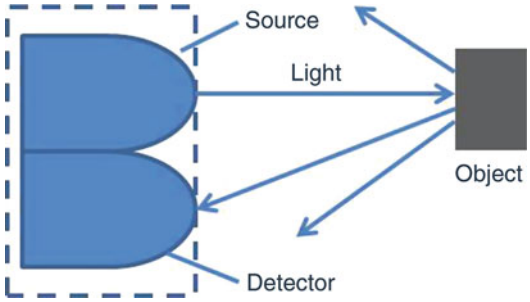




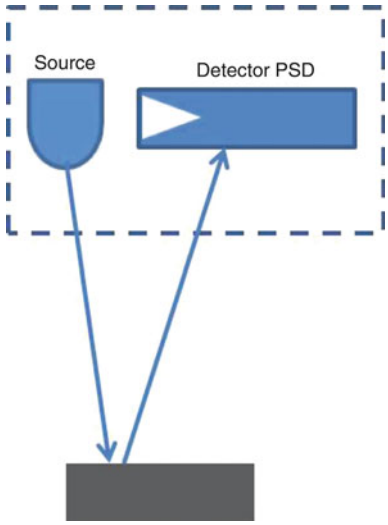
Sensor (Assembly), Fig. 2 Through-beam detection

Sensor (Assembly), Fig. 3 Reflex detection





Sensor (Assembly), Fig. 4 Diffuse detection



Sensor (Assembly), Fig. 5 Diffuse detection with PSD

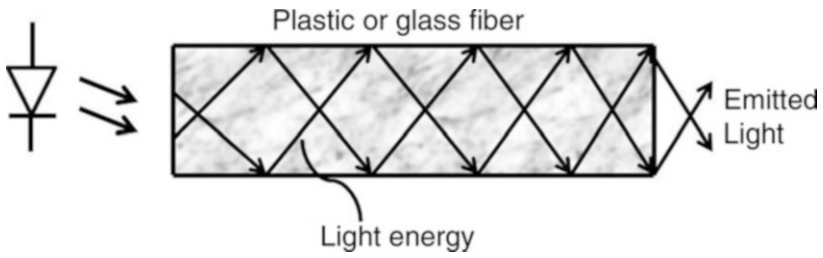
to set up and calibrate the color sensor and needs less than a few minutes to start operating the sensor.

Another type of photoelectric sensor is the fiber optics sensor, which uses a bundle of transparent fibers of glass or plastic to conduct and guide light energy using the principle of total internal reflection, as depicted in Fig. 6. Fiber optics serve as "light pipes" to transmit the light from the source to the detector. With a small diameter and flexibility of the fibers to be bent and twisted, fiber-optic sensors can be placed in confined places.

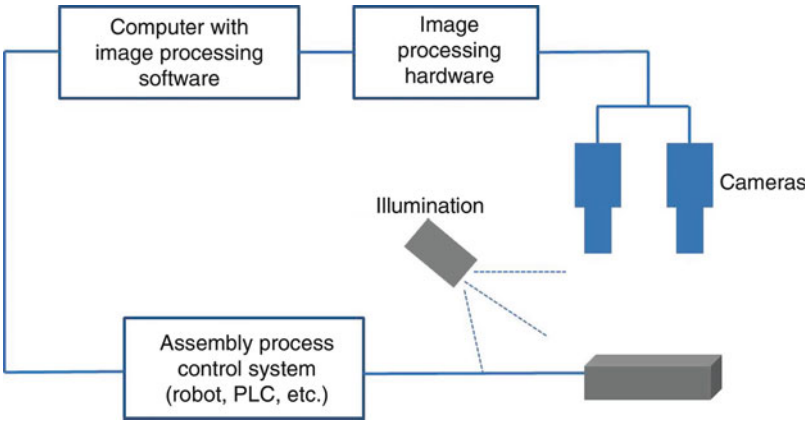
In industrial application, through-beam slot sensors have been embedded into the conveyor in an assembly line to control the material flow and feed equipment. Through-beam sensors are also used to control the product height such as stack height and lipstick height before capping. Reflex sensors are mostly applied to count discrete components through the conveyor in the assembly line such as bottles, cans, or cartons by detecting the presence of objects. The laser diffuse sensor with background suppression is usually used to detect small parts and ensure quality. For example, it has been used to determine the orientation of an IC chip and to detect items of varying heights with the help of PSD. The fiber-optic sensor has been successfully installed in some hard-to-access locations such as beneath the product surface to verify that screws are correctly seated (STEP 2012).

Vision Sensors

More and more assembly processes rely on vision sensors for automation. A vision system performs the following functions: image acquisition, image processing, feature extraction, and decision-making (Malamas et al. 2003). Typically, it includes single or multiple cameras, illumination devices, as well as image processing hardware and software, as shown in Fig. 7. Usually, the scene to be monitored needs to be illuminated for achieving better image features and facilitating image processing and classification. The cameras are configured and calibrated to get the



Sensor (Assembly), Fig. 6 Fiber optics with total internal reflection



Sensor (Assembly), Fig. 7 A typical vision sensor system

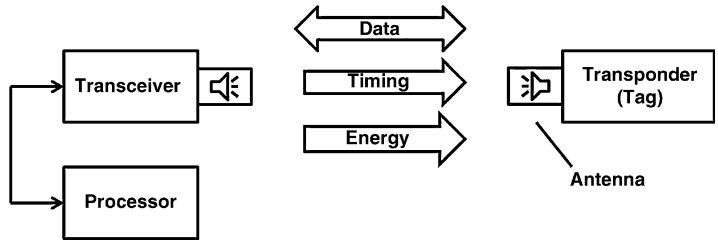
configuration data. Images containing the required information are acquired in a digital form and then are processed to remove the background noise. A set of uncorrelated features such as size, position, and contour and texture measurement are computed and analyzed by statistical or other computing techniques. Depending on the application, a decision is made by processing a set of features with thresholds, statistical or soft classification. As an economical solution, vision sensors are the ideal choice when different details need to be automatically scanned as part of a 100% quality inspection at the end of a production line.

Most industrial applications with vision sensors can be divided into the following four types: dimensional inspection, surface quality inspection, assembly inspection, and operation inspection. In the assembly line, vision sensors are widely used in robots as the machine vision to help them recognize the positions/orientations of objects to be handled or assembled, to determine the presence or absence of parts, and to detect parts which do not meet required specifications. Zhu et al. (2013) and Chadda et al. (2011) built vision systems using low-cost cameras such as Wii Remotes and Firefly MV cameras to track the object or human motion in an assembly line. Object motion data was used to generate assembly simulation for training, and human motion data was used to do ergonomic analysis to help design a better assembly cell.

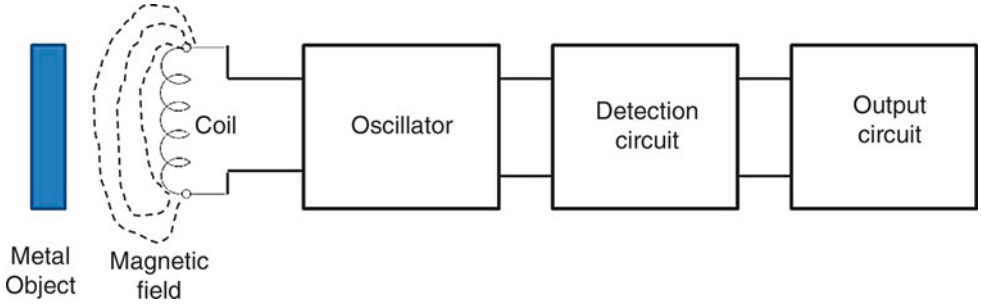
Radio-Frequency Identification (RFID) Sensors

RFID is a technology capable of providing wireless identification of objects. Figure 8 shows a typical RFID system that consists of a transponder (a tag) including an integrated circuit containing an RF circuitry and information to be transmitted, with a transceiver to read the radio frequency and transfer the information to a processing device.

RFID tags can be divided into three types: passive, active, and semi-passive. A passive tag does not have any power source and is only activated by an approaching transceiver. However, both active and semi-active tags need a power source such as a button cell battery.



Sensor (Assembly), Fig. 8 A typical RFID system



Sensor (Assembly), Fig. 9 Inductive sensor component and principle

When required by a transceiver, the tag uses power from its internal battery or from power harvested from the transceiver's electromagnetic field to modulate or demodulate the signal and transmits it through the antenna. The signal generated must not create interference with the transceivers. The transceiver picks up the tag's radio frequency and interprets it to meaningful information, and then sends it to a processing device.

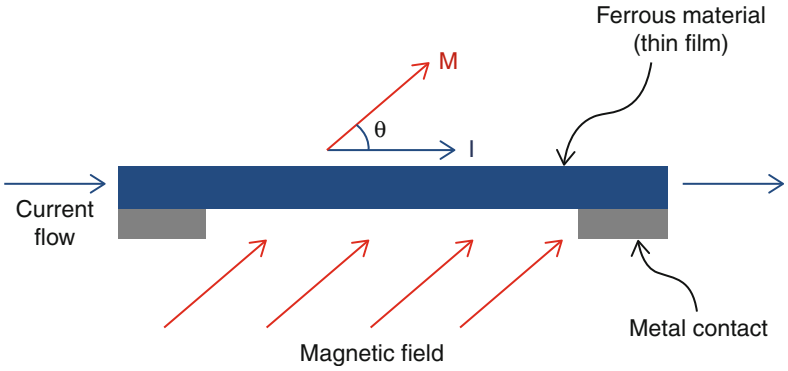
RFID has been used in manufacturing plants for more than a decade. It is used to track parts and work in process and to reduce defects, increase throughput, and manage the production of different versions of the same product. In an assembly line, RFID sensors ensure a reliable exchange of information between material flow and data processing, including all areas of production where materials need to be moved and identified. Tateno (2006) applied RFID onto assembled objects to automatically recognize human movement or assembly states in assembly training. The following have been detected using RFID sensors: (a) workpieces, (b) assembly tools, (c) work points, and (d) work time.

Inductive Sensors

Inductive sensors are used to detect metallic objects without contact between the sensor and the object. An inductive sensor consists of four components: coil, oscillator, detection circuit, and output circuit, as shown in Fig. 9. A small amount of energy is supplied to the coil, and the detection circuit uses this coil to produce an oscillator. The "kill oscillator" principle is implemented in inductive sensors. When the metals enter the changing magnetic field, eddy current is built up on the metal surface through drawing energy from the coil. When more metal enters the field, less energy is left in the coil, and the amplitude of the oscillator is reduced more. Finally, the oscillator is killed by inadequate energy, and an output signal is generated (Soloman 2010).

Inductive sensors often apply a simple principle of discrete on-off to determine a part's presence and for feature detection, hole presence, and nesting validation. Inductive sensors are reliable, accurate, and stable in operation. They can work in a wide range of temperature and are among the easiest sensors to deploy. An inductive sensor should be first considered if the target to be detected is metal (Balluff 2008).

Sensor (Assembly), Fig. 10 AMR principle

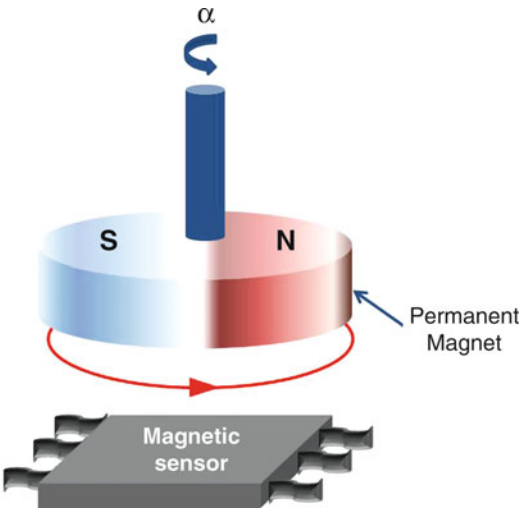


Inductive sensors have been used to detect the positions of rotary index tables, which are standard components used in assembly systems. Inductive sensors are also used to ensure quality control in packaged or bottled food and drink production by detecting metal parts to allow for accurate positioning in filling, labeling, or batch-counting applications.

Magnetic Sensors

Magnetic position sensors are based on anisotropic magneto-resistance (AMR) which exists in certain ferrous material and can be used as a resistive element with thin strip shape, as shown in Fig. 10. The resistance of AMR strip element changes in a relationship with $\cos^2\theta$, where Θ is the angle between the magnetic moment (M) and the current flow (I), as shown in Fig. 10. Four ARM elements can be connected together to form a Wheatstone bridge, and those side contacts in the bridge produce a differential voltage as a function of the supply voltage, MR ratio, and the angle Θ (Honeywell 2002).

A magnetic sensor can provide both position and angle information. For example, Fig. 11 shows a typical magnetic angular position sensor that includes AMR elements and an electronic part to achieve biasing, amplification, offset cancelation, and temperature stabilization. The permanent magnet is mounted on the shaft end of a rotating axis and is magnetized diametrically. The magnetic field through the sensor rotates when the magnet rotates with the shaft. The output can be transformed into the angle α between the magnet field direction and the current flow in the sensor plate, enabling the angular position of the shaft to be calculated (Meijer 2008).



Sensor (Assembly), Fig. 11 Magnetic angular position sensor

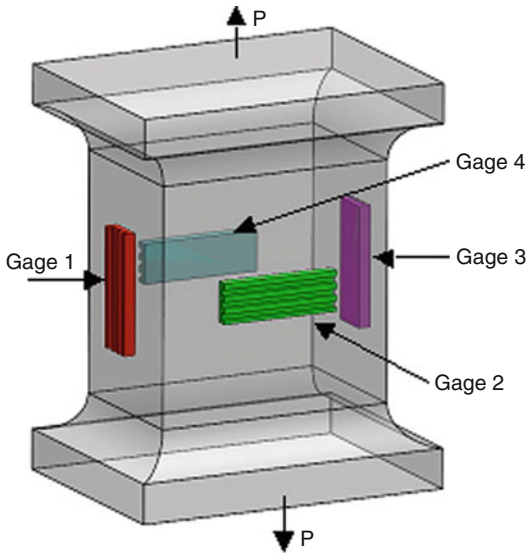
Magnetic sensors are becoming popular to locate moving objects without contact in the assembly line, such as a "pick and place" module which requires precise positioning and high accuracy at high speeds. The magnetic operating principle ensures process reliability even when contamination from oil or dust is present.

Force-Torque Sensors

Transducers for measuring force or torque usually contain an elastic member that converts the mechanical quantity to a linear or rotational

deflection that can be measured by a set of strain gages, which give an electrical signal proportional to the quantity of force or torque.

The commonly used force sensors are load cells. A typical link-type load cell with strain gages constituting the sensor is shown in Fig. 12, where the load P can be either tensile or compressive. The four strain gages are bonded to the link such that two of them (gages 1 and 3) are in the axial direction and the other two (gages 2 and 4) are in the transverse direction. The four gages are wired into a Wheatstone bridge. The load P is applied to the link that can be measured based on changes in the axial and transverse strains.



Sensor (Assembly), Fig. 12 Link-type load cell

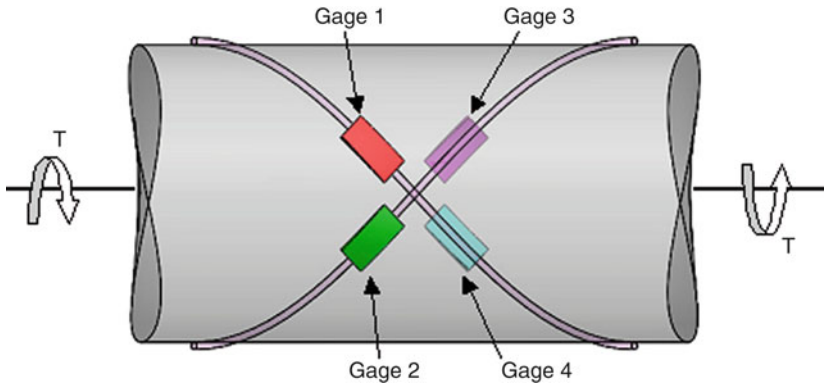
Figure 13 shows a torque cell consisting of a circular shaft mounted with four strain gages on two 45° helices that are diametrically opposite to each other; gages 1 and 3 mounted on the right-hand helix sense a positive strain, while gages 2 and 4 mounted on the left-hand helix sense a negative strain. The two helices define the principal stress and strain directions when the circular shaft is subjected to pure torsion.

Some applications require that force and torque can be measured simultaneously. The combined measurements can be accomplished by using two or more strain-gage bridges mounted on an elastic element or by using selected gage combinations to form a single bridge. An example is given in Fig. 14, which shows a link having a circular cross-section (Dally et al. 1993). Gages A and C are wired into a Wheatstone bridge to measure the axial force P, and gages B and D are wired into a Wheatstone bridge to measure the torque M.

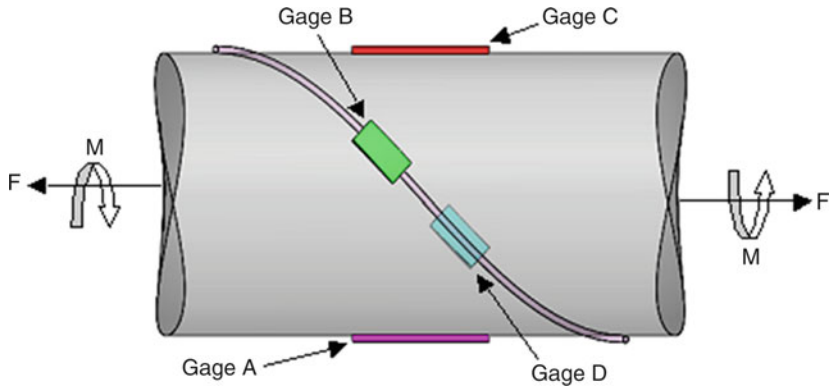
Force sensors are most commonly used in aerospace, medical, and automation measurement platforms, while torque sensors are often used in process monitoring and control. Different types of force and torque sensors can be used together to improve the quality of measurement in manufacturing processes.

Sensor Characteristics

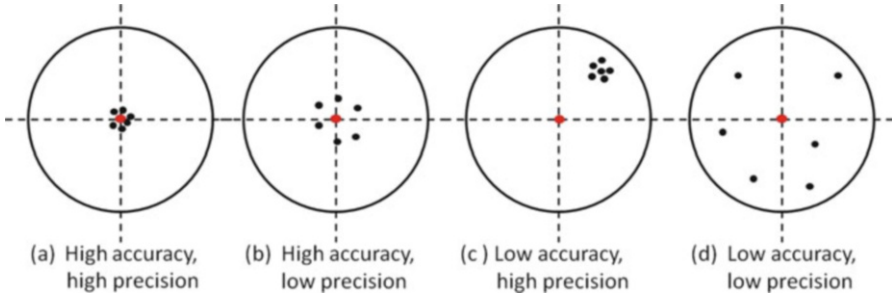
The following characteristics should be considered in the selection of actual sensors according to application requirements (Morris 2001).



Sensor (Assembly), Fig. 13 Torque sensor made by mounting strain gages on a circular shaft



Sensor (Assembly), Fig. 14 Transducer for measuring both axial load and torque



Sensor (Assembly), Fig. 15 Comparison of accuracy and precision

Accuracy

The accuracy of a sensor depicts how close the sensor output reading is to the correct value. The sensor inaccuracy can be represented by the percentage of the maximum reading of a sensor. For example, if a magnetic angular sensor having the range of 0°–360° has a quoted inaccuracy of ±0.1% of full-scale reading, then the maximum error to be expected in any reading is 0.36°.

Precision/Repeatability

Repeatability depicts how close the output readings are when all of them are read under same inputs over a short period of time. The degree of repeatability defines the sensor's precision. Figure 15 shows the difference between accuracy and precision. In this figure, the circle center point is the true value the sensor measures, and black dots are several readings from the sensor under same inputs.

Measurement Range

A sensor's measurement range describes the range of maximum and minimum values that the sensor is designed to measure.

Linearity

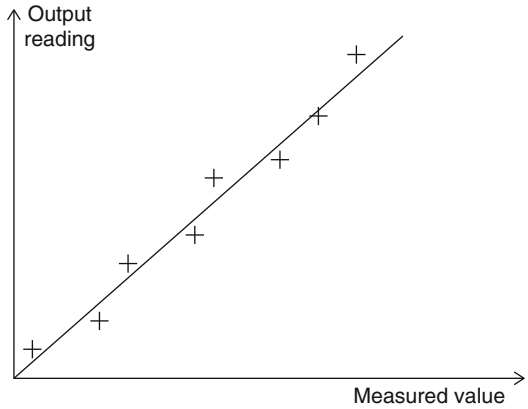
The output readings of a sensor should be linearly proportional to the measured values. A mathematical line-fitting technique can be used to draw a line fitting through output readings, as shown in Fig. 16. The nonlinearity is defined as maximum deviation of any output reading over the full-scale reading, expressed as a percentage of full-scale reading.

Sensitivity

The sensitivity is the ratio of the change in output reading to the change in the value of the measured parameters. It is the slope of the straight line shown in Fig. 16.

Drift (Bias)

Zero drift represents how much the zero reading of a sensor changes with a change in the surrounding condition. The zero drift can be removed by a calibration process. For some sensors, the repeated measurements under same conditions can drift with time.



Sensor (Assembly), Fig. 16 Sensor linearity and sensitivity

Hysteresis

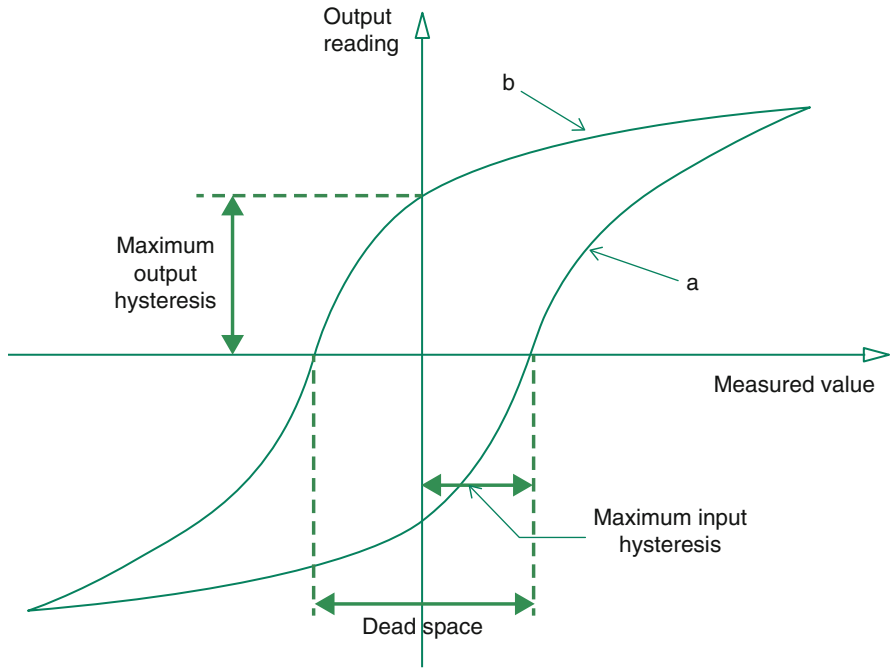
If the value of input parameter changes in sign, the output reading tends to form a loop as shown in Fig. 17, which may also result in a time lag between the input and the output. The non-coincidence between the curve *a* and curve *b* is called hysteresis. The maximum input hysteresis and the maximum output hysteresis are marked in Fig. 17.

Dead Space

Dead space is the range of input value at which the output value is zero. The dead space exists in Fig. 17 with the hysteresis effect, but even without hysteresis, some sensors may still suffer from dead space. Fox example, a sensor with backlash in mating gears has a dead space.

Cross-References

▶ Sensor (Machines)



Sensor (Assembly), Fig. 17 Sensor hysteresis

References

Almassri AM, Wan Hasan WZ, Ahmad SA, Ishak AJ, Ghazali AM, Talib DN, Wada C (2015) Pressure sensor: state of the art, design, and application for robotic hand. J Sens 2015:1

Balluff (2008) Sensors and RFID: the unbeatable team for advance error proofing (White paper). Balluff, Florence

Chadda A, Zhu W, Leu MC, Liu XF (2011) Design, implementation, and evaluation of optical low-cost motion capture system. In: Proceedings of the ASME 2011 international design engineering technical conferences & computers and information in engineering conference (IDETC/CIE 2011), 28–31 Aug 2011, Washington, DC. Paper no DETC2011-47270, pp 1451–1461. https://doi.org/10.1115/DETC2011-47270

CIRP Dictionary of production engineering III: Manufacturing Systems (2004)

Dally J, Riley W, McConnell K (1993) Instrument for engineering measurements. Wiley, New York

Honeywell (2002) Retrieved from application of magnetic position sensors. http://www51.honeywell.com/aero/common/documents/Applications-of-Magnetic-Position-Sensors.pdf. Accessed 4 Dec 2018

Malamas EN, Petrakis EG, Zervakis M, Petit L, Legat J-D (2003) A survey on industrial vision systems, applications and tools. Image Vis Comput 21(2):171–188

Meijer G (ed) (2008) Smart sensor systems. Wiley, Chichester

Morris AS (2001) Measurement & instrumentation principles. Butterworth-Heinemann, Woburn

Santochi M, Dini G (1998) Sensor technology in assembly systems. Ann CIRP 47(2):503–524

Saudabayev A, Varol HA (2015) Sensors for robotic hands: a survey of state of the art. IEEE Access 3:1765–1782

Soloman S (2010) Sensors handbook, 2nd edn. McGraw-Hill, New York

STEP (2012) STEP 2000 (Siemens Technical Education Program) Self-study courses. EandM, San Francisco

Tateno T (2006) Interactive system of work support in consideration of worker competency. JSME Int J Ser C 49(2):576–582

Zhu W, Vader AM, Chadda A, Leu MC, Liu XF, Vance JB (2013) Wii remote based low-cost motion capture for automated assembly simulation. Virtual Reality 17(2):125–136. https://doi.org/10.1007/s10055-011-0204-z

Sensor (Machines)

Robert X. Gao
Case Western Reserve University,
Cleveland, OH, USA

Synonyms

Encoders; Gauges; Probes; Transducers

Definition

The word "sensor" typically refers to a device that converts a change in a physical stimulus into output parameters that can be read by human or an electronic instrument. As an example, a thermocouple that converts the difference in temperature at a reference point and the measurement point into an electrical voltage forms a temperature sensor. The output of a sensor contains information about the state of a physical object (e.g., a machine tool) or a process (e.g., metal forming). Such information can be processed for improved observability in manufacturing machines and processes (Dornfeld 1992; Altintas and Park 2004; Gao et al. 2008; Altintas and Jin 2011), leading to better control and decision-making (Toenshoff and Inasaki 2000).

With the advancement of microelectronics and information technology over the past decades, the definition of "sensor" has been broadened by including the concept of "sensor nodes" or "sensing agents," which are sensing devices equipped with data communication and signal processing capabilities. Such self-contained sensing devices are capable of preprocessing analog signals acquired, extracting characteristic features hidden in the signal, and performing decision-making tasks. Further expansion of the definition of sensors is seen in the recent advancement of multisensor system or "sensor networks," in which a group of sensors cooperate to sense the physical environment and automatically deliver the extracted information to remote users.

Theory and Application

Sensors for Machine Monitoring

Before the widespread acceptance and application of sensors, machines and processes were monitored by human operators, by means of visual observation, hearing, smell, or touch. These inputs are processed by the brain for machine control. The advancement of sensing technologies aims to replace human-based, subjective perception by physical sensors that enable objective measurement with high accuracy, repeatability,

and efficiency and even the ability to sense phenomena that cannot be sensed by humans, such as infrared and ultrasonic waves.

Depending on the type of physical stimulus, sensors used for machine monitoring can be generally classified into five categories:

1. **Mechanical**: for measurement of position, velocity, acceleration, force, pressure, stress, strain, torque, flow velocity, viscosity, etc.
2. **Thermal**: for measurement of temperature, heat flux, thermal conductivity, etc.
3. **Electrical**: for measurement of charge, current, voltage, conductivity, permittivity, impedance, etc.
4. **Magnetic**: for measurement of magnetic field, magnetic flux, permeability, etc.
5. **Radiation (optical)**: for measurement of radiation energy, emissivity, reflectivity, intensity, etc.

Sensing Physics

Sensing physics is the mechanism underlying the operation of sensors that converts an input parameter (e.g., a mechanical force) into an output signal of different nature (e.g., an electrical charge), to enable electrical measurement of non-electrical parameters. Sensing physics is realized through either the properties of the material used for manufacturing the sensors (e.g., piezoelectric effect or Seebeck effect, as listed in Table 1) or the mechanical design of the sensor structure (e.g., MEMS structures for capacitive pressure sensing or acceleration sensing). A single sensor may be based on one or a multiplicity of such mechanisms to achieve the signal conversion from the detected physical phenomenon to the electrical signals to be displayed or transmitted to an end user.

Sensor Evaluation Metrics

The performance of a sensor can be evaluated using standard metrics that have been generally accepted by the international research community as reflected in the table below (Table 2).

In Fig. 1, some of the major metrics that have been commonly used for characterizing sensors are graphically illustrated.

Sensors for Machine and Process Monitoring

Thermocouple

Thermocouple (Fig. 2) is one of the most widely used sensors for temperature measurement and control. The mechanism underlying the function of thermocouples is the so-called thermoelectric

Sensor (Machines), Table 1 Material properties related to sensing mechanisms

	Mechanical	Thermal	Electrical	Magnetic	Radiant
Mechanical	Mechanical and acoustic effects	Friction effects	Piezoelectricity, piezoresistivity	Magnetomechanical effects	Photoelastic systems
		Cooling effects	Resistive, capacitive, and inductive effects		Interferometers
					Sagnac effect
					Doppler effect
Thermal	Thermal expansion		Seebeck effect, thermo-resistance, pyroelectricity		Thermooptical effects
	Radiometer effect		Thermal (Johnson) noise		Radiant emission
Electrical	Electrokinetic and electromechanical effects	Joule (resistive) heating	Charge collectors, Langmuir probe	Biot-Savart's law	Electrooptical effects
		Peltier effect			
Magnetic	Magnetomechanical effect	Thermomagnetic effect	Galvanomagnetic effect		Magneto-optical effects
Radiant	Radiation pressure	Bolometer, thermopile	Photoelectric effects		Photorefractive effects, optical bistability

Sensor (Machines), Table 2 Metrics for sensor evaluation

Metrics	Definition
Sensitivity	The minimum magnitude of input signal required to produce a specified output signal having a specified signal-to-noise ratio or other specified criteria (see Fig. 1a)
Full-scale range	The maximum and minimum values of the measured property
Offset (bias)	Magnitude of the output signal when the measured property is zero (see Fig. 1b)
Nonlinearity	The amount the output differs from ideal behavior over the full range of the sensor, often noted as a percentage of the full range (see Fig. 1c)
Drift	The change of sensor output signal independent of the measured property over a period of time (see Fig. 1d)
Hysteresis	A type of path-dependent error caused by the time lag of sensor response to the dynamics of measured properties (see Fig. 1e)
Noise	The random fluctuation in an electrical signal which can be further categorized as thermal noise, shot noise, flicker noise, burst noise, and avalanche noise, based on the effect of noise generation
Sampling rate	The number of samples taken from a continuous signal per unit of time (see Fig. 1f)

effect or Seebeck effect, named after the German-Estonian physicist Thomas J. Seebeck in 1821.

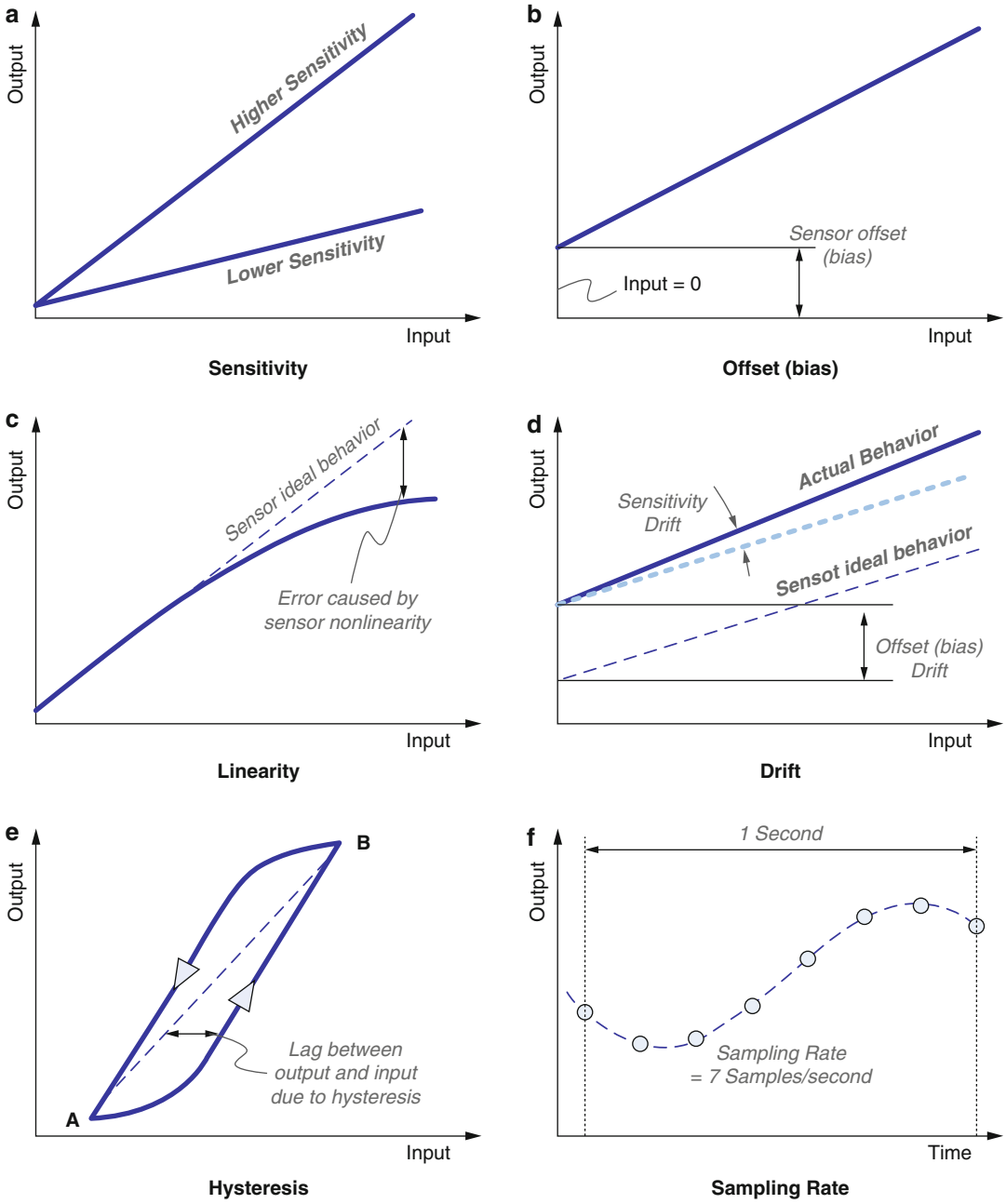
Each thermocouple consists of two different metal alloy conductors joined at two junctions, called the "hot junction" and "cold junction," respectively, as shown in Fig. 2. When the two junctions are maintained at different temperature ranges, the difference will produce an electromotive force (EMF), known as the Seebeck EMF. As a result, an electrical voltage proportional to the temperature will be generated by the thermocouple. The relationship between the voltage and temperature depends on the type of materials. For practical applications, such a relationship is utilized to produce different types of thermocouples for different temperature ranges. Thermocouples have been used in machining system (Arriola et al. 2011) for tool temperature monitoring and error compensation through temperature dependent axis motion.

Strain Gauge

Strain gauges, invented by Arthur C. Ruge and Edward E. Simmons in 1938, have been used for measuring strain on the surface of an object. The sensing physics of strain gauges is rooted in the dependence of the resistance of conductive materials on the material geometry. When the material is stretched within the limit of elasticity, the resistance will change with the deformation, based on the definition of conductivity. A practical strain gauge usually maximizes the length of the conductive material by arranging it in a zigzag pattern to maximize the changes in the output resistance. To convert resistance to voltage signals, strain gauges are typically configured as an arm in a half or full Wheatstone bridge. Strain gauges have also been used as the sensing element in the load cell, torque sensors, or pressure sensors (gas/liquid), as well as vibration detectors in microstructure sensors (Fig. 3).

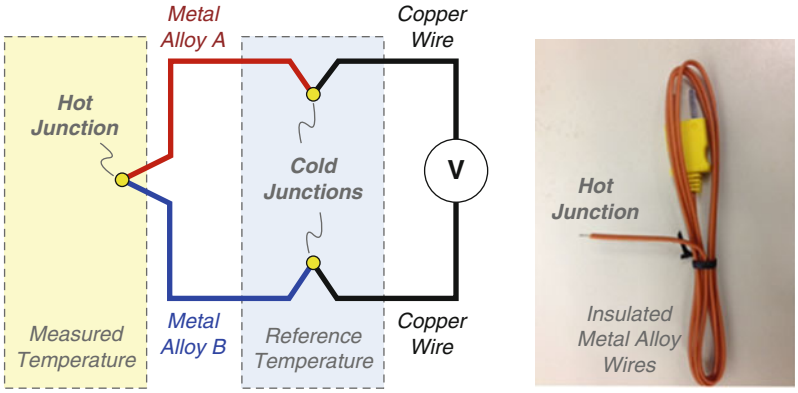
Acoustic Emission Sensors

Acoustic emission sensors are sensing elements that measure the acoustic wave generated by machine components such as metal-cutting tool, stamping die, or grinding wheel during machining processes. When acoustic wave arrives at the sensor surface, the sound pressure causes change in strain within the sensing element, e.g., piezoelectric material, and it produces a dynamic voltage signal according to the piezoelectric effect. Due to the nature of high internal impedance of piezoelectric materials, such a voltage signal needs to be amplified by a circuitry with high input impedance, e.g., a JFET amplifier. Because of the broad bandwidth (typically 100–900 kHz), AE sensors can detect most of the phenomena in machining, although significant data acquisition and signal processing are required to properly interpret the AE signals measured (Fig. 4) (Toenshoff and Inasaki 2000).

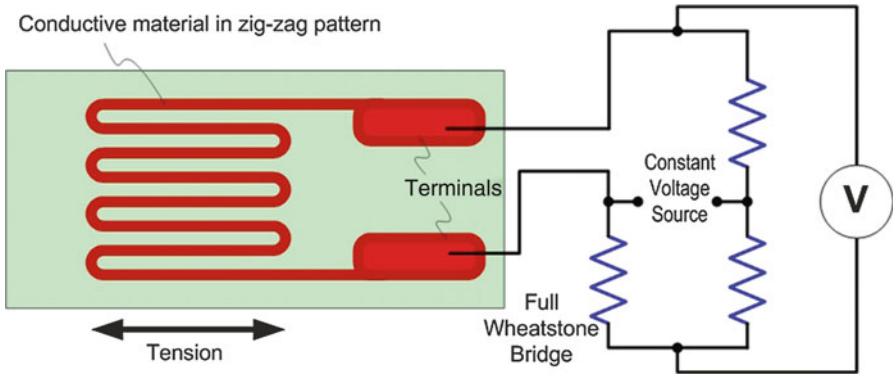


Sensor (Machines), Fig. 1 Illustration of major sensor metrics

Besides piezoelectric materials, AE sensors can also be constructed as having a parallel-plate structure, which converts incoming acoustic waves into variations in the distance between the two plates, which in turn results in the change of capacitance across the two plates. The capacitance change is then converted into voltage signal by a Wheatstone bridge. Compared with piezoelectric



Sensor (Machines), Fig. 2 Working principle and practical illustration of thermocouple



Sensor (Machines), Fig. 3 Strain gauges

acoustic sensors, capacitive sensors have higher accuracy but are very sensitive to sensor position and surface mounting quality, which limit its application in manufacturing process monitoring, where the operating environment is typically harsh for the sensor.

Displacement Sensors

Displacement sensors are typically noncontact sensors measuring the position or change of position of a target. Depending on the sensing physics, displacement sensors either are typically capacitive or use eddy current for displacement sensing. The former employs plates to form a pair of capacitor electrodes between itself and the target surface (metallic or other conductive materials). The change of distance will be converted into voltage by approximately following the principle of capacitance between parallel plates. Eddy current displacement sensors utilize

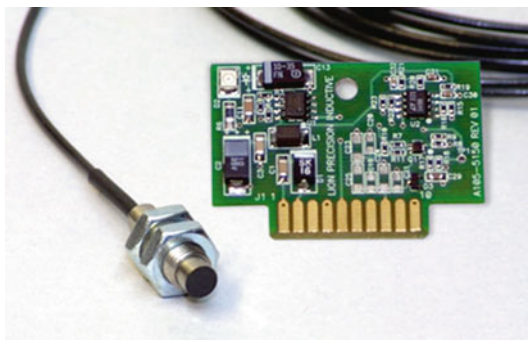
Sensor (Machines), Fig. 4 Acoustic emission sensor (From Kistler)

a high-frequency magnetic field, which is generated by passing a high-frequency current through the sensor head coil. When a metal target is present in the magnetic field, electromagnetic induction causes an eddy current perpendicular

to the magnetic flux passage to flow on the surface of the target. This changes the impedance of the sensor head coil. The sensors measure the distance between the sensor head and target based on the change in the oscillation status. Due to the difference in electrical and magnetic field, capacitive displacement sensors are more sensitive to the presence of particles such as oil or water drops in the sensing range and thus are not as robust as the eddy current sensors in an industrial environment (Toenshoff and Inasaki 2000). Other types of displacement sensors include acoustic and laser-based sensors that utilize the time-of-flight (TOF) of the waves to provide high-precision measurement. Applications of displacement sensors have been demonstrated for positioning objects in the nanometer scale (Fig. 5) (Gao and Kimura 2007).

Embedded Sensing for Stamping Monitoring

In stamping operations, sheet metal is formed into a desired shape by being pressed between suitably shaped dies in a hydraulic or mechanical press (Sah and Gao 2011). As a predominant manufacturing process, sheet metal forming has been widely used for the production of automobiles, aircraft, home appliances, beverage cans, and many other industrial and commercial products. A major effort on stamping processes monitoring has been focused on investigating variations in the press force. Given that the press force itself is an integral of the contact pressure distribution over the die and binder contact interfaces, measuring contact pressure distribution at the tooling-workpiece interface provides direct insight into the localized forming process and the quality of the stamped parts, as compared to measuring the total press force on the supporting structure of the press.



Sensor (Machines), Fig. 5 Eddy current displacement sensor (From Lion Precision)

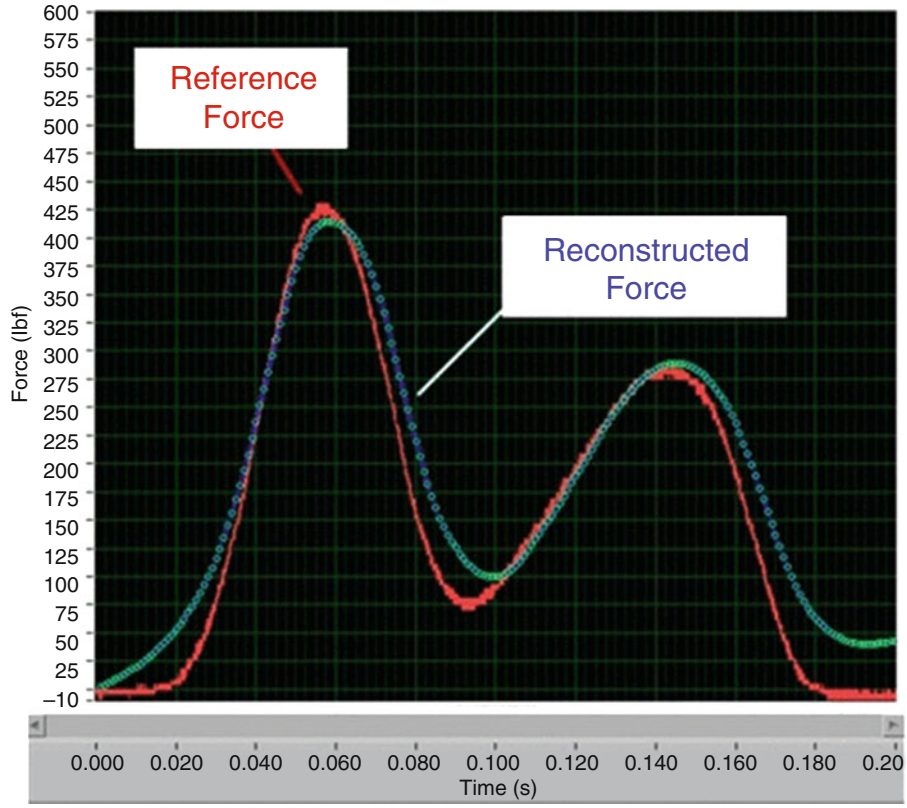
Tonnage sensing refers to the measurement of press forces (from the punch and binder) either directly by piezoelectric load cells or indirectly through piezoresistive strain gauges mounted on the press columns and/or linkages that determine the elastic strain in load-bearing members induced by the press forces during a stamping cycle.

Another approach to stamping process monitoring is the measurement of acoustic emissions (AE) (Liang and Dornfeld 1990). This technique used wideband piezoelectric acoustic sensors that are in direct contact with the sheet metal workpiece within the binder. The energy release rate of the acoustic emissions was used to identify different forming events such as initial impact, shear fracture, and rupture in the stamping operation. Large acoustic activity in the initial stages of plastic deformation can be well represented by AE measurements, but the generally low signal-to-noise ratios encountered in the later stages of the process make AE-based monitoring a difficult endeavor.

Recent effort has investigated tooling embedded sensors for measuring the temporal and spatial distributions of stamping forces on the punch and binder for improved fault diagnosis and process control. Piezoelectric sensors have been flush embedded into the die structure. Processed by numerical surface algorithms (Sah and Gao 2011), force output from the sensors provided input to building a temporal-spatial map representing the actual force distribution at the tool-workpiece interface for die misalignment detection, as illustrated in Fig. 6.

Cross-References

- ▶ Actuator
- ▶ Mechatronics
- ▶ Monitoring



Sensor (Machines), Fig. 6 Force curve measured by embedded sensor

References

Altintas Y, Jin X (2011) Mechanics of micro-milling with round edge tools. CIRP Ann Manuf Technol 60(1): 77–80

Altintas Y, Park SS (2004) Dynamic compensation of spindle-integrated force sensors. CIRP Ann Manuf Technol 53(1):305–308

Arriola I, Whitenton E, Heigel J, Arrazola PJ (2011) Relationship between machinability index and in-process parameters during orthogonal cutting of steels. CIRP Ann Manuf Technol 60(1):93–96

Dornfeld DA (1992) Acoustic emission feedback for precision deburring. CIRP Ann Manuf Technol 41(1): 93–96

Gao W, Kimura A (2007) A three-axis displacement sensor with nanometric resolution. CIRP Ann Manuf Technol 56(1):529–532

Gao R, Fan Z, Kazmer D (2008) Injection molding process monitoring using a self-energized dual-parameter sensor. CIRP Ann Manuf Technol 57(1):389–393

Liang SY, Dornfeld DA (1990) Characterization of sheet metal forming using acoustic emission. ASME. J Eng Mater Technol 112(1):44–51

Sah S, Gao R (2011) An experimental study of contact pressure distribution in panel stamping operations. Int J Adv Manuf Technol 55(1):121–132

Toenshoff HK, Inasaki I (eds) (2000) Sensors in applications. Sensors in manufacturing, vol 1. Wiley-VCH, Weinheim

Sequencing

► Scheduling

Servo Control

► Computer Numerical Control

Servo System

Kaan Erkorkmaz
Mechanical and Mechatronics Engineering,
University of Waterloo, Waterloo, ON, Canada

Synonyms

Feedback control system; Motion control system

Definition

In a broad context, a control system designed to follow a given reference input that changes as a function of time can be classified as a servo system (Franklin et al. 2005).

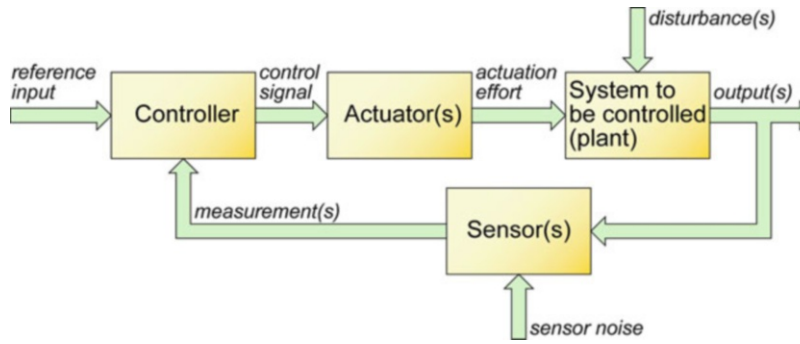
Theory and Application

Description

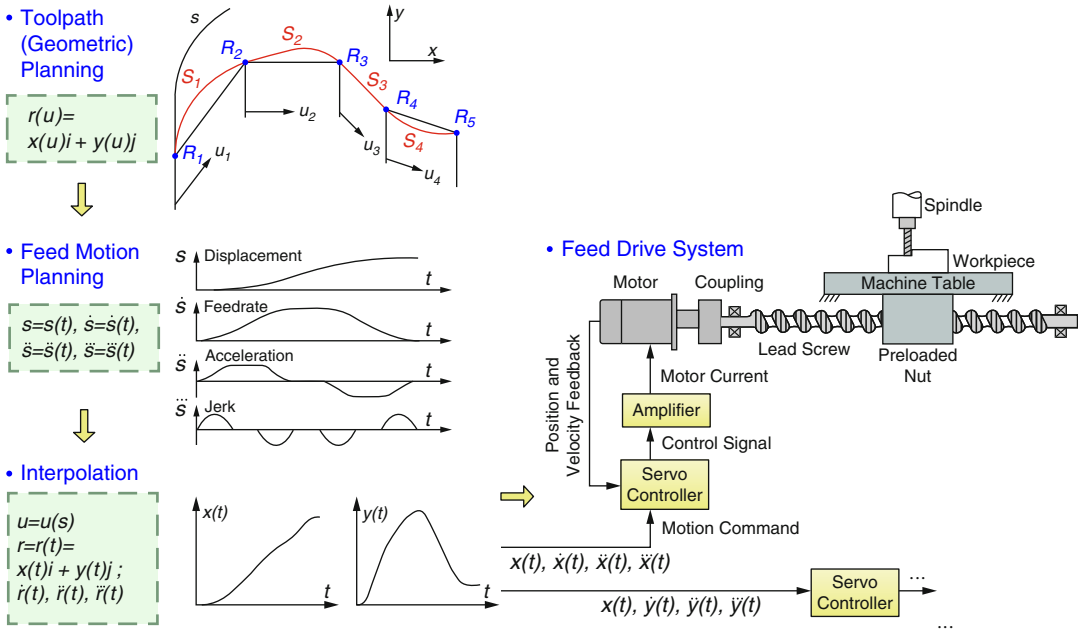
In practical implementation for industrial manufacturing systems, servo systems are frequently used in a feedback configuration, as shown in Fig. 1, where a sensor is (or multiple sensors are) used to measure variables of interest and an actuator is (or multiple actuators are) used to affect change in a system (i.e., plant) being controlled, so that all or part of its states and/or outputs follow the reference input as closely as possible, in spite of disturbances acting on the system, as well as measurement errors (i.e., sensor noise) and changes in the plant dynamics. The controller uses sensor measurement and reference values to determine the actuation effort that needs to be applied to the physical system, and communicates it to the actuators via the control signal. The main aspect that sets a servo system apart from a regulator is that in the latter, the reference input is constant over time, whereas in a servo system, the reference input is expected to change as a function of time, which makes it typically more challenging to track by the system states or outputs.

Examples

In machine tool or robot control, a plant consists of a motion delivery mechanism (e.g., translating or rotating components with specific inertia; a geared or ball screw drive-type motion transmission and conversion systems with inherent mechanical flexibility; and bearings and guiding components). The most commonly used actuators are typically electric motors. In some applications, they can also be pneumatic, hydraulic, or piezoelectric. These actuators are typically powered by electrical amplifiers and in some cases, pneumatic and hydraulic circuits. The sensors are typically position sensing encoders, velocity sensing tachometers, and in some cases, accelerometers. Each of these sensors has specific noise characteristics. Among some of the most commonly used controllers in industry are Proportional-Integral-Derivative (PID) position control, P-PI position-velocity cascade control, and PI motor current control (Franklin et al. 2005; Ogata 1997). Common sources of disturbances in drive systems can originate from the friction and cutting or process forces, especially



Servo System, Fig. 1 General overview of a servo system



Servo System, Fig. 2 Schematic of a servo system being used to control a ball screw-driven machine tool

in machine tools or machining/polishing/assembly robots (Altintas 2000).

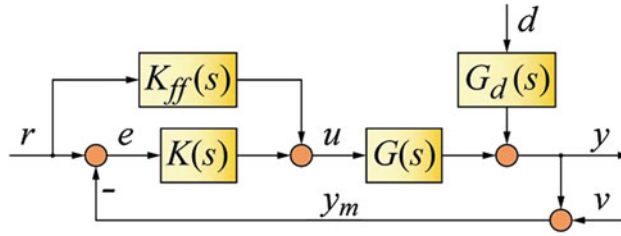
An overview of a servo system being used to control a ball screw driven machine tool has been shown in Fig. 2. In the figure, the trajectory planning is achieved by designing the toolpath geometry and the feedrate profile (i.e., timing information) in a coordinated manner. Interpolating the feed motion, representing the arc displacement with respect to time s(t), then enables the interpolation of the parametric toolpath expressions x(u(s(t))) = x(t), y(u(s(t))) = y(t), etc. This allows for the servo axis level trajectory commands (i.e., set points) to be obtained as a function of time. The position commands are then fed into their respective servo control loops, which enable execution of the desired multiaxis trajectory.

Theory and Practical Application Considerations

A generalized model for a servo-control system can be depicted as shown in Fig. 3 (Skogestad and Postlethwaite 2005). By applying block diagram algebra (Franklin et al. 2005; Ogata 1997), the true output $y(s)$ can be expressed in terms of the reference ($r(s)$), disturbance ($d(s)$), and sensor noise ($v(s)$) as:

$$y(s) = \underbrace{\frac{G(s)K_{ff}(s)}{1+G(s)K(s)}}_{G_{\text{track}}:\text{Tracking Transfer Function (TF)}} r(s) + \underbrace{\frac{1}{1+G(s)K(s)}}_{G_{\text{dist}}(s):\text{Disturbance TF}} G_d\, d(s) - \underbrace{\frac{G(s)K(s)}{1+G(s)K(s)}}_{G_{\text{noise}}(s):\text{Noise TF}} v(s) \quad (1)$$

In frequencies where it is desired to have the servo system output ($y(s)$) accurately replicate the reference input $r(s)$, it is essential to have $G_{\text{track}}(j\omega) \approx 1$. In addition, for successful disturbance rejection we need to maintain $|G_{\text{dist}}(j\omega)| \gg 1$. This is typically achieved by having



$G(s)$	plant response to control input (including actuator dynamics)	$r(s)$	reference profile
		$y(s)$	actual output
$G_d(s)$	plant response to disturbance input	$y_m(s)$	measured output
		$v(s)$	measurement noise
$K(s)$	feedback controller dynamics	$e(s)$	measured tracking error
		$u(s)$	control signal
$K_{ff}(s)$	feedforward controller dynamics	$d(s)$	disturbance input

Servo System,
Fig. 3 Functional block diagram of a servo-control system

the loop transfer function $L(s) = G(s)K(s)$ assuming large values in this frequency range (i.e., $|L(j\omega)| \gg 1$). As a trade-off, this causes the noise transfer function to also become $G_{\text{noise}}(j\omega) \approx 1$. This means that in the frequency range that we achieve successful tracking and disturbance rejection, sensor noise also directly disturbs our output. Hence, correct sensor selection with sufficient precision and accuracy is crucial in servo system design, in addition to selecting the actuator(s) to be sufficiently powerful in order to deliver the required actuation effort.

While one of the most critical factors that determine the performance of a servo system is the closed-loop bandwidth (Skogestad and Postlethwaite 2005) (i.e., active frequency range, within which disturbance inputs are successfully rejected and reference inputs are accurately followed), there are also important limitations that determine the achievable bandwidth in a servo system, arising typically from the dynamics of the plant and controller such as unstable zeros and delays, unmodeled dynamics, vibration modes, actuator saturations, and measurement noise (Franklin et al. 2005; Pritschow 1996). Therefore, in designing servo controllers, it is crucial to guarantee that the final system will also be stable and retain its stability and performance in the presence of the aforementioned factors. To this end, several methods are being used for this purpose, such as the Routh-Hurwitz and Nyquist stability criteria (Franklin et al. 2005; Ogata 1997; Skogestad and Postlethwaite 2005), Lyapunov's Stability Theorem (Pritschow 1996), and Structured Singular Value (μ-) Analysis (Skogestad and Postlethwaite 2005).

While traditional servo design for machine tool and robotic applications has focused primarily on controlling the rigid body dynamics, the requirements to achieve greater accuracy during high travel speeds and accelerations have pushed the research community to develop a new generation of servo designs that actively compensate for the structural flexibilities and vibration modes of drive systems and machines (Altintas et al. 2011). In return, this allows significantly greater positioning and disturbance rejection bandwidths to be realized. One key issue then becomes maintaining the stability and performance of such designs in the presence of variations in the machine's or drive's structural (i.e., vibratory) parameters (Pritschow 1996), which is still an open research topic.

Cross-References

▶ Computer Numerical Control
▶ Control

References

Altintas Y (2000) Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design. Cambridge University Press, Cambridge
Altintas Y, Verl A, Brecher C, Uriarte L, Pritschow G (2011) Machine tool feed drives. Ann CIRP 60(2):779–796
Franklin GF, Powell JD, Emami-Naeini A (2005) Feedback control of dynamic systems, 5th edn. Prentice Hall, Upper Saddle River
Ogata K (1997) Modern control engineering, 3rd edn. Prentice Hall, Upper Saddle River
Pritschow G (1996) On the influence of the velocity gain factor on the path deviation. Ann CIRP 45(1):367–371
Skogestad S, Postlethwaite I (2005) Multivariable feedback control: analysis and design, 2nd edn. Wiley, New York

Setting Up

► Assembly

SG Iron

► Machining of Spheroidal Ductile Iron

Shape Deviation

► Form Error

Shape Distortion

► Springback

Shape Error

► Form Error

Shaping

► Gear Cutting

Shave Grinding

► Honing

Shear Cutting

Wolfram Volk and Jens Stahl
Institute of Metal Forming and Casting, Technische Universität München, Garching, Germany

Synonyms

Blanking; Cutting; Die cutting; Piercing; Punching; Shearing

Definition

Shear cutting is the separation of workpieces by two blades moving in opposite directions past each other (DIN 8588 2013).

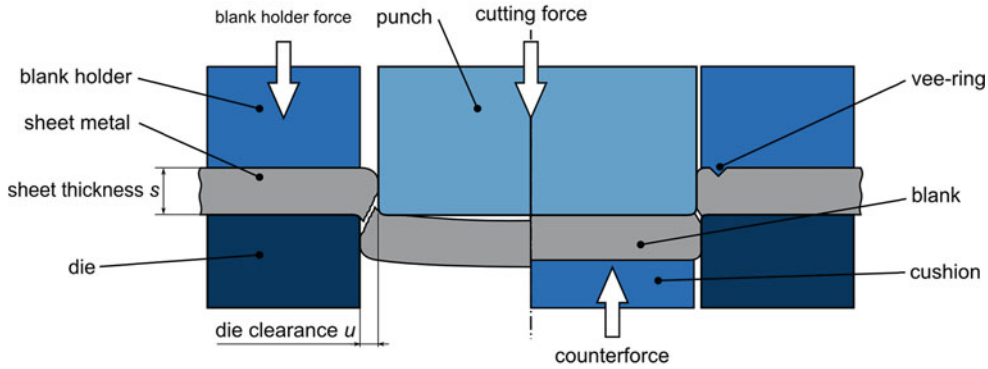
Theory and Application

Basics

The shear cutting processes are part of the separation manufacturing processes, standardized in (DIN 8580 2003) and more precisely characterized in (DIN 8588 2013) as the mechanical separation of workpieces without forming shapeless matter. Here, shapeless matter refers to chips or grinding dust, for example. Some parts manufactured by shear cutting are displayed in Fig. 1.

Figure 2 shows a typical tool with sheet metal. A shear cutting tool consists of at least a punch and a die, also referred to as die plate. The punch and die plate move in opposite directions past each other and are, therefore, responsible for the cutting. On account of this, they are also referred to as the tool's active elements. Both the punch

Shear Cutting, Fig. 1 Parts manufactured by shear cutting (Courtesy of Feintool Technologie AG, Lyss (Switzerland))

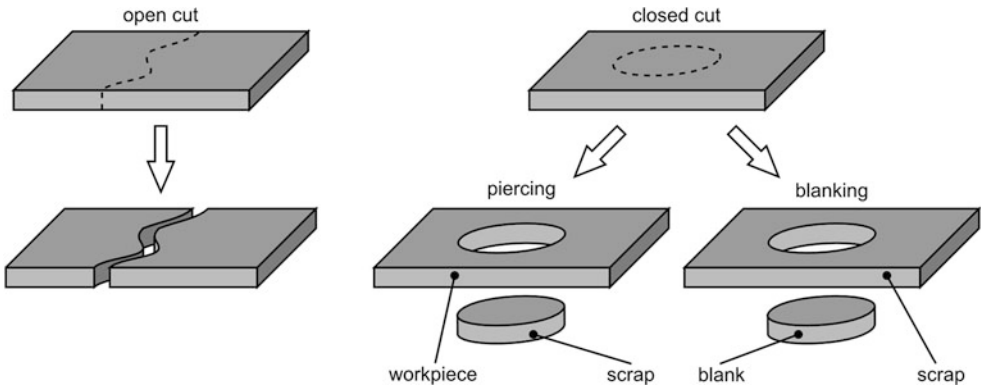


Shear Cutting, Fig. 2 Basic terms of the shear cutting process (*left*) and the fineblanking process (*right*) according to (Schmidt et al. 2007)

and die need to possess cutting edges. These are usually sharp or manufactured with a small radius of approximately 15 µm. Bigger radii or a chamfer may be used instead of the radius in special cases. As illustrated in Fig. 2, the punch's lower surface and the die's upper surface are both flat and parallel. Therefore, the whole cutting edge operates from the beginning of the cutting operation. However, this is not a necessity. The geometry of the punch and die plate does not have to be identical: a sloped punch may be used together with a flat die plate, for example.

The die clearance is the distance between punch and die plate, measured perpendicular to the area in which the cutting edge moves during the cutting process (DIN 8588 2013). It is common to specify the die clearance as a percentage of the sheet metal thickness. Usually, the die clearance ranges between 5 % and 12 % of the sheet metal thickness, depending on the workpiece material and the desired cut surface qualities, for example.

A blank holder, in the case of fineblanking with an additional vee-ring (see Fig. 2 on the right), is used to improve the quality of the manufactured parts. It fixes the sheet metal by pressing it against a stationary part before the cutting operation. Furthermore, it reduces warping of the sheet metal during the cutting operation by applying pressure on the workpiece, created by the blank holder force. Additionally, the blank holder is used to strip the sheet metal strip off the punch and may be used to guide the punch and act as a positioning unit.



Shear Cutting, Fig. 3 Open (*left*) and closed cut (*right*) together with piercing and blanking according to (DIN 8588 2013)

To reduce the bending of the workpiece under the punch, a cushion can be used (Smith 1994; Schuler GmbH 1998; Schmidt et al. 2007; Hörmann 2008; Altan and Tekkaya 2014a; b; Demmel 2014).

Subgroups

Subgroups of the shear cutting process can be distinguished according to several criteria:

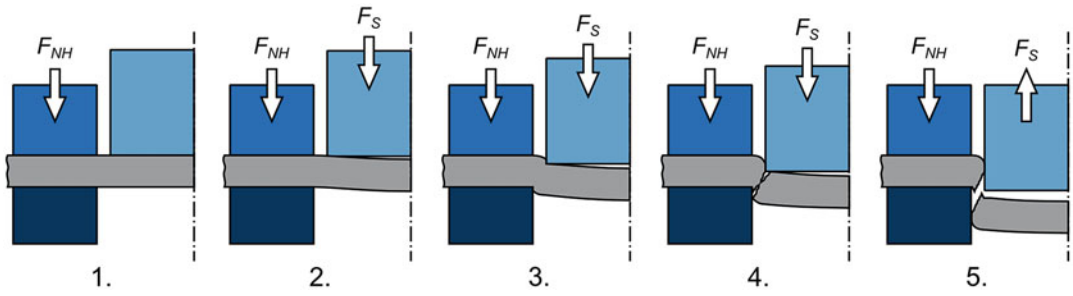
Open cut and closed cut	If the contour of the cut crosses the rim of the workpiece, it is an open cut; otherwise, it is considered a closed cut (see Fig. 3)
Piercing and blanking	The first describes a process where the cut-out part is scrap, whereas, with blanking, this piece is used. In this case, the punched out part is called a blank or a slug, and the remains are referred to as sheet metal strip (see Fig. 3) (DeGarmo et al. 2011)
Blanking and fineblanking	Fineblanking produces parts with smooth cut surfaces, usually as a result of using a vee-ring on the blank holder and a cushion (see the section on precision cutting processes). Conventional blanking on the other hand is characterized by a bigger die clearance, a blank holder without a vee-ring, and the absence of a cushion. This results in a smaller proportion of clean-cut compared to fineblanking
Continuity of the process	Refers to whether one or more strokes are needed for the cutting process or if it is continuous

Several other subgroups are listed in (DIN 9870-2 (1972)), and their classifications depend on the cutting line or the function of the cut-out or cut-off part.

Furthermore, a distinction is made between whether the full length of the cutting edges operates from the beginning of the process, or if the blades cross each other and one gradually cuts through the workpiece. The maximal force necessary for the cut is significantly lower for the case of the crossing blades, due to the fact that the active elements cut the workpiece gradually along the cutting line and not the whole contour at once. The sloping shape of an active element, necessary for the blades to cross, causes forces perpendicular to the punch's movement direction, which may cause the tool to distort, if not properly designed. This effect may also cause deformations on either the sheet metal strip, if the die plate is sloped, or the cut-out part, if the punch is sloped (Rao 1993; DIN 9870–2 1994; Schmidt et al. 2007; Demmel 2014).

Phases of the Shear Cutting Process

According to (Hoffmann et al. 2012), the shearing process can be divided into five phases. These are illustrated in Fig. 4. In this section, only the one-stroked closed cut, which produces a circular blank with a blank holder, is focused on. Furthermore, the full length of the blade edges operates from the beginning.



Shear Cutting, Fig. 4 Phases of the shear cutting process according to (Hoffmann et al. 2012)

Phase 1 – The blank holder fixes the sheet metal: After inserting the sheet metal, the blank holder touches the sheet metal and clamps down on it with the blank holder force F_{NH}. Simultaneously, the punch moves toward the sheet metal with a defined velocity.

Phase 2 – Elastic deformation of the sheet metal: After having established a firm contact with the sheet metal, the punch deforms the sheet metal elastically. Mainly due to die clearance, a bending torque is induced, which causes the workpiece to warp under the punch. The blank holder limits this effect on the rim of the sheet metal. This bending torque also reduces the contact area between the punch and sheet metal to a ring-shaped zone. Additionally, elastic energy is stored in both the tool and press (Hoffmann et al. 2012; Doege and Behrens 2010; Demmel 2014).

Phase 3 – Plastic deformation of the sheet metal: Once the yield point of the material is reached, plastic deformation occurs. Stress is highest at the blade edges, and material flows downward with the punch movement. The plastic deformation is located in an area near to the blade edges, which is referred to as shearing zone. Due to the material flow, a roll over is formed, in this case on the upper side of the sheet metal strip and on the lower side of the blank. The progressive penetration of the sheet metal by the punch results in the clean-cut zone, which is characterized by a very smooth surface. Furthermore, the previously mentioned bending torque leads to a permanent bend in the workpiece (Lange and Liewald 1990).

Phase 4 – Formation of cracks and separation of the sheet metal: Once the formability of the material is exhausted, cracks appear. Depending on different parameters, this can start from either the punch, the die, or both at the same time. If both blade edges are equally sharp, the crack usually initiates at the die plate, due to the combined stress from bending and plastic deformation. The cracks begin to grow with ongoing deformation, eventually merge, and thus separate the workpiece. If they do not coalesce, another scenario is possible. In this case, the bridge that still connects the material can be sheared off again by the previously mentioned mechanism, resulting in separation of the pieces. This results in one or more additional clean-cut zones. If these are separated from the roll over by a fracture zone, they are referred to as secondary clean-cut zones or double shear (Groover 2002; Rapien 2010; Hoffmann et al. 2012; Demmel 2014).

Phase 5 – Springback and punch movement to the initial position: The stored elastic energy is released after the separation of the workpiece, which causes the whole system to vibrate. The springback may also result in a change of the workpieces' shape and an interference fit between the punch and the sheet metal strip. After reaching the bottom dead center, the punch moves back to the initial position. The relative movement between the punch as well as the die and the sheet metal is responsible for wear on the respective active elements (Doege and Behrens 2010). The blank holder strips the sheet metal strip off the punch by preventing the sheet metal strip to move upward (Hoffmann et al. 2012; Demmel 2014).

Mechanics of the Shear Cutting Process

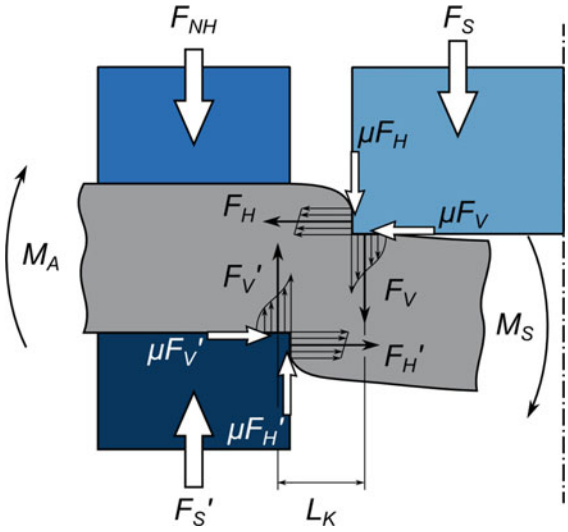
The forces and torques occurring during the shear cutting process are shown in Fig. 5.

These forces, necessary to overcome friction and the deformation resistance of the material, have to be provided by the press. Therefore, the occurring forces are of utmost importance for the tool's design.

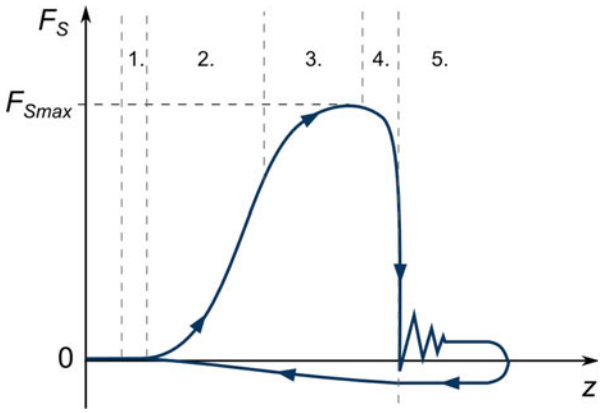
The cutting force F_S consists of, as illustrated in Fig. 5, the contact force F_V perpendicular to the bottom of the punch and the frictional force μF_H on its lateral surface.

The cutting force – punch travel curve, presented in Fig. 6, enables to evaluation and monitoring of the cutting process, due to the fact that the geometry of the cut surface can be retraced with this curve.

All of the abovementioned phases can be found in this curve. The almost linear, very steep increase at the beginning corresponds with the elastic deformation of tool, press, and workpiece. After reaching the sheet metals yield limit, a still increasing curve can be observed. Two competing mechanism, strain hardening and the reduction of the sheets thickness by the plastic deformation, are mainly responsible for the curves shape in that region. At first the increase in force necessary to compensate strain hardening is bigger than the decrease of the cross-sectional area, until a maximum of the cutting force is reached. Afterward,



Shear Cutting, Fig. 5 Force components during the shear cutting process according to (Hoffmann et al. 2012)



Shear Cutting, Fig. 6 A typical cutting force – punch travel graph with corresponding phases

the reduction of the cross-sectional area dominates, and, therefore, a decrease of the cutting force can be observed. After the formability limit is reached, cracks occur and the workpiece is separated. This results in a sudden decrease in the cutting force. The stored elastic energy is released and causes the whole system to vibrate, which results in the oscillating part of the curve after the abrupt decrease. Due to friction between the sheet metal strip and the punch, a force is necessary even after the separation until the sheet metal strip is stripped of by the blank holder (Rao 1993; Schmidt et al. 2007; Hoffmann et al. 2012; Demmel 2014).

According to (Klocke 2013), the cutting force is affected by the following factors:

- Shear strength τ_B of the workpiece material
- Sheet thickness s
- Die clearance u
- Length of the cutting line l_S
- Geometry of the cutting line
- Tool wear
- Surface condition of the tools
- Lubrication

Punch and die radii influence the occurring force as well (Schmidt et al. 2007).

An estimation for the maximum of the cutting force F_{Smax} can be calculated according to (Hoffmann et al. 2012):

$$F_{Smax} = s\, l_S\, k_S$$

with the sheet metal thickness s, the length of the cutting line l_S, and the cutting resistance k_S.

The cutting resistance k_S can be approximated by the ultimate tensile strength of the sheet metal material R_m and then multiplied by the factor c_S, which varies between 0.6 for brittle and 0.8 for ductile materials (Doege and Behrens 2010):

$$k_S = c_S\, R_m$$

The work required for the cutting operation, defined by the integral of the cutting force along the punch travel of one cycle, is another important parameter for the selection of a suitable press (Klocke 2013):

$$W_S = \int_0^{z_g} F_S(z)\, dz$$

with the work required for the cutting operation W_S, the cutting force F_S, the punch travel z, and the total punch travel z_g.

The equation above can be approximated by

$$W_S = F_{Smax}\, s\, c$$

with the work required for the cutting operation W_S, the maximum of the cutting force F_{Smax}, the sheet metal thickness s, and a coefficient c.

The coefficient c ranges between 0.3 and 0.7 and incorporates workpiece material properties as well as process parameters. The lower end of this range should be chosen for brittle sheet metal materials together with a big die clearance and a high sheet thickness. Consequently, the upper range allows to approximate the work for ductile sheet metal materials with a small die clearance and a thin sheet (Schuler GmbH 1998; Klocke 2013).

The horizontal force component on the punch, consisting of F_H and μF_V, is necessary for a complete understanding of the shear cutting process.

The cutting force is not evenly distributed on the contact surface between active element and sheet metal, but concentrated on a small area near the blades.

The resultants of the area loads in vertical direction, F_V and F_V', are located within a distance L_K from each other. This results in the bending torques M_A and M_S which are mostly determined by the die clearance, the sheet metal thickness, and the punch diameter (Hörmann 2008). The torque M_A causes the sheet metal to warp and lifts it of the die surface. This results in stresses on the lateral area of punch and die, which can be combined in the respective forces F_H and F_H'. Additional frictional forces in horizontal and vertical direction can be found on the contact surface between the active elements and the sheet metal. These frictional forces depend on contact pressure and several other conditions, lubrication and surface conditions, for example (Klocke 2013; Hoffmann et al. 2012; Altan and Tekkaya 2014b; Demmel 2014).

Especially in an open cut, the horizontal force components should be kept in mind. They can cause undesired displacements between punch and die plate which can lead to poor size accuracy. When cutting a circular hole or any other geometry with a rotationally symmetric cutting line, this effect cannot be observed due to the fact that the horizontal force components cancel each other (Hoffmann et al. 2012).

Cut Surface Characteristic Values

Characteristic values of the cut surface can be seen in Fig. 7. They are selection criteria for obtaining a workpiece with the desired quality. Here, the function of the part defines which cut surface characteristics are desired.

Due to the plastic deformation during the cutting process, the roll over and clean-cut zone are formed. This deformation also causes strain hardening in the shearing zone and, therefore, an elevated hardness in this region. The clean-cut zone is also referred to as burnish or shear zone.

A fracture zone is formed after the formability of the sheet metal material is exhausted. The shape of the burr is determined by several parameters, for example, the shape of the blade edges and the locus of crack initiation.

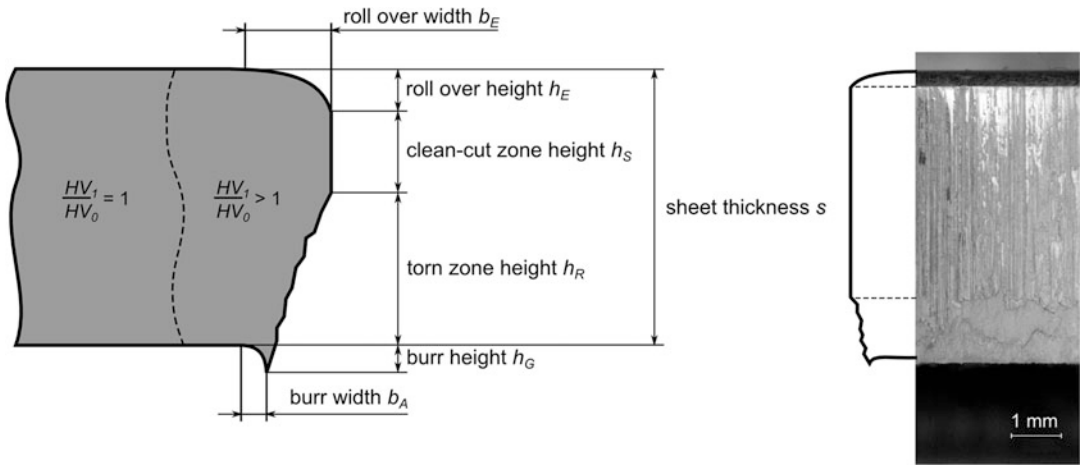
The vertical dimensions are usually evaluated as a percentage of the sheet metal thickness. This allows easier comparison between different materials (Groover 2002; Schmidt et al. 2007; Altan and Tekkaya 2014b; Demmel 2014).

Precision Cutting Processes

For some applications, the surface characteristics achieved by conventional shear cutting are not sufficient. The outer surface of a gear, for example, needs small tolerances and a smooth surface to be able to transmit the desired torque. This can be achieved by finishing the surface by grinding, milling, or additional shear cutting operations. Another possibility is to adjust the parameters and the set-up of the shear cutting process to achieve the desired workpiece quality in a single cutting operation.

One possibility is to cut the workpiece in a first step slightly bigger than the desired final contour. This additional material is cut off in a second shear cutting operation. Due to the small dimensions of the additional material, a higher percentage of clean-cut zone can be achieved. The combination of those two shear cutting operations is called sheaving or burnishing.

A burr-free cut surface with a higher percentage of clean-cut zone compared to the conventional shear cutting can be achieved by counterpunching. A tool with two sets of active elements, one on each side of the sheet metal, is required for this process. The first set is used to cut the sheet metal from one side and stopping the



Shear Cutting, Fig. 7 Cut surface characteristic values according to (VDI 2906 Blatt 2 1994) (*left*) and photography of a cut surface (*right*)

operation slightly before the material is separated. Afterward, the second set of active elements is used to cut through the sheet metal from the opposite side, this time by fully separating the parts.

For some applications, the surface characteristics achieved by shear cutting with a die clearance smaller than 5 % can be sufficient.

Parts which were produced by fineblanking can show a clean cut over the whole thickness of the sheet metal. The high quality of the surface is achieved by altering the stress condition in the shearing zone. An additional vee-ring on the blank holder penetrates the sheet metal and hinders the material flow during the cutting process which reduces the roll over. The pressure induced by this penetration delays the crack formation. Furthermore, a small die clearance of around 0.5 % and a cushion are used. The counterforce on the cushion opposes the punch force to reduce the bending of the sheet metal (Rao 1993; VDI 2906 Blatt 3 1994; VDI 2906 Blatt 5 1994; Schuler GmbH 1998; Szumera 2003; Schmidt et al. 2007; Tschaetsch 2007; Hörmann 2008).

Presses for Shear Cutting

Both mechanical and hydraulic presses are used for shear cutting. Gap-frame presses are usually used for the production of parts with low-quality requirements and press forces smaller than 2500 kN. For press forces higher than 4000 kN, straight side presses are used exclusively (Lange and Liewald 1990; Smith 1994; Schuler GmbH 1998).

To decide whether a mechanical or a hydraulic press best suits the production of the desired part, several factors should be taken into account:

- Number of production steps
- Tool design
- Sheet metal thickness
- Desired component quality
- Efficiency (Lange and Liewald 1990)

To improve the production rate, high-speed presses are used, which allow 2000 strokes per minute (Schuler GmbH 1998).

Fineblanking requires a blanking force, a force for the vee-ring, and a counterforce on the cushion. Therefore, triple action presses are used. To achieve the desired cut surface quality, the die clearance is not allowed to change under load: A stiff construction of tool and frame is needed to avoid undesired displacements. Furthermore, high precision of, for example, slide gibs and parallelism of the die clamping surfaces are necessary (Schmidt et al. 2007; Schuler GmbH 1998).

Cross-References

- ▶ Billet Shearing
- ▶ Forming
- ▶ Stamping

References

Altan T, Tekkaya AE (2014a) Sheet metal forming – fundamentals. ASM International, Russel Township

Altan T, Tekkaya AE (2014b) Sheet metal forming – processes and applications. ASM International, Russel Township

DeGarmo EP, Black JT, Kohser AR (2011) Materials and processes in manufacturing. Wiley, Hoboken

Demmel PM (2014) In-situ temperature measurement in shear cutting. Dissertation, Technische Universität München

DIN 8580:2003-09 (2003) Manufacturing processes – terms and definitions, divisions. DIN Deutsches Institut für Normung e. V., Beuth Verlag, Berlin

DIN 8588:2013-08 (2013) Manufacturing processes severing – classification, subdivision, terms and definitions. DIN Deutsches Institut für Normung e. V., Beuth Verlag, Berlin

DIN 9870-2:1972-10 (1972) Terms for stamping practice; production processes and tools for severing. DIN Deutsches Institut für Normung e. V., Beuth Verlag, Berlin

Doege E, Behrens B-A (2010) Handbuch Umformtechnik. Springer, Heidelberg

Groover MP (2002) Fundamentals of modern manufacturing – materials, processes, and systems. Wiley, Hoboken

Hoffmann H, Neugebauer R, Spur G (2012) Handbuch Umformtechnik. Carl Hanser Verlag, München

Hörmann F (2008) Influence of the process parameters on single-stage near-net-shape-blanking operations. Dissertation, Technische Universität München

Klocke F (2013) Manufacturing processes 4 – forming. Springer, Berlin

Lange K, Liewald M (1990) Umformtechnik – Handbuch für Industrie und Wissenschaft – Band 3: Blechbearbeitung. Springer, Berlin

Rao PN (1993) Manufacturing technology – foundry, forming and welding. Tata McGraw-Hill, New Delhi

Rapien BL (2010) Fundamentals of press brake tooling. Hanser Verlag, München
Schmidt R-A, Birzer F, Höfel P, Hellmann M, Reh B, Rademacher P, Hoffmann H (2007) Cold forming and fineblanking. Carl Hanser Verlag, München/Wien
Schuler GmbH (1998) Metal forming handbook. Springer, Heidelberg
Smith DA (1994) Fundamentals of pressworking. SME, Dearborn
Szumera J (2003) The metal stamping process: your product from concept to customer. Industrial Press Inc., New York
Tschaetsch H (2007) Metal forming practise: processes – machines – tools. Springer, Heidelberg
VDI 2906 Blatt 2:1994-05 (1994) Quality of cut faces of (sheet) metal parts after cutting, blanking, trimming or piercing; shearing, form of sheared edge and characteristic values. VDI Verein Deutscher Ingenieure, Düsseldorf
VDI 2906 Blatt 3:1994-05 (1994) Quality of cut faces of (sheet) metal parts after cutting, blanking, trimming or piercing; sheaving. VDI Verein Deutscher Ingenieure, Düsseldorf
VDI 2906 Blatt 5:1994-05 (1994) Quality of cut faces of (sheet) metal parts after cutting, blanking, trimming or piercing; fine blanking. VDI Verein Deutscher Ingenieure, Düsseldorf

Shear Forming

Omer Music
Mechanical Engineering Department, TED University, Ankara, Turkey

Synonyms

Hydrodynamic spinning; Power spinning; Shear spinning; Shear/flow turning

Definition

A metal-forming process is used for production of hollow, rotationally symmetric sheet metal components. To produce a component with a given shape and thickness distribution, the sheet is clamped to a rotating rigid mandrel and formed by a roller tracing the shape of the mandrel at a fixed distance.

Theory and Application

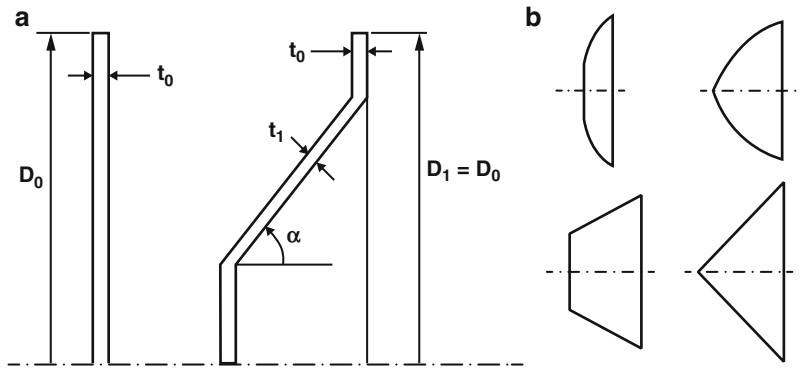
Introduction

Shear forming belongs to a group of processes used for production of hollow, rotationally symmetric parts. The term spinning is used to refer to the three processes in this group: metal (conventional) spinning, shear forming (shear spinning), and flow forming (tube spinning). The main difference between the three processes is the wall thickness of the formed part. While metal spinning preserves sheet thickness, shear-forming and flow-forming thickness is deliberately reduced to obtain a given part shape and thickness distribution.

Spinning processes have evolved from the craft of pottery using a potter's wheel, which can be traced back to ancient Egypt. The process has developed over the years; however, at the beginning of the twentieth century, it was still considered a craft process. In the middle of the past century, demand by industry for shorter production times and tighter dimensional tolerances led to an increase in power of the spinning machines and automation of the process, leading to an improvement in product quality and subsequently to the evolution of shear forming.

Components produced by shear forming are mainly parts for the automotive and aerospace industries, along with art objects, musical instruments, and kitchenware. Typical examples are jet engine and turbine components, dishes for radar antennas, and domestic utensils. The process is capable of forming components of diameters ranging from 3 mm to 10 m and thicknesses of 0.4–25 mm (Brown 1998). Figure 1 shows the main features of the process and examples of obtainable geometries.

The main alternative to shear forming is press forming. However, shear forming has three main advantages when compared to press forming; as the material deformation is localized under the roller, forming forces are low; simple and non-dedicated tooling provides flexibility; lastly, formed components generally have improved mechanical properties and a high-quality surface finish.



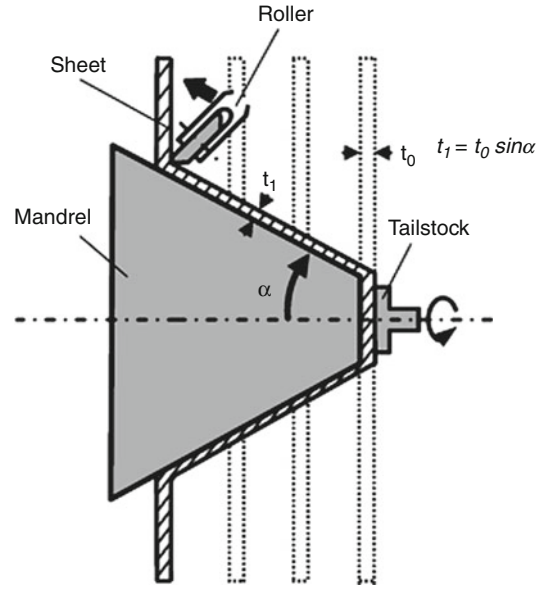
Shear Forming, Fig. 1 Shear forming: main feature, blank (*left*) and product (*right*), (**a**) and examples of obtainable geometries (**b**) (D_0 and D_1, diameter before and after forming; t_0 and t_1, thickness before and after forming; α, product wall angle)

Process Description

In contrast to metal spinning, in shear forming, the wall thickness is deliberately reduced while keeping the sheet diameter equal to the diameter of the original blank. The sheet is formed by the roller which traces the shape of the rigid mandrel, reducing the thickness from the initial thickness t_0 to a thickness t_1. The sine law (Fig. 2) relates the final thickness t_1 to the angle between the wall of the component and the axis of rotation (α). The mechanism of deformation in shear forming is predominantly shear along the axis of rotation.

The blank is typically formed in a single roller pass. Blank shape can be either a flat sheet or a preform produced by other means. Generally, a single roller is used to form the sheet; however, in case of high-strength and thick-walled components, a multiple roller configuration may be used.

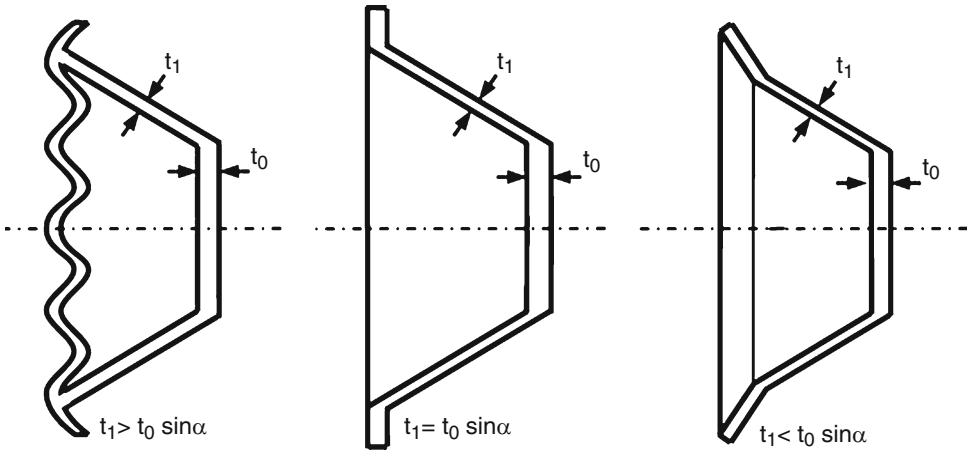
The sheet in conventional shear forming is clamped in the central section, allowing for production of parts with a flat central section only. To form the sheet over the mandrel end, a slightly different configuration is employed; instead of clamping the central section, the sheet is supported around the edges, with the support moving together with the roller.



Shear Forming, Fig. 2 Shear forming

Defects in Shear Forming

Defects in shear forming may occur in two cases. In the first case, defects occur when forming components with low wall angles. As the wall angle decreases, the sheet becomes thinner, as imposed by the sine law, eventually leading to severe thinning and sheet fracture. The minimum wall angle for a given material depends mainly on its ductility and the clearance between the roller and the mandrel. Reducing the clearance between the roller and the mandrel applies an additional compressive stress on the sheet and allows the sheet to be formed to lower wall angles. However, reducing the roller-mandrel clearance requires a deviation from the sine law – the second case in which defects may occur. Ideally, the final sheet thickness t_1 should be set to the value obtained from the sine law. Any deviation from the sine law may cause defects in the flange. This is because in shear forming to sine law value, the internal stresses are confined to the immediate area around the roller, leaving the flange stress free. Deviation from sine law causes the stresses to extend to the



Shear Forming, Fig. 3 Deviation from sine law in shear forming

flange and can lead to bending or wrinkling of the flange (Fig. 3).

Recent Developments

In spite of its age, the original shear-forming configuration had seen little or no change to the original configuration until recently. In the last two decades, the industrial trend toward low-volume, small batch production has driven a range of innovations in shear forming aimed toward making the process more flexible, producing a wider range of shapes, and forming challenging materials.

Flexible Shear Forming

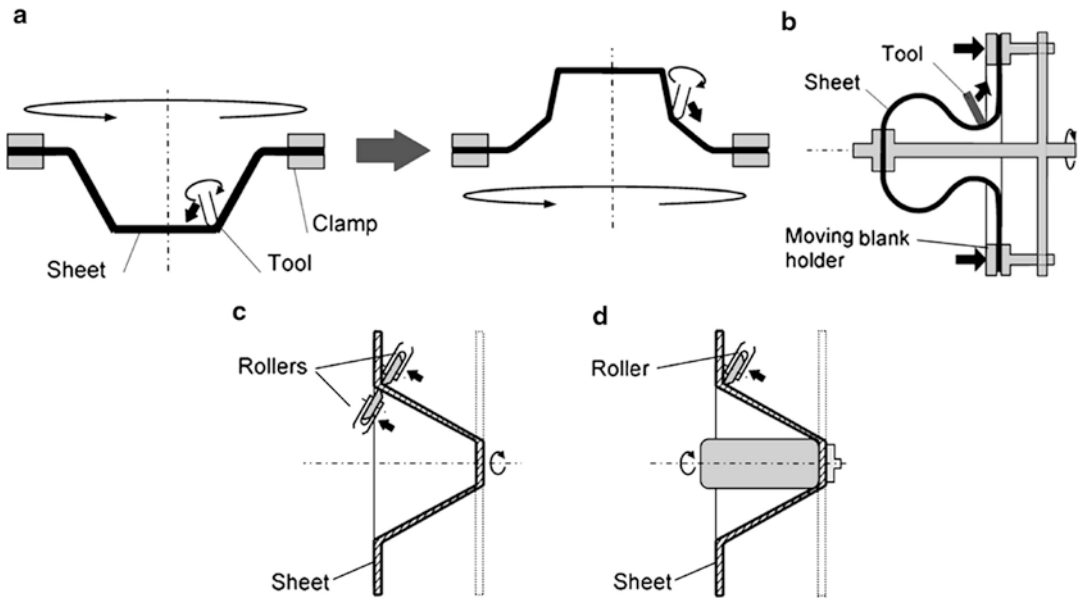
Conventional shear forming is not flexible, since a specific mandrel is required for each product. To overcome this limitation, four different configurations of shear forming without a mandrel (mandrel-free shear forming) have been proposed.

In the first configuration, rotating sheet is supported around the edges, while a tool gradually forms it into final shape (Kitazawa et al. 1994), shown in Fig. 4a (left). To allow forming of sharp corners and complex shapes, the sheet can be preformed and then reversed (Fig. 4a – right). The main feature of this configuration is its ability to form vertical walls, which is not possible in the conventional shear-forming configuration. In the second configuration, a moving holder supports the sheet around its edges and moves along the axis of rotation together with the roller (Matsubara 2001). This configuration allows forming of extreme, reentrant shapes from a flat sheet (Fig. 4b). In the third approach, the mandrel is replaced by a roller on the inner side of the sheet, located opposite the conventional roller (Shima et al. 1997), shown in Fig. 4c. This configuration creates a highly localized, controllable deformation in the sheet. The fourth configuration shown in Fig. 4d is similar to the first one; however, in this case, the sheet is not clamped around the edges but in the center, leaving the edges free (Kawai et al. 2001).

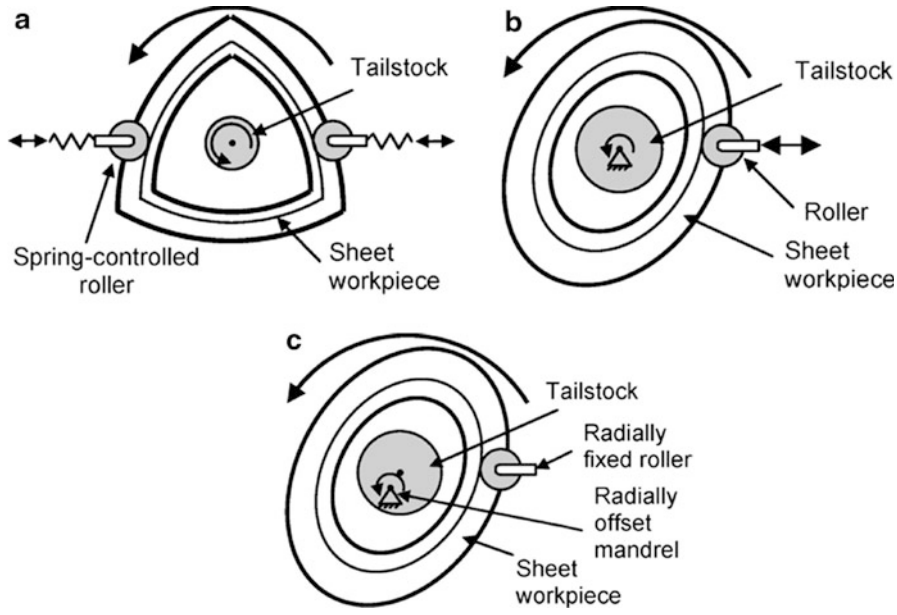
Asymmetric Shear Forming

Conventional shear forming is limited to production of rotationally symmetric components. To produce asymmetric components and therefore increase the range of shapes that can be produced by shear forming, four asymmetric configurations have been proposed.

The first configuration (Fig. 5a) involves shear forming with a pair of spring-controlled rollers which trace the shape of the asymmetric mandrel as it rotates (Awiszus and Meyer 2005). In the second approach (Fig. 5b), a single roller oscillates radially, tracing the shape of the mandrel as it rotates (Amano and Tamura 1984). Radial roller oscillation is achieved through a copying mechanism, linked to the main motor to synchronize the radial motion with mandrel rotation. Similar relative motion between the mandrel and the roller can be achieved

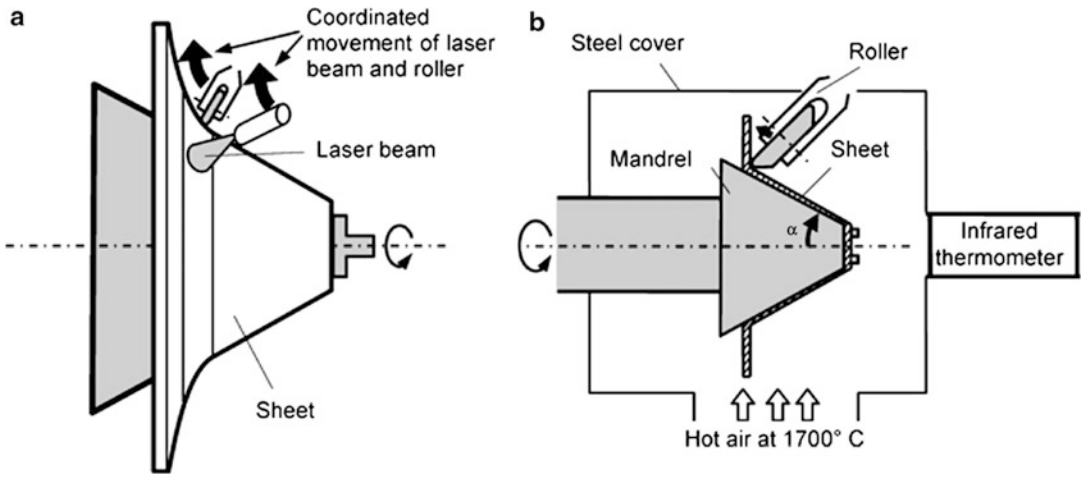


Shear Forming, Fig. 4 Flexible shear-forming configurations. (**a**) Shear forming of preformed shells. (Adapted from Kitazawa et al. 1994). (**b**) Shear forming with a moving blank holder. (Adapted from Matsubara 2001). (**c**) Shear forming with two rollers. (Adapted from Shima et al. 1997). (**d**) Shear forming with a simple cylindrical mandrel. (Adapted from Kawai et al. 2001)



Shear Forming, Fig. 5 Asymmetric shear forming. (**a**) Spring-controlled rollers. (**b**) Radially offset roller. (**c**) Radially offset mandrel

by oscillating the mandrel instead of the roller and keeping the roller fixed radially, shown in Fig. 5c (Gao et al. 1999). The last approach is based on a hybrid force/position control system which allows the roller to track the mandrel shape automatically (Arai 2005).



Shear Forming, Fig. 6 Hot shear forming with (**a**) laser, shear forming with laser. (Adapted from Klocke and Wehrmeister 2003), and (**b**) hot air, shear forming with hot air. (Adapted from Mori et al. 2009)

Hot Shear Forming

Shear forming is typically performed cold, but in case of high-strength materials and thick components, heating is applied to reduce the roller forces. In industry, manual heating is used widely; sheet is heated by an oxyacetylene flame as it is formed. However, this improvised approach has its disadvantages; sheet temperature cannot be controlled accurately; strength-reducing diffusion can occur along with reactions with atmosphere; and lastly, the workpiece needs to be removed, annealed, and replaced several times during the process. Two hot shear-forming configurations have been proposed to overcome these challenges.

The first approach (Fig. 6a) uses a laser to heat the sheet, just before it contacts the roller, allowing localized and controllable heating of the sheet (Klocke and Wehrmeister 2003). This approach has been applied in conventional spinning; however, it is also applicable in shear forming. In the second approach, shown in Fig. 6b, the shear-forming setup is enclosed in a chamber, and hot air is used to heat the sheet (Mori et al. 2009).

Cross-References

- ▶ Flow Forming
- ▶ Metal Spinning

References

Amano T, Tamura K (1984) The study of an elliptical cone spinning by the trial equipment. In: Proceedings of the third international conference on rotary metalworking processes, Kyoto, 8–10 Sept, pp 213–224

Arai H (2005) Robotic metal spinning – forming non-axisymmetric products using force control. In: Proceedings of the 2005 I.E. international conference on robotics and automation (ICRA 2005), Barcelona, 18–22 Apr 2005, pp 2702–2707. https://www2.lirmm.fr/lirmm/interne/BIBLI/CDROM/ROB/2005/ICRA%202005/papers/a361.pdf. Accessed 30 Apr 2013, and in the J Robot Soc Jpn 24(1):140–145 (in Japanese)

Awiszus B, Meyer F (2005) Metal spinning of non-circular hollow parts. In: 8th international conference on technology of plasticity, Verona, 9–13 Oct 2005, pp 353–355

Brown J (1998) Advanced machining technology handbook. McGraw-Hill, New York

Gao X-C, Kang D-C, Meng X-F, Wu H-J (1999) Experimental research on a new technology – ellipse spinning. J Mater Process Technol 94(2–3):197–200

Kawai K, Yang L-N, Kudo H (2001) A flexible shear spinning of truncated conical shells with a general-purpose mandrel. J Mater Process Technol 113(1–3): 28–33

Kitazawa K, Wakabayashi A, Murata K, Seino J (1994) A CNC incremental sheet metal forming method for producing the shell components having sharp components. J JSTP 35(406):1348–1353. (in Japanese)

Klocke F, Wehrmeister T (2003) Laser-assisted metal spinning of advanced materials. In: Proceedings of the 4th LANE (Laser assisted net shape engineering) conference, Erlangen, 21–24 Sept 2003, pp 1183–1192

Matsubara S (2001) A computer numerically controlled dieless incremental forming of a sheet metal. Proc Inst Mech Eng B J Eng Manuf 215(7):959–966

Mori K, Ishiguro M, Isomura Y (2009) Hot shear spinning of cast aluminium alloy parts. J Mater Process Technol 209(7):3621–3627
Shima S, Kotera H, Murakami H (1997) Development of flexible spin-forming method. J Jpn Soc Technol Plast 38(440):814–818 (in Japanese)

Shear Spinning

▶ Flow Forming
▶ Shear Forming

Shear Stress Modeling

▶ Cutting Force Modeling

Shear/Flow Turning

▶ Flow Forming
▶ Shear Forming

Shearing

▶ Billet Shearing
▶ Shear Cutting

Sheet-Bulk Metal Forming

Marion Merklein
LFT, Institute of Manufacturing Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Synonyms

Plate forging

Definition

Sheet-bulk metal forming (SBMF) processes are defined as forming of sheets with an intended three-dimensional material flow as in bulk forming processes (Merklein et al. 2012). As semifinished product, sheets with an initial thickness of 1–5 mm are used and subjected to one or several conventional bulk forming operations. Typical applications of SBMF include the forming of local functional elements on blank parts or the intended and locally restricted alteration of the sheet thickness in order to produce highly functional integrated parts out of sheet metal (Fig. 1) (Mori 2012). Further complexity can be achieved by the combination of SBMF processes with traditional sheet forming operations like bending or deep drawing, which is generally possible for most applications (Merklein et al. 2012). Besides the fabrication of end products, SBMF can also be used to produce tailored blanks with adjusted properties for usage in subsequent manufacturing processes (Tan et al. 2008).

Theory and Application

Classification

According to DIN 8582 (2003), forming processes are categorized with respect to their stress states during the forming operation. This classification is not applicable to SBMF processes due to the fact that characteristics of multiple processes can merge in some cases and due to the missing consideration of the geometry of the semifinished part in question. Kudo (1980) proposed a different approach that enables a general classification of SBMF processes as a function of the semifinished parts in question, the resulting product geometry, and the strain state in relation to the main axis of the part. Standring (1999) proposed another, more detailed classification, which also uses the contact normal vector, as well as the ratio of the contact areas in the forming zone in comparison to the remaining contact area as criteria. However, since both approaches are kept very general, an exact assignment of process characteristics to the process class SBMF is not possible. In order to solve this challenge, Merklein et al. (2011) proposed a classification in respect to the tool motion as is displayed in Fig. 2. According to this classification, upsetting, ironing, forging, and coining can

Sheet-Bulk Metal Forming, Fig. 1 Examples of sheet-bulk metal forming parts with functional elements (Feintool 2015)

	Linear tool motion				Rotational tool motion		
	Upsetting	Ironing	Forging	Coining	Flow forming	Orbital forming	Boss forming
Change in thickness	↑↓	↓	↑↓	↑↓	↓	↑↓	↑↓
Combination with sheet forming known	✓	✓	✓	✓	✓	✗	✗
Forming force	High	Medium	High	High	Low	Medium	Medium

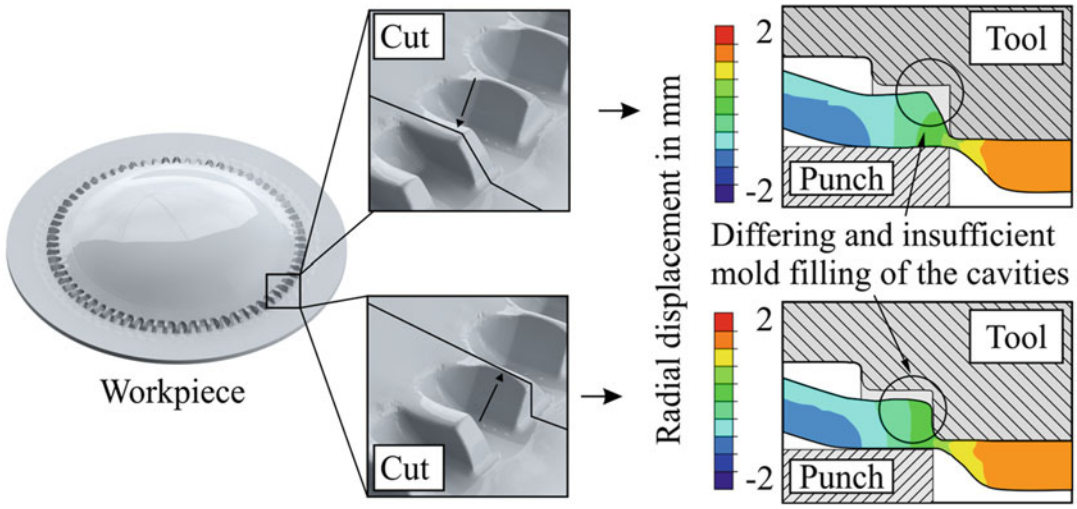
Sheet-Bulk Metal Forming, Fig. 2 Process classification and characteristics of common sheet-bulk metal forming operations (Merklein et al. 2011)

be assigned to the group of linear tool motion, whereas flow forming, orbital forming, and boss forming processes possess a rotational tool motion. As Fig. 2 depicts, with the exception of orbital forming and boss forming, combinations with conventional sheet forming processes are possible in all cases. Additionally, nearly all process variants allow the selective, local thickening and thinning of the sheet in use. Regarding typical forming forces, big differences between the individual process variants exist. The force demands are generally determined by the size of effective contact area between the tool and the forming zone. In principle, smaller contact areas result in less process forces. Hence, the lowest force demands can be found in incremental forming operations like flow forming or orbital forming.

Material Flow

Most applications of SBMF aim to produce parts that possess local functional elements on sheet metal with an element height in the dimension of the initial sheet thickness. Due to the complex three-dimensional material flow necessary for the forming of the singular elements, the stress and strain states during the forming operation are complex and three dimensional as well. In contrast to this, the remaining blank area shows comparatively low two-dimensional stress conditions. The high strains in the forming zone of the elements result in high and locally restricted strain hardening that is surrounded by large areas of low strain-hardened material.

Due to the strain gradient, the material flow tends to be directed away from the functional elements and therefore results in incomplete forming. Figure 3 demonstrates this circumstance by means of an exemplary SBMF forging process that enables the forming of two different element shapes on a circular blank. In order to allow proper forming of the functional elements despite this, the material flow for a particular SBMF process has to be considered in detail during the process design, and measures have to



Sheet-Bulk Metal Forming, Fig. 3 Insufficient mold filling during a SBMF forging process (Gröbel et al. 2015)

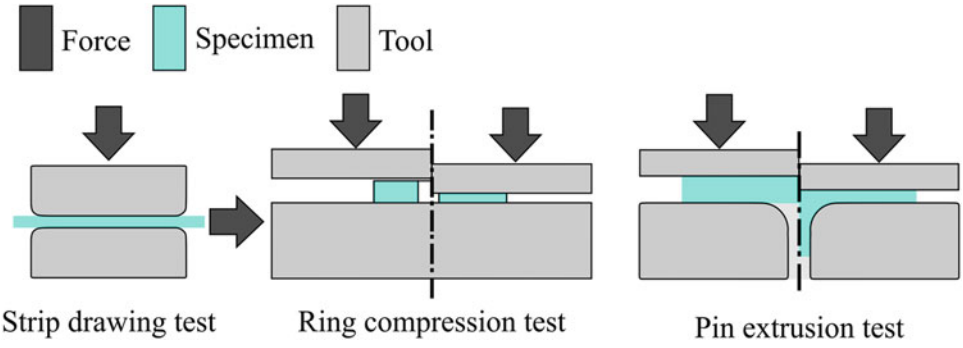
be taken to control the material flow as desired. Current research shows that this can be achieved either by means of geometrical flow restrictors (Koch and Merklein 2013) or by a local modification of the friction between the workpiece and the tool as demonstrated by Löffler et al. (2015).

Tribology

Tribological conditions are of major importance in SBMF processes, due to their influence on the process quality and accuracy (Merklein et al. 2012). Hence, detailed knowledge on the friction behavior is important to allow proper tool and process design by means of FE simulations. In order to determine the required friction factors and friction coefficients, laboratory friction tests are utilized. However, since SBMF processes possess both characteristics of conventional sheet and bulk forming operations, the tribological conditions during SBMF operations are of high complexity and cannot be represented by one friction test alone (Hetzner et al. 2012). Tribological conditions vary locally in dependence of the stress and strain conditions present at the contact surface in question. For this reason, the combined usage of the three friction tests shown in Fig. 4, strip drawing, ring compression, and pin extrusion, has been proposed in order to represent the dominant stress and strain states during SBMF processes. Both the ring compression test and the pin extrusion test have been adapted for this purpose by Vierzigmann et al. (2013) to meet the conditions of SBMF.

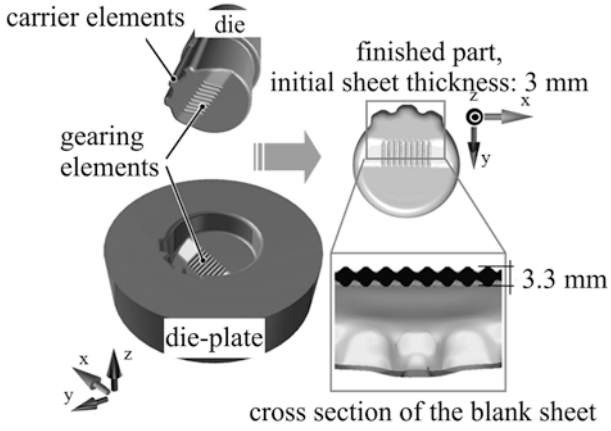
Tooling

Due to the complexity and locally varying stress and strain conditions during SBMF operations, the requirements on tool design and load capacity are demanding. Especially forging processes result in very high tool contact stresses as has been demonstrated for example by Nakano (2009). Furthermore, forming of asymmetrical parts leads to inhomogeneous tool loading and increased horizontal forces. An example for an asymmetrical product and the corresponding tool system is displayed in Fig. 5. Furthermore, since tool load in the forming zone of the elements can be as high as in conventional cold forging that show values up to 2750 MPa, tool systems for SBMF have to be reinforced in many cases in order to increase the load capacity and prolong tool life. For the same reason, the use of modern powder metallurgical high strength steels and innovate tool design, like non-circular reinforcements, are oftentimes necessary (Merklein et al. 2012).



Sheet-Bulk Metal Forming, Fig. 4 Different friction tests proposed for SBMF processes after Vierzigmann et al. (2012) and Vierzigmann et al. (2013)

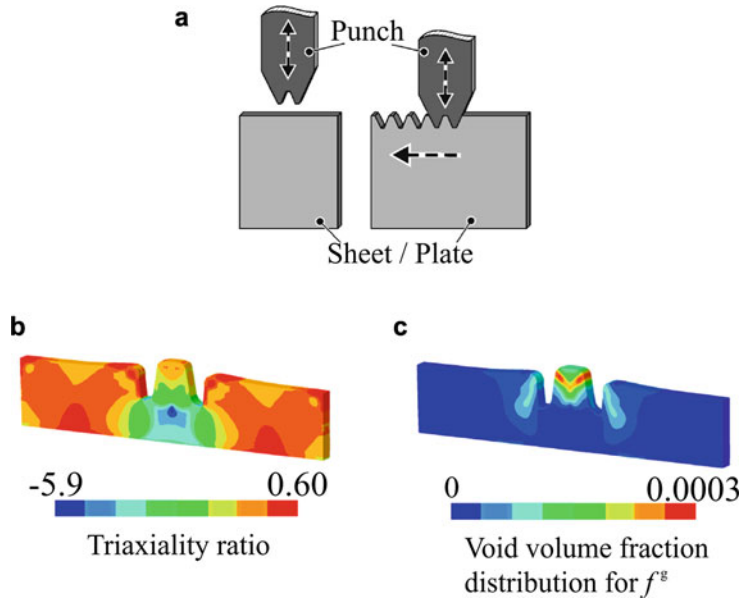
Sheet-Bulk Metal Forming, Fig. 5 Forming die and finished part with functional elements (Behrens et al. 2011)



Modeling

Since material flow and the resulting stress and strain conditions, as well as the tribological interaction between workpiece and tool, are complex and locally varying, process design by use of FE simulation is advisable. However, the modeling of SBMF operations currently presents a challenge due to limited computational capacity. Asymmetrical products prevent in part the exploitation of symmetries in order to reduce model size and therefore increase the number of necessary finite elements. Furthermore, the three-dimensional stress and strain states, which lead to the forming of the functional elements, require volumetric finite elements and fine meshing. If the same element strategy is used for the remaining blank area, a very high number of integration points will be necessary, which, in turn, will make the calculation difficult. A powerful strategy to overcome this problem is the use of adaptive modeling based on the dual weighted residual method, which enables the local modification of the numerical mesh (Becker and Rannacher 2002). This way, it is possible to heighten the mesh efficiency in respect to the numerical demands without distinctly reducing the solution precision. Since SBMF processes typically possess high areas of contact between the workpiece and tool, efforts have been made to apply the model adaptivity on friction problems as well, in order to further reduce calculation time (Rademacher 2015).

Besides the simulation of material flow and tool load, damage criteria have to be deployed to effectively use the forming potential in a given process. However, the conventional forming limit



Sheet-Bulk Metal Forming, Fig. 6 (**b**, **c**) FE simulation result of an indentation process (Soyarslan et al. 2014), (**a**) corresponding incremental forming process (Sieczkarek et al. 2014)

diagram cannot be utilized since the plane stress assumption cannot be applied to SBMF processes. In order to overcome this challenge, a topic of current research is the adaption of alternative damage models for the purposes of SBMF. For example, Soyarslan et al. (2014) have shown that this is possible for the Gurson model, as has been successfully demonstrated by means of a process where a w-shaped punch penetrates in the edge of a sheet metal with a thickness of 2 mm (Fig. 6).

Cross-References

- ▶ Cold Forging
- ▶ Tailored Blanks

References

Becker R, Rannacher R (2002) An optimal control approach to a posteriori error estimation in finite element methods. Acta Numerica 10:1–102

Behrens B-A, Bouguecha A, Krimm R, Matthias T, Salfeld V (2011) Characterization of horizontal loads in the production of asymmetrical parts. Key Eng Mater 473:223–228

DIN 8582 (2003) Fertigungsverfahren Umformen-Einordnung; Unterteilung, Begriffe, Alphabetische Übersicht [Manufacturing process forming-classification, subdivision, terms and definitions, alphabetical index]. DIN German Institute for Standardization, Beuth, Berlin (in German)

Feintool AG (2015) Taumeltechnologie von Feintool – Hohe Qualität und Wirtschaftlichkeit beim Umformen, "Orbital forming technology of Feintool – High quality and economic efficiency for forming operations". Booklet, Jona

Gröbel D, Hildenbrand P, Engel U, Merklein M (2015) Influence of tailored blanks on forming of cold forged functional elements in a sheet bulk metal forming process. In: Stahl-Institut VDEh (ed) Proceedings of METEC, Düsseldorf

Hetzner H, Koch J, Tremmel S, Wartzack S, Merklein M (2012) Improved sheet bulk metal forming processes by local adjustment of tribological properties. J Manuf Sci Eng 133(6):1–11

Koch J, Merklein M (2013) Cold forging of closely-tolerated functional components out of blanks – possibilities of the new process class sheet-bulk metal forming. In: Ishikawa T (ed) Proceedings of the 6th JSTP international seminar on precision forging, Kyoto

Kudo H (1980) An attempt for classification of metal forming operations. CIRP Ann 29(2):469–476

Löffler M, Freiburg D, Gröbel D, Loderer A, Matthias, Stangier D, Weikert T (2015) Untersuchung von Tailored Surfaces hinsichtlich ihres tribologischen Einflusses auf Prozesse der Blechmassivumformung [Investigation of Tailored Surfaces regarding its influence on sheet-bulk forming processes.]. In: Tekkaya AE, Liewald M, Merklein M, Behrens B-A (eds) Tagungsband zum 18. Workshop Blechmassivumformung & 3. Industriekolloquium Blechmassivumformung 2015 – DFG Transregio 73, Dortmund,

February 2015. Shaker, Aachen, pp 132–146 (in German)
Merklein M, Tekkaya AE, Brosius A, Opel S, Koch J (2011) Overview on sheet-bulk metal forming processes. In: Hirt G, Tekkaya AE (eds) Proceedings of the 10th international conference on technology of plasticity, Aachen, September 2011. Stahleisen, Düsseldorf, pp 1109–1114
Merklein M, Allwood JM, Behrens B-A, Brosius A, Hagenah H, Kuzman K, Mori K, Tekkaya AE, Weckenmann A (2012) Bulk forming of sheet metal. CIRP Ann 61(2):725–745
Mori K (2012) Bulk forming of sheet metals for controlling wall thickness distribution of products. Steel Res Int SI:17–24
Nakano T (2009) Introduction of flow control forming (FCF) for sheet forging and new presses. In: Yoshida Y, Matsumoto R (eds) Proceedings of the 5th JSTP international seminar on precision forging, Kyoto
Rademacher A (2015) Mesh and model adaptivity for frictional contact problems. In: Ergebnisberichte Angewandte Mathematik. Fakultät für Mathematik TU Dortmund, Dortmund
Sieczkarek P, Isik K, Ben Kahlifa N, Martins PAF, Tekkaya AE (2014) Mechanics of sheet-bulk indentation. J Mater Process Technol 214:2387–2394
Soyarslan C, Plugge B, Faßmann D, Isik K, Kwiatkowski L, Schaper M, Brosius A, Tekkaya A (2014) An experimental and numerical assessment of sheet-bulk formability of mild steel DC04. ASME 133(6):1–9
Standring PM (1999) Advanced technology of rotary fording for novel automotive applications. Adv Technol Plast 3:1709–1718
Tan CJ, Mori K, Abe Y (2008) Forming of tailored blanks having local thickening for control of wall thickness of stamped products. J Mater Process Technol 202:443–449
Vierzigmann U, Koch J, Merklein M, Engel U (2012) Material flow in sheet-bulk metal forming. Key Eng Mater 504–506:1035–1040
Vierzigmann U, Schneider T, Koch J, Gröbel D, Merklein M, Engel U, Hense R, Bierman D, Krebs E, Kersting P, Lucas H, Denkena B, Herper J, Stangier D, Tillmann W (2013) Untersuchungen von Tailored Surfaces für die Blechmassivumformung mittels angepasstem Ringstauchversuch, [Investigations on Tailored Surfaces for Sheet-Bulk Metal Forming by means of an adapted ring compression test]. In: Merklein M, Behrens B-A, Tekkaya AE (eds) 2. Workshop Blechmassivumformung, Meisenbach (in German)

Shot Peening

► Peening

Silver Die Attach Bonding

► Silver Sintering

Silver Sintering

Aarief Syed-Khaja[1,2] and Jörg Franke[1]
[1]Institute for Factory Automation and Production Systems (FAPS), Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Nuremberg, Germany
[2]Heraeus Electronics, Heraeus Deutschland GmbH & Co. KG, Hanau, Germany

Synonyms

Low temperature silver die-attach; Silver die attach bonding

Definition

Silver sintering is a method to form a solid mass which does not require a phase change by application of heat and with/without pressure.

Extended Definition

Forming of solid mass of silver particles which acts as an interconnect medium between the electronic component and circuit board. The usage of this technique has been used mainly in high performance electronic applications where high thermal and electrical conductivity are demanded.

Theory and Application

Theory and Mechanism

Sintering has been used over hundreds of years for making ceramic components (Johnson 1978; Kingery et al. 2000). This technique finds nowadays application in various sectors of machining,

automotive, aerospace, medical applications, etc. (Upadhyaya 2000). Sintering is usually classified into several types based on the shrinkage or densification mechanism, mainly solid-state and liquid-phase sintering. Polycrystalline materials are usually sintered by solid-state diffusion and amorphous materials by viscous flow, i.e., undergo viscous sintering. Sintering method that makes use of a transient second phase that exists as a liquid at a specific sintering temperature is known as liquid-phase sintering (German 1985; Kwon 2006). The sintering process is considered to be occurring in three stages (German 1985; Bai 2005).

1. During the initial stage, in response to process parameters, a rearrangement of particles occurs by sliding and rotation. This leads to volume shrinkage by increased particles contact area followed by neck formation between particles resulting in an overall increase of the density of the material due to diffusion process.
2. The intermediate stage begins when the neck radius starts further increasing and equilibrium shapes of the pores are attained as driven by the surface and interfacial energies. The random-shaped pores at this stage can be continuous or interconnected. Densification at this stage is assumed to take place by reduction in pore volume where the pores become unstable and are pinched off from each other.
3. During the last stage, the isolated pores are completely eliminated, and a density value closed to the theoretical value is achieved. In addition, a possible grain growth can occur depending on the materials and sintering conditions.

The driving force for the sintering mechanism is mainly the tendency of the interlayer material to reduce its chemical potential or energy. There are various sintering mechanisms according to the nature of the materials; however, densification of the materials occurs mainly due to volume diffusions, i.e., grain boundary and lattice diffusion at the particle grain boundaries and dislocations. Due to the solid-state diffusion of bonding material, the void or pore distribution is more uniform compared to solder joints.

The main difference between soldering and silver sintering in electronic applications is that there is no phase change during sintering process. In the soldering process, solder paste is heated up until it reaches its melting point and transformed into liquid phase. Subsequently, it cools down to form a solid metal and intermetallic phase at boundaries. In the silver sintering process, heat is applied to a silver sinter paste and subsequently densification of solid by diffusion occurs resulting in a pure silver sintered joint.

Application of Silver Sintering

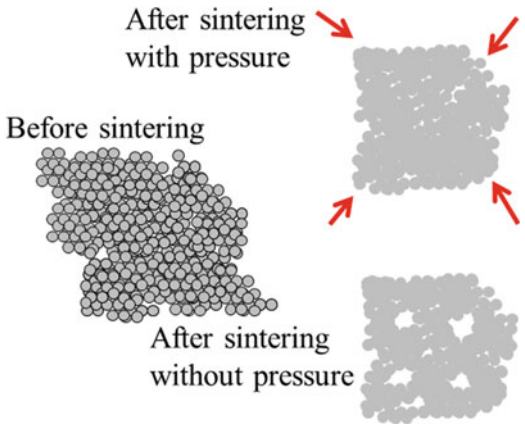
Sintered silvers are employed in a various range of applications such as solar cells, printed electronics, dental implants, and semiconductor packages (Layani and Magdassi 2011; Peng et al. 2015; Siow 2012). This article will focus on the application of silver sinter as die attach material for power electronics packages. The increasing demand for power electronics packaging with higher power density, higher voltage, and smaller size requires the replacement of silicon semiconductor components by wide band gap (WBG) semiconductor devices such as Silicon Carbide (SiC) and Gallium Nitride (GaN) (Kim 2014). The WBG devices require high operating temperature for better performance. Solder materials as the common die attach materials are inadequate for high temperature applications because of their low melting temperature. In recent years, die attach on power semiconductor using lead-free technique has attracted considerable interest. Silver sinter has demonstrated significant development over the past years to be considered one of frontrunner non-lead containing die attach solution mainly due to its high melting temperature (961 °C) and its excellent properties such as high thermal and electrical conductivity.

Silver sintering processes are generally classified as pressure and non-pressure sintering process based on the presence or absence of applied pressure during the process. Non-pressure sintering process is preferred over pressure sintering process mainly due to lower operating cost because a pressure sintering equipment is not required. However, pressure sintering process by far offers superior thermal and electrical

conductivity properties. The main difference between pressure and non-pressure sintering process is the porosity of silver sinter layer. The porosity of sinter layer obtained by non-pressure sintering process is higher than that obtained by pressure sintering process as explained in Fig. 1. The porosity of sinter layer can be decreased by applying pressure during sintering process: the higher the process pressure, the lower the porosity. It is known that thermal and electrical conductivity increase with decreasing the porosity of silver sintered layer (Wereszczak et al. 2012).

Sintering Process Flow

The process flow of non-pressure sintering is rather similar to soldering process in which paste is dispensed or printed on the substrate and the die is then placed on the deposited paste. Subsequently, non-pressure sintering process is performed in a programmable convection. The total process time of non-pressure sintering process can be about 4 h. The temperature profiles are generally in two steps with drying at 160 °C for 30 min and sintering at 230 °C for 60 min. This ensures reduced sintering defects and encourages maximum diffusion properties. At present, non-pressure sintering process is limited to small die sizes of maximum 25 mm^2. With bigger die sizes larger than 25 mm^2, ore voids and drying channels are observed when sintered under the standard non-pressure sintering profile.



Silver Sintering, Fig. 1 Silver sintered layer obtained by non-pressure and pressure sintering process

This is because it is difficult for the organic solvents to evaporate when a large die is placed on the wet sinter paste especially under the central region of the die. The long process time and die size limitations are the main reasons that non-pressure sintering process is not a favorable process in industry.

In case of pressure sintering process, there are four steps in the process flow including paste printing, predrying, component placement and pressure sintering. The predrying step is to ensure the organic solvents in the silver sinter paste are removed completely before die placement. Subsequently, die is placed on the predried sinter paste using a standard die bonder. After die placement, pressure sintering is performed under a sinter press at 230 °C for about 3 min. The total process time of pressure sintering process is considerably shorter than that of non-pressure sintering process. Predrying step before die placement reduces the risk of voids and drying channels formation. As a result, pressure sintering process is suitable for all die sizes.

Silver Sinter Pastes

Based on the literature review, silver sinter pastes can be classified into three categories: (1) micron-Ag paste, (2) nano-Ag paste, and (3) hybrid Ag pastes. Micron-Ag paste contains silver particles with sizes larger than 500 nm, whereas nano-Ag paste consists of silver particles with sizes less than 100 nm. The paste consisting of mixture of nano- and micro-silver particles is considered hybrid Ag-paste. It is believed that sintering of silver nano-particles occurs at lower temperature and lower pressure than silver micro-particles, and this phenomenon is attributed to the higher surface area and curvature of silver nanoparticles (Siow and Lin 2016). However, formulation of nano-Ag paste is more critical than micron-Ag paste as silver nanoparticles are reactive in air atmosphere so they can self-sinter in the absence of coatings around the particles or solvents in the paste.

Surface Metallization for Silver Sintering

Previous studies (Krebs et al. 2013; Schmitt and Chew 2017; Chew et al. 2017; Schmitt et al. 2017)

have revealed the feasibility of semiconductor devices attachment on silver and gold surfaces by either non-pressure or pressure sintering. Die attachment on bare Cu surface has attracted rising attention in recent years because eliminating precious metal finishing on substrate would represent significant compatibility to present supply chain and lower the entry barrier to adopt silver sintering solution. It has been demonstrated that it is feasible to produce silver sintered joint with high thermal cycle reliability on bare Cu surface by pressure sintering (Chew et al. 2018). In the study, it is observed that the initial die shear strength for Ag metallized substrate is higher than that for Au metallized and bare Cu. This relates to the self-diffusion of Ag faster than the silver/gold and silver/copper interdiffusion. An interesting phenomenon was discussed in the contribution was that the die shear strength for all the samples increased significantly after temperature cycling test with a condition of −40 °C/+150 °C and after a long-term storage at 250 °C. It is strongly believed that the sintering process is not yet completed under the mild sintering process conditions (230 °C, 10 MPa, 3 min) used in the study and consequently Ag, Au, and Cu continued to diffuse during temperature cycling test and high temperature storage and as a result strengthen the sintered joint. It is worth noting that after 2000 thermal cycles, the die shear strength for bare Cu substrate is relatively similar to the die shear strength for Ag metallized substrates. It was also observed that there is no further increase in die shear strength after 250 h storage at 250 °C indicating that the sintering process is completed after a certain time of storage.

Outlook

Cu Sinter Paste: Several studies (Kim 2014; Zhao et al. 2014; Zheng et al. 2014) have demonstrated that silver sintering is a promising die attach technology for high temperature power electronics packaging in which highly reliable bonding can be formed on surface with precious metal metallization and on bare Cu surface; however, the high cost of silver increases the entry barrier to adopt silver sintering in mass production. In recent years, Cu sinter paste has drawn increasing attention as an alternative sinter paste. It is believed that the material cost of Cu sinter paste is lower than that of Ag sinter paste. However, coatings around the Cu powder are important to prevent copper oxidation during handling or during the production of Cu sinter paste. As a result, manufacturing cost for Cu sinter paste might be higher than that for Ag sinter paste. A recent research (Nakako et al. 2017) reported that a Cu sintered layer on Cu, Ni, Au, and Ag surfaces with high thermal cycle reliability was obtained by pressureless sintering process with temperature above 225 °C under H_2 atmosphere. Nevertheless, sintering process performed under reducing atmosphere is challenging for high volume production environment. In order to increase the attractiveness of Cu sinter paste, Heraeus is focusing on developing a Cu sinter paste which is able to form a good bond by sintering under inert atmosphere.

Infrared Radiation Sintering: As mentioned previously, conventional non-pressure sintering process performed in a convection oven is not a favor process mainly due to the long process time which requires up to several hours and additionally the limitation in die size. Recent developments enabled an alternative so-called infrared radiation (IR) sintering for conventional non-pressure sintering process. The recent studies (Schmitt et al. 2017; Schmitt et al. 2018) demonstrated that IR sintering is feasible, and the total process time takes only about 90 min. The results show that die shear strengths for 90 min IR sintering and 4 h conventional non-pressure sintering are rather similar. In addition, a silver sintered joint formed between a die with a size of 26 mm^2 and Ag metallized lead frame showed no delamination and with void rates less than 1% after 96 h high-pressure test with conditions of 100% RH at 110 °C. There is no significant difference in the die shear strength before and after 96 h pressure cooker test. This study demonstrated that the total process time of non-pressure sintering can be reduced significantly using infrared radiation. Investigations of IR sintering for dies with sizes >25 mm^2 is ongoing. Further optimization of IR equipment such as installation of nitrogen supply to provide inert atmosphere for sintering process is in progress.

Silver Sinter Paste with a Reduced Young's Modulus: Silver sintering is an appropriate die attach technology for wide band gap semiconductor devices such as SiC and GaN which require high operating temperature. However, these WBG devices are harder than silicon components and a brittle die attach material might crack the components due to the CTE mismatch of the used materials. As a result, a sinter paste with low Young's modulus is desired in order to obtain a flexible bonding layer to avoid cracks of the WBG components. For this, silver sinter paste is developed by adding non-silver filler in the sinter paste to reduce the Young's modulus and to adjust the CTE of the sinter paste. The study performed in which standard silver sinter paste and the newly developed silver sinter paste were used as interconnect materials to attach SiC dies on Ag metallized substrate (Schmitt and Chew 2017). Thermal shock test with conditions of −65 °C/+150 °C under liquid to liquid was performed on the samples. It was observed that delamination in the sintered joint occurred after 1000 cycles for the standard paste, whereas, visible delamination in the sintered joint was not observed even after 2500 cycles for the newly developed sinter paste. It was observed that the crack initiated and propagation started and followed along the non-silver fillers towards the center of sintered layer. Presumably, the crack propagation along the fillers increases the reliability of sintered joint. The results from this study illustrate that a sinter paste with low Young's modulus is desired for wide band gap semiconductor devices. Further investigations are in progress in order to gain a better understanding of the failure mechanism of sintered joints for the newly developed sinter paste.

Cross-References

- ▶ Brazing and Soldering
- ▶ Diffusion Soldering
- ▶ Solder Paste Printing
- ▶ Wave Soldering

References

Bai G (2005) Low temperature sintering of nanoscale silver paste for semiconductor. Doctoral thesis, Virginia Polytechnic Institute and State University, Blacksburg

Chew LM, Schmitt W, Nachreiner J (2017) Sintered Ag joints on copper lead frame TO220 by pressure sintering process with improved reliability and bonding strength. Power conversion intelligent motion Europe, pp 1421–1425

Chew LM, Schmitt W, Schwarzer C, Nachreiner J (2018) Micro-silver sinter paste developed for pressure sintering on bare Cu surfaces under air or inert atmosphere. In: IEEE 68th electronic components and technology conference

German RM (1985) Liquid phase sintering. Plenum Press, New York

Johnson DL (1978) Fundamentals of the sintering of ceramics. In: Palmour H, Davis RF, Hare TM (eds) Processing of crystalline ceramics. Materials science research, vol 11. Springer, Boston, pp 137–149

Kim SS (2014) Are sintered silver joints ready for use as interconnect material in microelectronic packaging? J Electron Mater 43(4):947–961

Kingery WD, Bowen HK, Uhlmann DR (2000) Introduction to ceramics, 2nd edn. Wiley, New York

Krebs T, Duch S, Schmitt W, Kötter S, Prenosil P, Thomas S (2013) A breakthrough in power electronics reliability – new die attach and wire bonding materials. In: IEEE 63rd electronic components and technology conference, pp 1746–1752

Kwon O-H (2006) Liquid phase sintering. In: Cahn RW, Haasen P, Kramer EJ (eds) Materials science and technology. Classic softcovered. Wiley-VCH, Weinheim

Layani M, Magdassi S (2011) Flexible transparent conductive coatings by combining self-assembly with sintering of silver nanoparticles performed at room temperature. J Mater Chem 21(39):15378–15382

Nakako H, Ishikawa D, Sugama C, Kawana Y, Negishi M, Ejiri Y (2017) Sintering copper die-bonding paste curable under pressureless conditions. Power conversion and intelligent motion Europe, pp 91–95

Peng P, Hu A, Gerlich AP, Zou G, Liu L, Zhou YN (2015) Joining of silver nanomaterials at low temperatures: processes, properties, and applications. ACS Appl Mater Interface 7(23):12597–12618

Schmitt W, Chew LM (2017) Silver sinter paste for SiC bonding with improved mechanical properties. In: IEEE 67th electronic components and technology conference

Schmitt W, Chew LM, Miller R, Wolf A (2017) A new alternative non-pressure silver sinter process by using IR. Power conversion and intelligent motion Europe, pp 1426–1431

Schmitt W, Chew LM, Miller R (2018) Pressure-less sintering on large dies using infrared radiation and optimized silver sinter paste. In: IEEE 68th electronic components and technology conference

Siow KS (2012) Mechanical properties of Nano-Ag as die attach materials. J Alloy Compd 514(c):6–14

Siow KS, Lin YT (2016) Identifying the development state of sintered silver (Ag) as a bonding material in the microelectronics packaging via a patent landscape study. J Electron Packag 138:020804
Upadhyaya GS (2000) Sintered metallic and ceramic materials: preparation, properties and applications. Wiley, Chichester
Wereszczak AA, Vuono DJ, Wang H, Ferber MK, Liang Z (2012) ORNL/TM-2012/130 report: properties of bulk sintered silver as a function of porosity
Zhao Y, Wu Y, Evans K, Swingler J, Jones S, Dai X (2014) Evaluations of Ag sintering die attach for high temperature power module applications. In: IEEE 15th electronic packaging technology
Zheng H, Berry D, Ngo K, Lu GQ (2014) Chip-bonding on copper by pressureless sintering of nanosilver paste under controlled atmosphere. IEEE Trans Compon Packag Manuf Technol 4(3):377–384

Simulation

▶ Finite Element Analysis

Simulation of Machining Processes

▶ Modeling in Cutting

Simulation of Manufacturing Systems

Aydin Nassehi[1] and Marcello Urgo[2]
[1]Department of Mechanical Engineering, University of Bristol, Bristol, UK
[2]Dipartimento di Meccanica, Sezione Tecnologie Meccaniche e Produzione, Politecnico di Milano, Milan, Italy

Synonyms

Emulation

Definition

Simulation is the dynamic observation of an abstract model of a system through time with particular attention to the system's key attributes. The term is used extensively in manufacturing to refer to different types of such observations ranging from visual simulation of factories or individual machine tools to stochastic simulation of entire supply chains. Today, most simulation activities are carried out by computer software systems. Simulation systems are often categories into discrete-event simulation systems and continuous time simulation systems based on the approach they take with regard to advancing the time of the simulation forward.

Theory and Application

Discrete-Event Simulation

Discrete-event simulation is a type of simulation where the operation of a system is represented as a number of labelled points in time in chronological order. Each of these "points" is defined as an "event" and signifies a change in an aspect of the system's state. Each event is instantaneous. The state of the system is only determined when an event takes place; the time of the simulation, thus, jumps from one event to the next. The interim state of the system is either assumed to be unchanged or nondeterministic (Banks et al. 2009).

Method

(a) A determined state of the system is chosen as the *initial* state of the simulation, and the values for all variables that denote this state are stored. The timestamp is set at 0.
(b) All possible events at the initial state are then tabulated. The most immediate of these events is selected as the impending event.
(c) The timestamp is advanced to that of the impending event.
(d) The updated state of the system is generated by calculating the new values for all variables of the system at the updated timestamp.
(e) Newly possible events at the updated state of the system are tabulated, and the list of possible events is updated. The most immediate is selected as the impending event.
(f) Steps c–e repeated until the events for the entire timeframe of interest have been generated.

DES can be used for deterministic as well as stochastic modelling. For modelling stochastic systems, the probability distribution of the various variables that define the state of the system should be known in order to sample random values and generate the list of possible events in steps b and e (Sokolowski and Banks 2010).

Example

Consider the simple manufacturing planning problem shown in Fig. 1. Precut sheets of metal are delivered to a press (one sheet every 30 s). Loading and unloading the sheets take 2 s. If the pressing time is considered to be a random variable with normal distribution with a mean of 25 s and standard deviation of 5 s, how long will the first five components take to be manufactured?

The variables denoting the state of the system at each point in time can be specified as follows (Table 1):

The simulation can then be carried out as follows (Table 2):

In this particular simulation, it took 146.1 s to produce and unload the five products. This number will change if the simulation is run again as the random numbers sampled from the normal distribution used to model press time will be different.

Constituents of the DES Approach

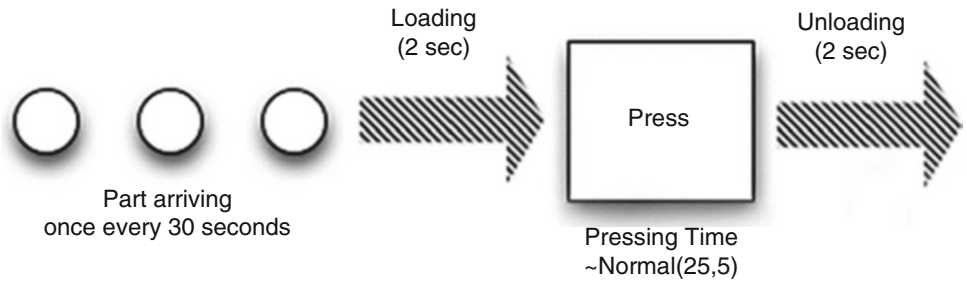
Regardless of the actual tools used to carry out DES, be it manually or using a computer, the following items are necessary:

The Clock It is essential in the simulation to keep the track of time and record the timestamp of events. The principal difference between real-time and discrete-events simulation techniques is that in continuous simulation, the timestamp advances by a fixed amount at each stage of the simulation, and the new state of the system is computed, whereas in discrete-event simulation, the timestamp is advanced immediately to the next event.

Events List It is necessary to be able to keep a list of impending event, (i.e., the list of possible events in one state). Without this list, it would be impossible to determine the impending event to be able to advance the clock.

Random Number Generators In order to carry out stochastic simulations, random number generators capable of generating random numbers with various probability distributions are necessary. Typical probability distributions used in DES include normal or Gaussian (to model the process time, maintenance time, or mean parts before failure for breakdowns), exponential (to model the period between arrivals and some service times), and uniform distributions among others. Pseudorandom number generators are often used in lieu of random number generators to allow rerunning simulations with behavior identical to that of a previous run.

Statistics The simulation tool should be able to track the statistics of interest and provide them as output to the user. For example, the total number of parts that go through a system should be maintained as part of the simulation.



Simulation of Manufacturing Systems, Fig. 1 A simple manufacturing system with a single machine (with infinite buffers before and after the process)

Termination Conditions As the simulation loop can go on forever, it is necessary to establish some conditions that would terminate the loop. These conditions can be formed as rules such as "finish if timestamp reaches x" or "finish after n parts are produced" or, in general, "when statistic S reaches Ω."

Simulation of Manufacturing Systems, Table 1 Variables modelling the state of the example system

Variable	Values	Description
p	0, 1, 2	p = 2 indicates that the press is blocked (i.e., there is a sheet on the press, but no operation is being carried out)
		p = 1 indicates that the press is busy
		p = 0 indicates that the press is idle
q	Integer	Number of sheets waiting to be pressed
ld	0 or 1	ld = 1 indicates that a new sheet is being loaded
		ld = 0 indicates that no loading is taking place
uld	0 or 1	uld = 1 indicates that the pressed sheet is being unloaded
		uld = 0 indicates that no unloading is taking place

Software Packages

A number of commercial (and academic) software programs have been developed to carry out discrete-event simulation. Among these, Arena (by Rockwell Automation), Witness (by Lanner), Simio (by Simio simulation software), and Tecnomatix Plant simulation system (by Siemens) are well known in manufacturing.

Continuous Time Simulation

In this type of simulation, the state of the system after the lapse of a small time period is defined based on the previous state of the system (i.e., modelled). By knowing the state of the system at the initial time, it is possible to advance the time in small increments and calculate the updated state of the system after each advance (Law 2006). The time increment can be chosen infinitely small, so, in theory, this type of simulation can be used to approximate the continuous lapse of time. This

Simulation of Manufacturing Systems, Table 2 Simulation event log

Timestamp	Events	System state
0	First sheet arrives, loading starts	p = 0, q = 1, ld = 1, uld = 0
2	First sheet is loaded to the machine, press starts	p = 1, q = 0, ld = 0, uld = 0
27.9	Press finishes (press time = 25.9), unloading starts	p = 2, q = 0, ld = 0, uld = 1
29.9	Unloading finishes	p = 0, q = 0, ld = 0, uld = 0
30	Second sheet arrives, loading starts	p = 0, q = 1, ld = 1, uld = 0
32	Second sheet is loaded to the machine, press starts	p = 1, q = 0, ld = 0, uld = 0
60	Third sheet arrives, waits	p = 1, q = 1, ld = 0, uld = 0
61.2	Press finishes (press time = 29.2), unloading starts	p = 2, q = 1, ld = 0, uld = 1
63.2	Unloading finishes, loading starts	p = 0, q = 1, ld = 1, uld = 0
65.2	Loading finishes, press starts	p = 1, q = 0, ld = 0, uld = 0
86	Press finishes (press time = 20.8), unloading starts	p = 2, q = 0, ld = 0, uld = 1
88	Unloading finishes	p = 0, q = 0, ld = 0, uld = 0
90	Fourth sheet arrives, loading starts	p = 0, q = 1, ld = 1, uld = 0
92	Fourth sheet is loaded to the machine, press starts	p = 1, q = 0, ld = 0, uld = 0
118.3	Press finishes (press time = 26.3), unloading starts	p = 2, q = 0, ld = 0, uld = 1
120	Fifth sheet arrives, waits	p = 2, q = 1, ld = 0, uld = 1
120.3	Unloading finishes, loading starts	p = 0, q = 1, ld = 1, uld = 0
122.3	Press starts	p = 1, q = 0, ld = 0, uld = 0
144.1	Press finishes (press time = 21.8), unloading starts	p = 2, q = 0, ld = 0, uld = 1
146.1	Unloading finishes	p = 0, q = 0, ld = 0, uld = 0

approach is used extensively in system dynamics and control engineering.

Software Packages

Software packages for continuous time simulation include Vensim, Stella, and Matlab Simulink. The use of continuous simulation for modelling manufacturing systems has been limited thus far.

Agent-Based Simulation

The universe of discourse is modelled as a number of distinct autonomous software entities (Leitão et al. 2012). Each individual software entity, or agent, has its own time-based state. State changes may be initiated by internal mechanisms (such as stochastic or deterministic delays) or through reaction to external stimuli or reception of messages. This methodology is powerful in modelling systems where emergent collaboration, cooperation, or competition comes to exist as the result of interaction between the entities. Smart manufacturing systems, for example, where the products and machines negotiate production processes, are a good candidate for modelling with agents. In most multi-agent environments, the timing of the events across different agents is asynchronous. In other words, the overall order of execution of tasks across different agents is not guaranteed, and additional constructs are required to ensure a sequence of events. In practice, with a very high number of agents executing in parallel, the management of the sequence of changes in states is a complex task requiring significant effort.

Software Packages

Software packages for agent-based simulation include Anylogic and NetLogo.

Common Applications for Simulation

Simulation is widely used across different sectors and industries for performing a wide variety of analysis including the following (Law and McComas 1998; Smith 2003):

- Process troubleshooting: The simulation is used to diagnose problems in complex systems. It is well understood that improvements in a system are only meaningful if the "bottlenecks" are improved. Simulation helps to identify the bottlenecks.
- Planning a new system: When a new system is being designed, it is possible to assess how well it will perform using a simulation and to use this data to choose the best design from a number of alternatives.
- Investigate improvement ideas: Simulation can be used to understand the effects of proposed changes and assess their effectiveness before being put into place.
- Stress test a system: Many systems perform adequately at normal circumstances but could fail in occasional scenarios. Using simulation, it would be possible to test existing systems against such scenarios.

Cross-References

- ▶ Computer-Integrated Manufacturing
- ▶ Manufacturing System
- ▶ Production Planning
- ▶ System

References

Banks J, Carson JS, Nelson BL, Nicol DM (2009) Discrete-event system simulation, 5th edn. Prentice Hall, Upper Saddle River

Law AM (2006) Simulation modeling and analysis, 4th edn. McGraw-Hill, New York

Law AM, McComas MG (1998) Simulation of manufacturing systems. In: Proceedings of the 30th conference on winter simulation. IEEE Computer Society Press, Atlanta, pp 49–52

Leitão P, Barbosa J, Trentesaux D (2012) Bio-inspired multi-agent systems for reconfigurable manufacturing systems. Eng Appl Artif Intell 25(5):934–944

Smith JS (2003) Survey on the use of simulation for manufacturing system design and operation. J Manuf Syst 22(2):157–171

Sokolowski JA, Banks CM (2010) Modeling and simulation fundamentals: theoretical underpinnings and practical domains. Wiley, Hoboken

Simultaneous Engineering

▶ Cooperative Engineering

SK

▶ Tool Holder

Skiving

▶ Gear Cutting

Sliding Mode Control

▶ Control

Slot Milling

▶ Groove Milling

Slotting

▶ Groove Milling

Smart Materials

▶ Actuator

Smart Products

Michael Abramovici
Lehrstuhl für Maschinenbauinformatik, Ruhr-Universität Bochum, Bochum, Germany

Synonyms

Smart system

Definition

Smart Products are cyber-physical products/systems (CPS) which additionally use and integrate internet-based services in order to perform a required functionality. CPS are defined as "intelligent" mechatronic products/systems capable of communicating and interacting with other CPS by using different communication channels, i.e., the internet or wireless LAN (Lee 2010; Rajkumar et al. 2010).

Theory and Application

Definition of Products

The term product [lat. productus] has a diversity of meanings in the context of different disciplines. For example, in mathematics, a product is the result of a multiplication. In chemistry, the term "product" refers to substances resulting from chemical reactions. In economics, the gross domestic product (GDM) measures the economic performance of a country or region. In marketing, a product is defined as anything that can be offered to a market and might satisfy a want or need (Kotler and Keller 2006).

The definition of products observed in this entry is geared to the engineering-specific perspective described in the EN ISO 9000 guideline. In accordance with this guideline, "a product is a physical or digital good, resulting from a value adding process (i.e., manufacturing process) and offered to the market to satisfy customer needs" (DIN EN ISO 2005). This definition makes a clear distinction between products (goods) and product-related services as complementary counterparts to these products. Services are defined as activity chains intended to generate customer value. In the last decades, products and product-related services have also been offered as integrated solutions aiming to provide customer satisfaction and value. These bundled offerings are defined as Product Service Systems (Meier et al. 2010).

The spectrum of industrial products is very large, ranging from single components (i.e., screws, axes, or chips) to extremely complex

systems (i.e., cars, airplanes, or factories), from nano to macro products, from mass-produced to mass-customized or personalized products. For very complex or networked products, the term "product system" is used as a synonym. According to the International Council of System Engineering INCOSE, a system is a construct or a collection of different elements that together produce results not obtainable by the elements alone (INCOSE 2014). In the following definition of Smart Products, the generic term "product" will be used for both single, simple products and complex product systems.

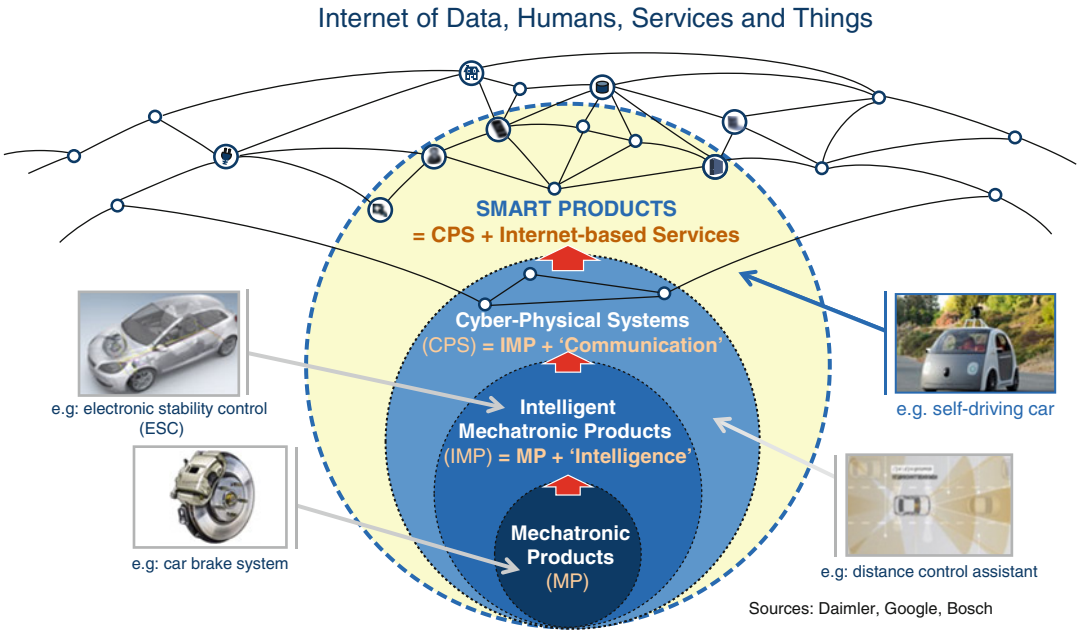
Definition of "Smart"

The meaning of the adjective "smart" is described in several dictionaries as related to the attributes of "clever, intelligent, bright, agile" etc. (Webster Dictionary 2014).

Evolution of Traditional Products Towards "Smart Products"

In order to better understand the definition of Smart Products, it is helpful to follow the evolution of traditional industrial products in the last decades, which was mainly driven by the tremendous advances of the information and communication technologies (Abramovici et al. 2013), (Fig. 1).

The core of each industrial product is a mechanical component or structure. The functional enhancement of mechanic products by electronic products has led to so-called mechatronic products (VDI 2206 2004; Neugebauer et al. 2007), e.g., a car brake system. Over the past four decades, mechatronic products have increasingly incorporated embedded microcomputer devices and software improving their behavior and functionality. The majority of current state-of-the art products are mechatronic products. Mechatronic products/components provide the foundation for Smart Products. The dramatic miniaturization of micro-embedded devices and the advances of the embedded software within mechatronic products have continuously improved their capabilities regarding their autonomy, self-optimization, and real-time interaction with their environment by using performant sensors and actors. These features describe the first evolutionary step of mechatronic products, the so-called "intelligent" mechatronic products (Neugebauer et al. 2007). An example of an intelligent mechatronic product is an electronic stability



Smart Products, Fig. 1 Evolution of traditional products towards Smart Products

control system in a car. The next evolutionary step of mechatronic products was the extension of intelligent mechatronic products by the capability to communicate and to network with other products using different embedded communication units and channels. This category of products is defined as cyber-physical products or cyber-physical systems (CPS) (Rajkumar et al. 2010; Baheti and Gill 2011; NSF 2012). A distance control assistant can be considered an example of a CPS in a car.

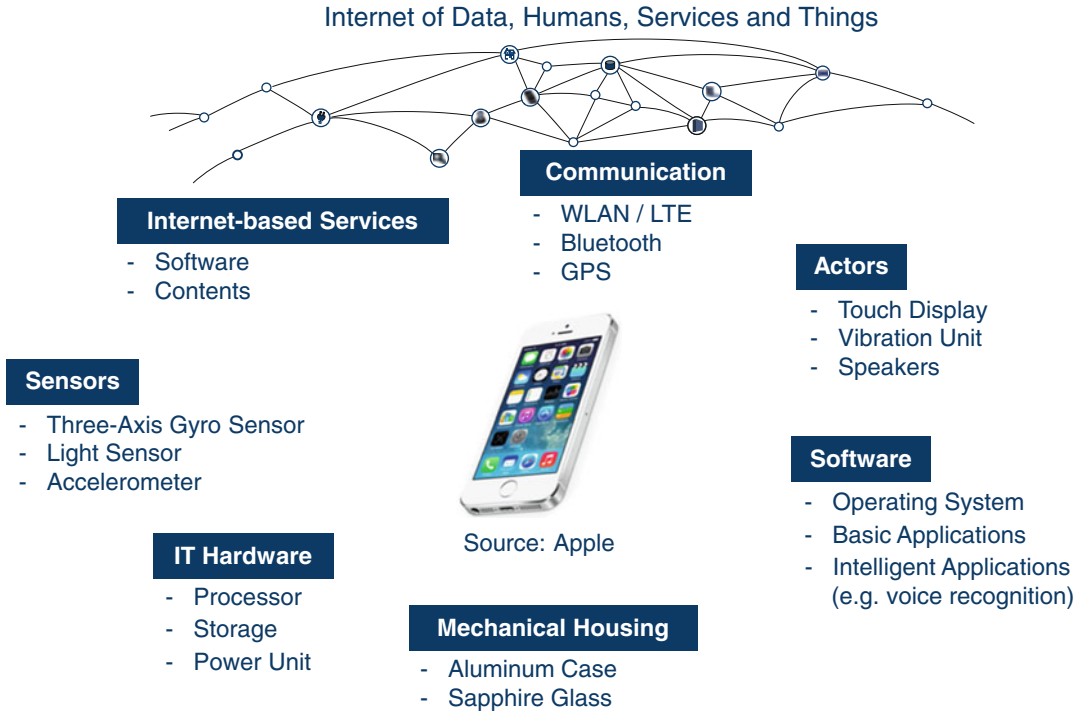
The synergies and convergence of different ICT innovations like the Internet IPv6 standard, Service Oriented Architectures, Web Services, semantic technologies, and the emerging number of cyber-physical products have extended the traditional Internet of Data (IoD) and of Humans (IoH) by an Internet of Services (IoS) and an Internet of Things (IoT), Fig. 1.

This new Internet of Data, Humans, Services, and Things enables CPS not only to communicate with each other but also to use and to integrate available internet services in order to perform required functions. The combination and integration of CPS with internet-based services has led to a new generation of intelligent, agile, flexible, and networked products. As all these features are related to the previously described meanings of the adjective "smart," these products are called "Smart Products" (Abramovici et al. 2014), Fig. 1. Products like a self-driving car illustrate the evolutionary stage of Smart Products.

The boundaries between the four described product types are blurred. Each evolutionary stage of a product is a prerequisite and a fundament for the next evolutionary stage.

Components and Characteristics of Smart Products

Smart Products break the boundaries between physical and virtual components as well as between products and internet-based services. The main components of a Smart Product are shown in Fig. 2 by using the example of the best known Smart Product – the smart phone. These components are as follows:



Smart Products, Fig. 2 Components of Smart Products using the example of a smart phone

- Mechanical housing (i.e., case, sapphire glass, display, buttons)
- IT hardware (i.e., processor, storage unit, power unit)
- Embedded software (i.e., operating system, voice recognition)
- Sensors (i.e., gyro sensor, light sensor, accelerometer)
- Actors (i.e., speakers, vibration unit)
- Communication unit (i.e., WLAN, Bluetooth, NFC)
- Internet-based services (i.e., apps)

The main characteristics of Smart Products are as follows:

- Huge number of heterogeneous components
- Context awareness
- Real-time reactivity
- High degree of autonomy
- High degree of connectivity
- High degree of personalization
- Self-learning capabilities
- Dynamic reconfiguration during the whole lifetime
- Human centration
- Ease of use

Applications

The described converging ICT innovations and the Internet of Data, Humans, Services, and Things penetrate all the traditional products in all industry sectors. Therefore, Smart Products rise in all industrial areas such as:

- The media and entertainment industry (i.e., smart phones, smart glasses, smart tablets, smart TV)
- Manufacturing (i.e., smart machines, smart robots, smart factories)
- Mobility (i.e., smart car, self-driving vehicles)
- Logistics (i.e., smart packagers, smart containers)
- Healthcare (i.e., smart clothes, smart hospitals)
- Energy (i.e., smart energy grid)

Due to their "intelligence" and networking capabilities, Smart Products allow the development of integrated trans-disciplinary and trans-sectorial solutions like the Smart House or the Smart City.

Most mature Smart Products on the market are consumer products like mobile smart devices using media, retail, or payment services. Huge national and international research programs like "Industrie 4.0" (Kagermann et al. 2013) and "Smart Service Welt" (Kagermann et al. 2014) in Germany and "Smart Manufacturing Leadership Coalition" (2014) in the USA or "Horizon 2020" by the European Union (European Comission 2014) will accelerate the development of Smart Products in all industrial sectors. The transition of traditional products to Smart Products and the development of totally new Smart Products, cloud services, and solutions are on the agenda of most industrial companies and are supported by a variety of large initiatives such as "Industrial Internet" by General Electric.

In the near future, Smart Products will also integrate traditional services provided by humans, which will lead to complex Smart Product Service Systems (Abramovici et al. 2014). Smart Products are considered a driving force of the so-called fourth Industrial Revolution.

Cross-References

- ▶ Industrial Product-Service System
- ▶ Mass Customization
- ▶ Mechatronics
- ▶ Industrial Product-Service System

S

References

Abramovici M, Stark R (eds) (2013) Smart product engineering: proceedings of the 23rd CIRP design conference, Bochum, Germany, March 11–13, 2013. Springer, Heidelberg/New York/Dordrecht/London

Abramovici M, Göbel JC, Neges M (2014) Smart engineering as enabler for the 4th industrial revolution. Paper presented at the 6th international conference on Integrated Systems Design and Technology (ISDT), Berkeley, 26–27 Mar 2014

Baheti R, Gill H (2011) Cyber-physical systems. In: Samad T, Annaswamy AM (eds) The impact of control

technology: overview, success stories, and research challenges. IEEE Control Systems Society, pp 161–166
Deutsches Institut für Normung (2005) DIN EN ISO 9000: Qualitätsmanagementsysteme, Grundlagen und Begriffe [Quality management systems – fundamentals and vocabulary]. Beuth, Berlin
European Commission (2014) Horizon 2020. http://ec.europa.eu/programmes/horizon2020. Accessed 22 Nov 2014
INCOSE (2014) International council on systems engineering. http://www.incose.org. Accessed 22 Nov 2014
Kagermann H, Wahlster W, Helbig J (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Acatech, Berlin
Kagermann H, Riemensperger F, Hoke D, Helbig J et al (2014) Smart service welt: recommendations for the strategic initiative web-based services for businesses. Acatech, Berlin
Kotler P, Keller KL (2006) Marketing-management, 12th edn. Pearson Prentice Hall, Upper Saddle River
Lee EA (2010) CPS foundations. In: Sepatnekar S - (ed) Proceedings of the 47th Design Automation Conference (DAC), Anaheim CA, USA, June 13–18, 2010. ACM/IEEE, New York, pp 737–742
Meier H, Roy R, Seliger G (2010) Industrial product-service systems (IPS^2). CIRP Ann Manuf Technol 59(2):607–627
Neugebauer R, Denkena B, Wegener K (2007) Mechatronic systems for machine tools. CIRP Ann Manuf Technol 56(2):657–686
NSF (2012) NSF 12–520 cyber-physical systems (CPS) program solicitation. National Science Foundation, Arlington. www.nsf.you/pubs/2012/. Accessed 6 Nov 2013
Rajkumar R, Lee I, Sha L, Stankovic J (2010) Cyber-physical systems: the next computing revolution. In: Sepatnekar S (ed) Proceedings of the 47th Design Automation Conference (DAC), Anaheim CA, USA, June 13–18, 2010. ACM/IEEE, New York, pp 731–736
SMLC Smart Manufacturing Leadership Coalition (2014) Smart manufacturing leadership. https://smartmanufacturingcoalition.org. Accessed 22 Nov 2014
VDI Verein Deutscher Ingenieure (2004) VDI 2206: design methodology for mechatronic systems (VDI guideline). Beuth, Berlin
Webster Dictionary (2014) Smart. http://www.webster-dictionary.org. Accessed 22 Nov 2014

Smart Structure

► Actuator

Smart System

► Smart Products

Smart Systems

► Cyber-Physical Systems

SMD Component Placement

Michael Pfeffer and Jörg Franke
Institute for Factory Automation and Production Systems (FAPS), Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Nuremberg, Germany

Synonyms

Pick and place; Surface mounted device (SMD) placement

Definition

The placement of surface mount technology (SMT) components on printed circuit boards (PCB) and its assembly task is defined as follows:

- Grip the right component from the pick-up position (respectively feeder)
- Place the component on the right position (target position) on the surface of printed circuit boards (PCB) with appropriate precision and accuracy

Extended Definition

The process of placing a surface mounted device (SMD) is structured in following steps:

1. Feeding
   (a) Depending on component shape and according to the packaging type of the components (paper or plastic tape and reel, tray, tubes, and bulk case), there exist different individual feeder types.
2. Gripping

(a) The components are gripped at the top-side by individually designed vacuum grippers respectively nozzles. These are customized according to the shape and weight of the component to hold the components reliable even at high acceleration forces.
(b) One or more vacuum nozzles are arranged at the placement head depending on the working principle of the placement head (pick and place, collect and place).

3. Alignment correction
   (a) A digital vision system with integrated cameras measures the real position of the PCB on the PCB transport system of the machine and the position of the component at the tip of the nozzle. The target positions of the components are adjusted depending on the individual measured offsets of the components and the PCB.
   (b) Position marks (fiducials) on the PCB define its zero point are captured by the PCB camera.
4. Transporting
   (a) The placement head is attached to a gantry system that moves the head (in x- and y-direction) to the particular target position of the component on the PCB.
5. Placing
   (a) The components are placed with a defined force onto the solder paste depots that were printed before on the pads of the circuitry.
   (b) The placement head moves the components in z-direction and places the component with right orientation on the PCB, so that the pins of the components are ideally aligned to the pads of the circuitry (Fig. 1).

Theory and Application

After printing solder paste on the pads of the circuitry on the PCB ("▶ Solder Paste Printing"), the placing of electric components is the second process step of the reflow soldering process (Klein Wassink and Verguld 1995). Thus, the components are placed subsequently in these solder depots. The final soldering process forms the electrically conductive and mechanically solid connections between the component pins and the pads of the circuitry (Feldmann et al. 2014).

In electronics production, SMD placement machines are used for automatically assembling electronic components in surface mount technology (SMT) on PCB (Scheel and Hanke 1999). These machines are capable to place a very high variety of components at an immense speed of up

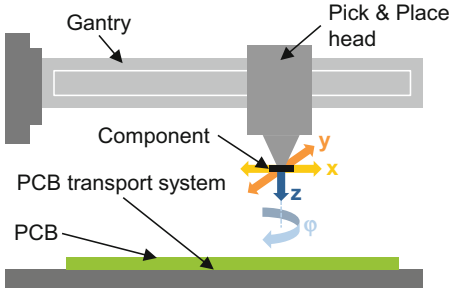
SMD Component Placement, Fig. 1 Functional elements of SMD placement machines

to 150,000 parts per hour. On one side, very small components, e.g., chip components like capacitors or resistors with the chip shape 0201 (metric) with a size of 0.250 mm by 0.125 mm or flip-chips with bump sizes down to 30 µm (pitch 100 µm), are placed precisely. On the other side, very big components like BGA with several hundred I/O-pins at the bottom-side or large connectors with a length of up to 200 mm are placed due to the high flexibility of the machine in handling a high component spectrum.

Placement machines generally have one or more Cartesian gantry system that moves the placement head. The gantry is set up by high dynamic linear motors, which allow high speeds and a high positioning accuracy. The head picks the components from the feeders and places the components on the exact position and orientation on the PCB. The precision and accuracy in placing components is guaranteed by the integrated vision system (Wohlrabe 2009). The PCB camera registers position marks (fiducials) on the PCB to measure the position of the PCB in the machine. The component camera optically measures the position of each component at the nozzle tip in x- and y-direction and orientation before placing on the PCB. With the offsets of the PCB and the components, the target positions of the components are adjusted (Wormuth and Zapf 2001).

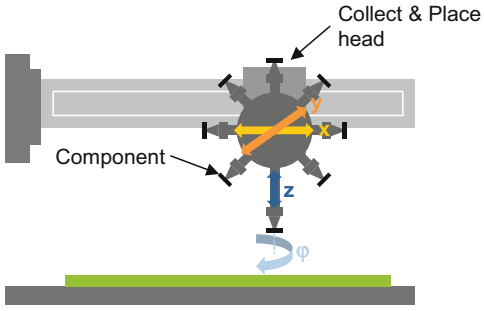
Two types of placement heads exist that are characterized regarding their working principle: the pick and place head and the collect and place head (Feldmann et al. 2014). The pick and place head (see Fig. 2) picks only on component one at a time, measures the component position and places the component on the PCB. This head achieves the highest placement accuracy of, e.g., 10 µm (3σ) due to the simple and rigid construction. However, the placement speed is reduced as only one component is picked and the process steps picking, measuring, and placing are not parallelizable.

The collect and place head (see Fig. 3) picks sequentially several components (from 6 up to 24, depending on the type) from the component feeders, simultaneously measures the component



Functionality Pick & Place	Components are picked up, are optically captured and placed separately and sequentially on the PCB.
Advantages	▪ High placement accuracy ▪ Big components and fine pitch components
Challenges	▪ Low placement speed

SMD Component Placement, Fig. 2 Pick and place placement head



Functionality Pick & Place	Several Components are picked up sequentially, are simultaneously optically measured and sequentially placed on the PCB.
Advantages	▪ Very high placement speed ▪ Chip components and small IC
Challenges	▪ Low placement accuracy ▪ Limited component spectrum

SMD Component Placement, Fig. 3 Collect and place placement head

positions, and places them sequentially on the PCB. Up to 30,000 components per hour at an accuracy of 40 µm (3σ) can be realized by this type of placement head. At this high speed, the time for picking, measuring, and placing the components is limited to a couple of milliseconds. Due to the complex kinematics of the head, the achievable placement accuracy is reduced.

The flexibility in placing a wide range of component shapes is given by specific vacuum nozzles that pick the component by vacuum. They are automatically changed by the machine for each job. The nozzles are customized to the component shapes. For example, 0201 (metric) components require a very fine nozzle with a tip size of 0.2 mm. Additional sensors can be integrated in the machine like a laser triangulation sensors to measure the planarity of the contacts of large-area components. A component height sensor measures the height of mechanically sensitive components.

Defects in the placement process (e.g., misaligned components in x- and y-direction and rotation, incorrect components, missing components, and tilted components) can be detected by means of automatic optical inspections systems.

Cross-References

- ▶ Assembly
- ▶ Feeding
- ▶ Gripping
- ▶ Solder Paste Printing

References

Feldmann K, Schöppner V, Spur G (eds) (2014) Handbuch Fügen, Handhaben, Montieren [Handbook joining, handling, assembling]. In: Edition Handbuch der Fertigungstechnik [Production engineering handbook], vol 5. Hanser, München (in German)

Klein Wassink R, Verguld M (1995) Manufacturing techniques for surface mounted assemblies. Electrochemical Publications, British Isles

Scheel W, Hanke H-J (eds) (1999) Montage [Assembly]. Baugruppentechnologie der Elektronik [Electronics assembly technology]. Verlag Technik, Berlin (in German)

Wohlrabe H (2009). Qualitätsoptimierung bei der Fertigung Elektronischer Baugruppen Mittels Statistischer Analysemethoden [Quality optimization in electronics production by means of statistical analysis methods]. In: System integration in electronic packaging, vol 7. Detert, Templin (in German)

Wormuth D, Zapf J (2001) Grundlagen der Surface Mount Technology [Basics of surface mount technology]. Siemens Dematic AG, München (in German)

Smoothing

▶ Polishing

Solder Application

▶ Solder Paste Printing

Solder Paste Application

▶ Solder Paste Printing

Solder Paste Printing

Aarief Syed-Khaja[1,2] and Jörg Franke[1]
[1]Institute for Factory Automation and Production Systems (FAPS), Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Nuremberg, Germany
[2]Heraeus Electronics, Heraeus Deutschland GmbH & Co. KG, Hanau, Germany

Synonyms

Solder paste application; Solder application; Solder printing

Definition

Application of the solder material on the circuit carrier in electronics production.

Extended Definition

The solder material is applied as a medium for connecting the electronic components on the

circuit board in surface mount technology. The paste material is applied either by using a stencil or screen on two-dimensional circuit carriers. The main task of this process step is to make available appropriate volume of solder material for a reliable post-soldering process.

Theory and Application

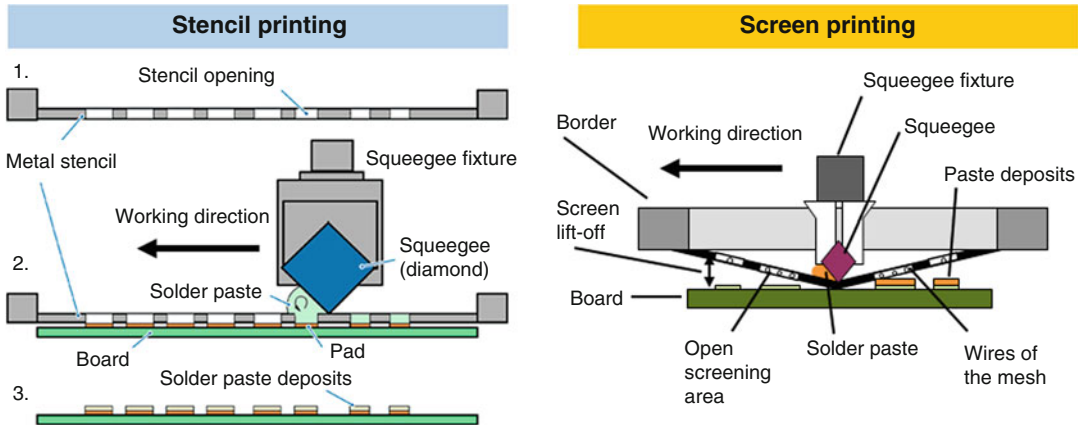
Theory

The solder paste printing is the first process step in the electronics production. Solder paste which is generally a suspension of metal particles in flux material is used as interconnection material. The flux material is a temporary adhesive consisting of activators, resins, and acid solvents. This holds the components until the soldering process melts the solder and makes an electrical and mechanical interconnection. The flux constituents mainly support in the removal of oxide layer on solder particles and circuit pads simultaneously protecting the components against oxidation. These also support the flowing as well as wetting of the solder melt to the component. The requirements for a solder paste are available in the IPC regulations (IPC 2012).

The solder material is thixotropic in nature, which means that its viscosity changes over time with applied shear force. Printing enables compatibility for large circuit boards and is also selected for its mass production efficiency. Application of the solder paste can be done by stencil printing (STP) or screen printing (SCP) as shown in Fig. 1. Both processes require similar machine platform, controlled motion, vision fiducial recognition, and alignment of the circuit carrier to the "layout" which is either designed in a stencil or mesh screen. The main differences between STP and SCP are tabularized as follows (Table 1):

Stencil printing generally involves the use of "trailing edge" metal squeegee blades with "on-contact" print (the circuit carrier and stencil come together during the squeegee deposition of solder paste). Screen printing is generally off-contact with a gap between the mesh and substrate; the squeegee blades are generally poly materials (not metal). In some cases, due to material viscosity, screen printing may employ a "flood-print, print-flood" cycle, meaning the material to be printed is spread thickly over the mesh by a squeegee "flood" blade and allowed to penetrate the metal mesh before a final print stroke generates the image/material deposit.

For many years the design of stencil apertures has been based on the original IPC7525 specification (IPC 2007) which recommended that aperture area ratios should be greater than 0.66 for acceptable stencil printing (to achieve in excess of 70–75% transfer efficiency). In the last decade, a tremendous amount of research and development has taken place with solder paste materials, stencil technologies, and process enhancements to improve paste transfer efficiency appropriate for the product and corresponding soldering process. The latest specifications are also summarized in the latest guidelines (IPC 2011). This



Solder Paste Printing, Fig. 1 Overview of the stencil printing and screen printing processes

Solder Paste Printing, Table 1 Overview of the paste printing processes

	Stencil paste printing	Screen printing
1.	Rigid stencils in thickness ranges of 50 µm to 400 µm are used	Flexible screens with wide range of mesh openings and wire diameters
2.	No space between the stencil and circuit board	Defined distance between mesh and circuit board
3.	Generally used for printing solder pastes	Generally used for printing epoxy-based conductive materials

corresponds to the stencil manufacturing, materials, and stencil coating (Burkhalter et al. 2007; Fleck and Chouta 2003; Coleman 2001; Coleman and Burgess 2006).The application of appropriate paste volume according to the electronic component is defined mainly by the applied force and velocity of the squeegee during printing as well as the selected dimensions such as thickness and apertures of the stencil/screen and the dimensions of the contact pads on the circuit board. The research shows that more than 60% of the errors in the electronic production can be traced back to solder paste printing. The quality of the paste printing process is often evaluated by a solder paste inspection step. Here the main quality parameters such as paste volume, transfer efficiency, and positional inaccuracies are traced, and the printing process parameters are subsequently optimized. Due to the errors in the printing process such as positional offsets, low/high paste volume, defects such as open connections, short connections, component offset, and component rotation occur. This technique is also applied in printing of conductive/non-conductive adhesives, sinter pastes.

Applications

Surface mount technology, Thick film technology, Wafer bumping, Solar cell technology, LTCC, Fine line printing

References

Burkhalter G, Leak E, Shea C, Tripp R, Wade G (2007) Transfer efficiencies in stencil printing. In: Proceedings of i-Connect007, SMT007, May 2007. http://smt.iconnect007.com/index.php/article/45167/transfer-efficiencies-in-stencil-printing/45170/?skin=smt. Accessed 14 May 2018

Coleman WE (2001) Stencil technology and design guidelines for print performance. Circuits Assembly 12:38

Coleman WE, Burgess MR (2006) Stencil performance comparison/AMTX electroform vs laser-cut electroform nickel foil. In: SMTA International 2006 and Assembly Technology Expo: "the power is you", September 24–28, 2006, Rosemont, Illinois, USA. Conference proceedings of the surface mount technology association, vol 1. Curran Associates, Red Hook, pp. 51–57

Fleck I, Chouta P (2003) A new dimension in stencil print optimization. J Surf Mount Technol 16(1):24

IPC (2007) IPC-7525A stencil design Guidelines. IPC, Bannockburn

IPC (2011) IPC-7525B stencil design guidelines, 3rd edn. IPC, Bannockburn

IPC (2012) IPC J-STD-005A requirements for soldering pastes. IPC, Bannockburn

Solder Printing

► Solder Paste Printing

Spark Erosion

► Electric Discharge Machining

Specific Energy

Konstantinos Salonitis[1], Apostolos Fysikopoulos[2,3], John Paralikas[2] and George Chryssolouris[2]
[1]Manufacturing Department, Cranfield University, Cranfield, UK
[2]Laboratory for Manufacturing Systems and Automation (LMS), Department of Mechanical Engineering and Aeronautics, University of Patras, Patras, Greece
[3]Automation Systems – Materials & Process Technologies, COMAU SpA, Grugliasco, Italy

Synonyms

Energy density; Energy per unit mass; Specific power

Definition

Specific energy is defined as the ratio of the energy required for the processing of a unit volume of material. It is a very important parameter, especially for the machining processes, and can be used as metric of comparing the energy requirements between different manufacturing processes. It has been defined in detail for almost all conventional manufacturing processes and research has been focused in estimating this in detail (indicatively for grinding process, Mishra and Salonitis 2013 calculated empirically the specific energy for grinding processes).

Specific energy can be defined for nonconventional manufacturing processes as well. Electrophysical and chemical processes are in general material removal processes. Specific energy (*SE*) is defined in their case as the ratio of the required energy (*E*) for removing a specific volume of material, to the volume of material removed (*V*):

$$SE = \frac{E}{A} \tag{1}$$

This parameter can be used to estimate the portion of the energy that is used for removing material. The higher machine tool efficiency and removal rate, the better the use of consumed power for removing material, for the same power.

Theory and Application

Electrical Discharge Machining (EDM)

Electrical discharge machining (EDM) is a thermal material removal process caused by electrical discharges (Salonitis et al. 2009). Material is removed utilizing two electrodes separated by a dielectric liquid and powered by an electric voltage (Dekeyser et al. 2003). There are different types of EDM, such as wire, sinker, and drill EDM. Each electrical voltage discharge is an output of energy. The energy distribution between the workpiece-wire electrodes and the actual effective removal energy is related to the distance between the electrode and the workpiece, flushing pressure, conductivity of the machining liquid, and discharge on time.

Specific discharge energy (*SDE*) is defined as the actual energy required for removing a unit volume of material (Liao und Yu 2004). An amount of real energy required to remove a unit volume of a specific material can be linked with the rate of energy distribution and the rate of volume of material removed:

$$SDE = \frac{(\text{Energy for discharge}) * (\text{Discharge Frequency})}{(\text{Rate of material removed volume})} \tag{2}$$

Specific energy requirements are presented for wire and drill EDM (Gutowski et al. 2006) for different rates of material removed volume. Regarding wire EDM, specific energy is 6.39 10^6 J/cm^3 for processing rate of 2.23 10^{-3} cm^3/s, and for drill EDM specific energy is 1.54 10^{10} J/cm^3 for processing rate of 1.7 10^{-7} cm^3/s.

Electrochemical Machining (ECM)

Electrochemical machining (ECM) is a material removal process into which pulsed current is dissipated through a conductive electrolytic solution between the tool and the workpiece. Such chemical interaction causes material to be removed from the workpiece according to the shape of the tool. ECM is based on thermal effects by extremely quick heating, melting, and vaporizing. The heat sources are the energy transfer between the plasma and the electrodes (Tönshoff et al. 1996). Dielectric fluid role is twofold, as it is building a discharge channel and increases the energy density between the electrode and the workpiece and it is as well removing the material out of the gap.

Specific pulsed energy of pulsed ECM is based on the voltage, currency, the pulse on time, and gap between tool and workpiece (Kozak 2004).

$$SDE = \frac{(\text{Power of electrical energy}) * (\text{“Pulse on time”})}{(\text{Volume of material removed})} \tag{3}$$

Specific pulsed energy consumption of ECM is affected by the pulse on time, which is increased with the increase of voltage and increase of gap

size. Experimental data have shown that the specific pulsed energy of ECM is 1 kJ/mm^3 for pulsed time of 2 ms and gap size of 0.2 mm (Kozak 2004).

Electro-Optical-Thermal Processes

Electron beam machining (EBM) and laser beam machining (LBM) are material removal processes by thermal processing. The main difference is that electrical energy is used to generate high-energy electrons in case of electron beam machining (EBM) and high-energy coherent photons in case of laser beam machining (LBM).

In EBM process, electrons are accelerated to a velocity nearly 75% that of light (~200,000 km/s). The process is performed in a vacuum chamber to reduce the scattering of electrons by gas molecules in the atmosphere. The electron beam is aimed using magnets to deflect the stream of electrons and is focused using an electromagnetic lens. The stream of electrons is directed against a precisely limited area of the workpiece; on impact, the kinetic energy of the electrons is converted into thermal energy that melts and vaporizes the material to be removed, forming holes or cuts (Chryssolouris 1991).

Laser is an acronym for light amplification by stimulated emission of radiation. LBM is accomplished by precisely manipulating a beam of coherent light to vaporize unwanted material (Chryssolouris 1991). Laser machining can replace mechanical material removal methods in many industrial applications, particularly in the processing of difficult-to-machine materials, as in the case of hardened metals, ceramics, and composites. In general, laser processes can be divided into one- (e.g., laser drilling), two- (e.g., laser machining, joining, heat/surface treatment), and three-dimensional processes (e.g., laser machining with two beams, laser processing with 5-axis cutting head, remote laser processing, laser-assisted machining, and laser shaping) (Tsoukantas et al. 2002).

Both in LBM (laser drilling, laser cutting, and laser machining with two beams) and EBM, the specific energy is defined as follows (Ahmadi et al. 2011, 2012; Xu et al. 2003; Thawari et al. 2005):

$$SPE = \frac{E}{V} \tag{4}$$

where E is the input energy (per joules) and V the removed volume of the material (per cubic centimeters); E is calculated as:

$$E = P * t \tag{5}$$

where P is the laser power and t is the interaction time. Typical values for rock drilling are 50–200 kJ/cm^3.

However, lasers are not used only in machining but also in other processes. Laser forming refers to the sheet metal forming processes such as bending. This is primarily based on establishing steep temperature gradients in the sheet by laser heating (during scanning) such that differential thermal expansion results in thermal stresses. The bending of sheet is caused by the plastic deformation of the material, when the thermal stress exceeds the temperature-dependent yield stress. Laser-based rapid prototyping processes such as stereo lithography (SL), selective laser sintering (SLS), laminated object manufacturing (LOM), and laser engineered net shaping (LENS) are used for the fabrication of a variety of complex shapes for a wide range of materials (Dahotre und Harimkar 2008; Tian et al. 2009). In addition to this, lasers also have been extensively used for joining (welding, soldering, and cladding) a variety of materials. The various laser welding processes involve spot welding, seam welding, and deep penetration welding. Laser welding generally involves the formation of keyhole by the surface vaporization of material (Mackwood und Crafer 2005; Coelho et al. 2000; Salonitis et al. 2013). In addition to this during laser cladding, which is a deposition welding process, a layer of powder is deposited on the substrate material, and the two materials are fused by metallurgical bonding through the action of a laser beam (Zeng et al. 1996; Choi et al. 2000; Lalas et al. 2007; Salonitis et al. 2016).

Weld quality is highly affected by the specific energy and that is why the scientific community is highly interested in this subject (Zeng et al. 1996; Choi et al. 2000; Dahotre und Harimkar 2008). Here, the point of interest is the scanning surface

so the definition of specific energy changes slightly (Coelho et al. 2000; Zeng et al. 1996). The energy density, E/A, can be considered the specific energy of the process, and is a parameter that combines the values of process variables with laser incident power, displacement speed, and laser spot area required for the welding of each type of plastic. During the scanning, the spot takes an elliptical shape and the *SPE* can be defined as follows (Coelho et al. 2000):

$$SPE = \frac{E}{A} = 1.27 * \frac{P}{\upsilon d} \tag{6}$$

where E is the energy input, A is the scanning area, P is the laser power, υ is the scanning velocity, and d should be considered as its minor axis when using an elliptical beam spot. The specific energy alone is not a suitable parameter for explaining laser clad or welding properties. Some values of the specific energy for laser cladding of cobalt-based coatings on low-alloy steel, for average interaction time and power density (e.g., 0.5 s and 15 kW/cm^2 respectively), can be about in the range of 0.5–15 kJ/cm^2. On the other hand, for welding a typical ferritic stainless steel, AISI 430 typical values are from 20 to 50 J/mm^2.

Cross-References

- ▶ Electric Discharge Machining
- ▶ Electron Beam Machining
- ▶ Laser Beam Machining

References

Ahmadi M, Erfan MR, Torkamany MJ, Safian GA (2011) The effect of interaction time and saturation of rock on specific energy in Nd:YAG laser perforating. Opt Laser Technol 43:226–231

Ahmadi M, Erfan MR, Torkamany MJ, Sabbaghzadeh J (2012) The effect of confining pressure on specific energy in Nd:YAG laser perforating of rock. Opt Laser Technol 44(1):57–62

Choi J, Choudhuri SK, Mazumder J (2000) Role of pre-heating and specific energy input on the evolution of microstructure and wear properties of laser clad Fe-Cr-C-W alloys. J Mater Sci 35(13):3213–3219

Chryssolouris G (1991) Laser machining-theory and practice. Springer, New York

Coelho JP, Abreu MA, Pires MC (2000) High-speed laser welding of plastic films. Opt Lasers Eng 34:385–395

Dahotre NB, Harimkar SP (2008) Laser fabrication and machining of materials. Springer, New York

Dekeyser W, Snoeys R, Jennes M (2003) A thermal model to investigate the wire rupture phenomenon for improving performance in EDM wire cutting. J Manuf Syst 4(2):179–109

Gutowski T, Dahmus J, Thiriez A (2006) Electrical energy requirements for manufacturing processes. In: Proceedings of the 13th CIRP international conference on life cycle engineering (LCE2006), Leuven, 31 May–2 June 2006, pp 623–628

Kozak J (2004) Thermal models of pulse electrochemical machining. Bull Pol Acad Sci Tech Sci 52:4

Lalas C, Tsirbas K, Salonitis K, Chryssolouris G (2007) An analytical model of the laser clad geometry. Int J Adv Manuf Technol 32:34–41

Liao YS, Yu YP (2004) Study of specific discharge energy in WEDM and its application. Int J Mach Tools Manuf 44:1373–1380

Mackwood AP, Crafer RC (2005) Thermal modelling of laser welding and related processes: a literature review. Opt Laser Technol 37:99–115

Mishra VK, Salonitis K (2013) Empirical estimation of grinding specific forces and energy based on a modified Werner grinding model. Procedia CIRP 8:287–292

Salonitis K, Stournaras A, Stavropoulos P, Chryssolouris G (2009) Thermal modeling of the material removal rate and surface roughness for die-sinking EDM. Int J Adv Manuf Technol 40:316–323

Salonitis K, Stavropoulos P, Fysikopoulos A, Chryssolouris G (2013) CO2 laser butt-welding of steel sandwich sheet composites. Int J Adv Manuf Technol 69:245–256

Salonitis K, D'Alvise L, Schoinochoritis B, Chantzis D (2016) Additive manufacturing and post-processing simulation: laser cladding followed by high speed machining. Int J Adv Manuf Technol 85:2401–2411

Thawari G, Sarin Sundar JK, Sundararajan G, Joshi SV (2005) Influence of process parameters during pulsed Nd:YAG laser cutting of nickel-base superalloys. J Mater Process Technol 170:229–239

Tian X, Günster J, Melcher J, Li D, Heinrich JG (2009) Process parameters analysis of direct laser sintering and post treatment of porcelain components using Taguchi's method. J Eur Ceram Soc 29(10):1903–1915

Tönshoff HK, Eggerl R, Klockez F (1996) Environmental and safety aspects of electrophysical and electrochemical processes. CIRP Ann Manuf Technol 45(2):553–568

Tsoukantas G, Salonitis K, Stavropoulos P, Chryssolouris G (2002) An overview of 3D laser materials' processing concepts. In: Proceedings of SPIE – the international society for optical engineering, vol 5131, pp 224–228

Xu Z, Reed CB, Konercki G, Parker RA, Gahan BC, Bataresh S, Graves RM, Figueroa H, Skinner H (2003) Specific energy for pulsed laser rock drilling. J Laser Appl 15(1):25

Zeng X, Zhu B, Tao Z, Cui K (1996) Analysis of energy conditions for laser cladding ceramic-metal composite coatings. Surf Coat Technol 79:162–169

Specific Power

▶ Specific Energy

Spheroidal Graphite Iron

▶ Machining of Spheroidal Ductile Iron

Spherulitic Graphite Cast Iron

▶ Machining of Spheroidal Ductile Iron

Spindle

Christian Brecher
Werkzeugmaschinenlabor WZL der RWTH Aachen, Aachen, Germany

Synonyms

Main spindle; Working spindle; Milling spindle; Turning spindle; Grinding spindle; Drilling spindle

Definition

A spindle is a rotating shaft with a fixture for holding a tool (in the case of a milling, grinding, or drilling spindle) or a workpiece (in the case of a turning spindle). The spindle shaft serves as a support, a positioner, and a rotary drive for the tool or workpiece.

Theory and Application

Spindle Types and Applications

The spindle shaft must take up any machining forces arising during cutting with the lowest possible deformation response, generate/transmit the cutting power provided by an internal or external drive for machining, and exhibit high positioning and running accuracy. In machine tools, various types of main spindles are used to satisfy different requirements. Turning and grinding spindles must achieve extremely high concentricity at a high stiffness and usually medium speeds, whereas milling and drilling spindles are used at (in part) high speeds under changing operating conditions.

Main spindles can be classified as directly driven (motor spindles) or externally driven versions. In directly driven spindles, the power is generated via a motor integrated into the spindle housing, with the stator seated on the spindle shaft. In contrast, the drive in the externally driven version is positioned outside of the housing and connected to the spindle shaft by a coupling or gear. Normally synchronous or asynchronous machines are used as drives.

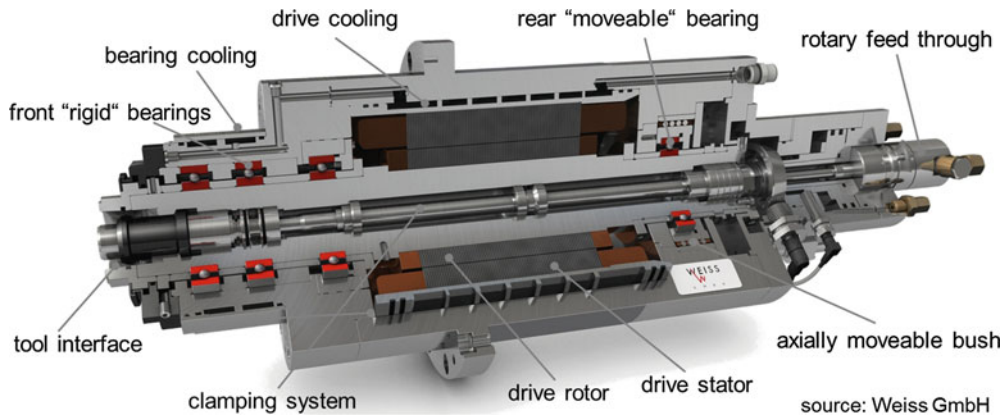
Requirements

The overall machine productivity is essentially determined by main spindle performance parameters such as maximum speed, available torque, maximum cutting depths yielded from stiffness and damping, and the resultant stock removal volumes.

The main spindle is a key component in workpiece cutting. Especially in high-speed cutting (HSC) and high-performance cutting (HPC), it represents a performance-limiting factor in many applications. In high-speed cutting applications, it must guide the tool axially and radially with a very high precision and take up medium forces at extremely high speeds. In HPC, the requirements are shifted from high-speed capabilities to medium speeds and high moments and forces. Spindles can be designed optimally for either HSC or HPC. If a spindle needs to be used in both areas, a compromise between stiffness and load capacity on one hand and achievable speeds on the other hand must be made.

Structure

Spindles are made up of a large number of individual components. A typical design of a

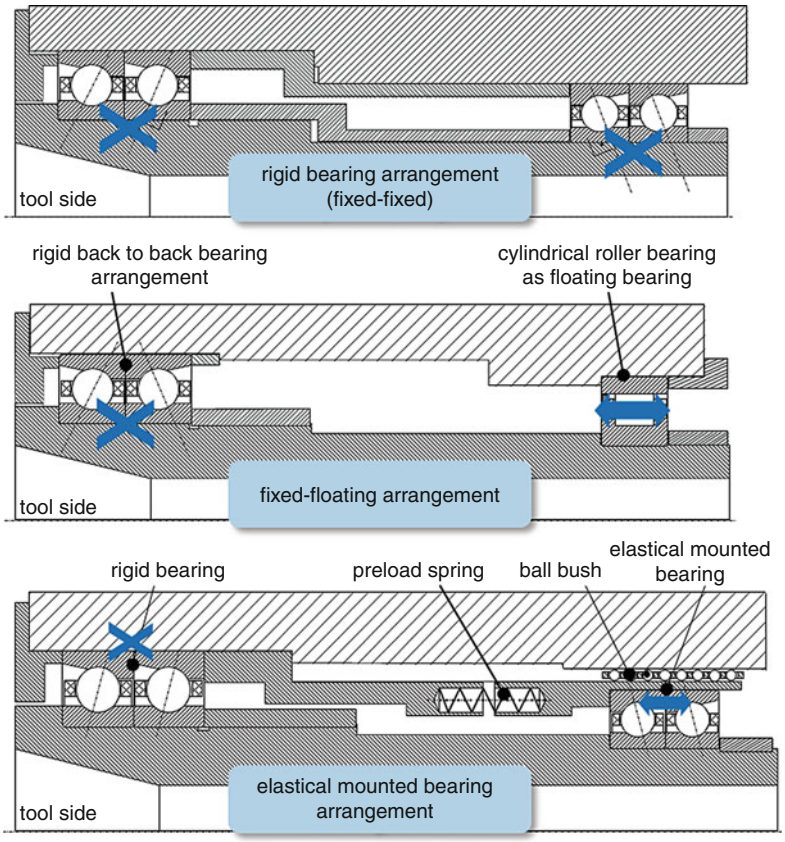


Spindle, Fig. 1 Structure of a motorized milling spindle

main spindle is shown in Fig. 1. The rotating spindle shaft is supported by two or more bearings. The bearing system is the most important performance-determining unit in the spindle, besides the motor. High-performance spindles are usually designed as motorized spindles with drives realized by single synchronous or asynchronous motors situated between the bearings. Due to the high power density of the motor, cooling by an external water cooling system is necessary. The outer bearing rings are often also cooled. A clamping system in the shaft enables rapid, automatic tool changes via a standard *tool interface*. Running longitudinally in the center of the clamping system is a pipe, in which cooling lubricant can be transferred to the rotating tool via a rotary feed-through at the end of the spindle. Noncontact seals, typically designed to be gas barrier seals, ensure that no contaminants can penetrate into the spindle. Integrated sensors monitor various operating parameters, for example, the angular position of the rotor via a rotary encoder; the position of the tool clamping system, motor and, in part, bearing temperatures; and others.

The spindle bearing has a great effect on the operating properties and the performance. In general, spindles can have hydrostatic, hydrodynamic, aerostatic, or electromagnetic *bearings* as well as roller bearings, depending on the given application. Hydraulic bearings are used in special cases, for example, in cases in which constant operating conditions prevail and extremely high concentricity are required. Aerostatic and electromagnetic bearings are suitable for low forces and extremely high speeds.

Spindle bearings, with high-speed capabilities and good stiffness properties, are mainly used in practice. They are also the most inexpensive of the bearing variants. However, roller bearings exhibit speed- and temperature-dependent kinematic properties, which are discussed in detail in the literature (Spachtholz 2008; Tüllmann 1999). This requires careful design of the bearings and the surrounding structure. If extremely high stiffness values and moderate to medium speeds are required, cylindrical roller and tapered roller bearings are also used. Through their line contact, they generate more friction and thus tend to produce more heat, which is difficult to dissipate, especially in the inner ring. Cylindrical roller bearings react to the resulting differences in thermal expansion with an increase in internal stresses and are hence more prone to failure. Apart from the selection of the bearing itself, the arrangement is especially decisive for the operating behavior of the spindle. During operation the spindle shaft, the bearing rings, and the housing are subjected to temperature changes and centrifugal forces. For no undesired changes to occur in the bearing preload through the resultant thermal expansion and kinematic displacements, the bearing arrangement must be appropriately designed. Basically the three possibilities shown in Fig. 2 exist for designing the main spindle bearings: fixed-fixed bearing, fixed-floating bearing, and (for extremely



Spindle, Fig. 2 Bearing arrangements

high speeds) elastically mounted bearings (Weck and Brecher 2006). The axially movable arrangement of one bearing side enables compensation for thermal and kinematic displacements and is hence more suitable for high speeds. The achievement of this movability is extremely important. Bearing sleeves with sliding fit, ball bushings, diaphragm spring bushings, etc. are used for this (Butz 2007). The elastic preload is normally achieved using springs, but hydraulic or piezoactuator solutions are also possible.

Ceramic rolling elements are widely used to increase the maximum allowable bearing speed. They not only exhibit better tribological properties in contact with steel but also are subjected to lower centrifugal forces thanks to their lower densities.

To minimize friction and wear, it is necessary to separate the rolling elements by a stable lubricant film. Hence, reliable supply of lubricant is crucial. In machine tool main spindles, bearings are predominantly operated with grease lubrication due to the low costs, low complexity, and low-friction running thanks to the minimum quantity lubrication (MQL). If, however, dn factors of greater than n x $d_m = 2*10^6$ mm/min are required, oil-air lubrication enabling dn factors of up to $3*10^6$ mm/min to be reached is usually used (Schaeffler 2014). The effect of different MQL methods on the operating behavior of spindle bearings should not be ignored and has been investigated extensively in the literature (e.g., Koch 1996).

The particular interface between spindle shaft and machining tool has a decisive effect on the machining result in the manufacturing process. Apart from the widespread steep taper (SK) interface, other standardized interfaces (HSK, PSC, TS) with considerably improved stiffness, accuracy, and speed capabilities have become established.

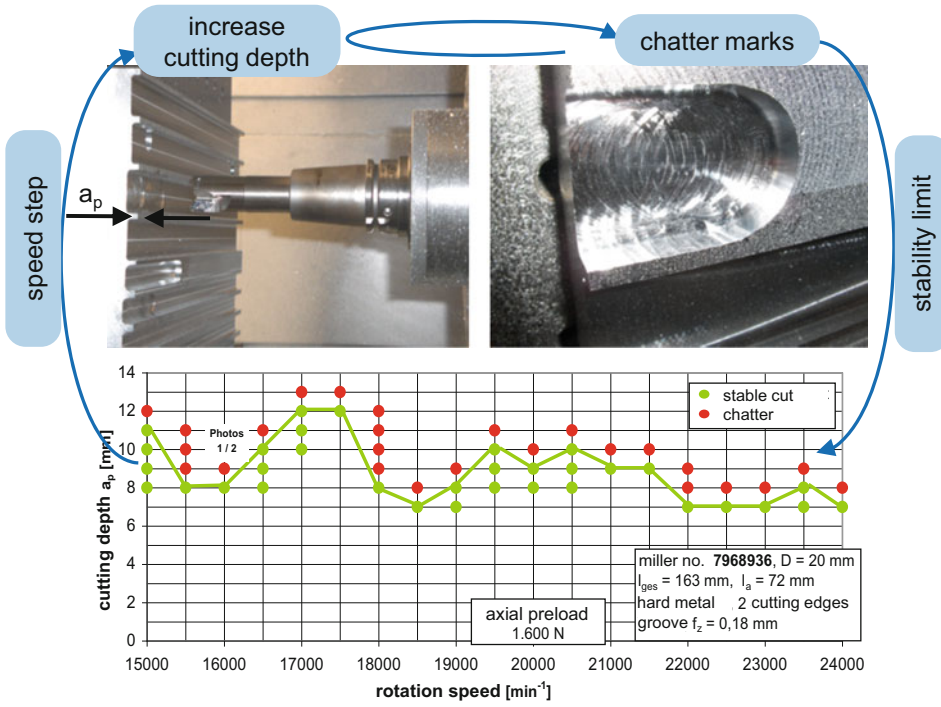
Design and Calculation

A number of factors limit the performance of a main spindle. The limitation of the allowable static and dynamic loads acting on the spindle via the tool tip (TCP) is an obvious one. Whereas the static limit plays more of a subordinate role in practice, the dynamic load capacity is extremely important. It is yielded out of the interplay between stiffness and damping characteristics of all components that affect the distribution of forces. The determining factors are the load transfer of the tool and the shaft, the stiffness and damping values of the bearings, and the effect of additional masses such as the rotor mass (Kreis 2008). If a frequency-dependent excitation threshold is exceeded in the cutting process, the spindle-bearing-machine system starts to vibrate at a frequency close to one of its resonant frequencies, and the process becomes unstable. The maximum allowable cutting depths at which stability is still given are speed-dependent and can be illustrated by a so-called chatter stability map, as shown in the example in Fig. 3. In general, high stiffness values, low masses, and good damping properties of the components have a positive effect on performance.

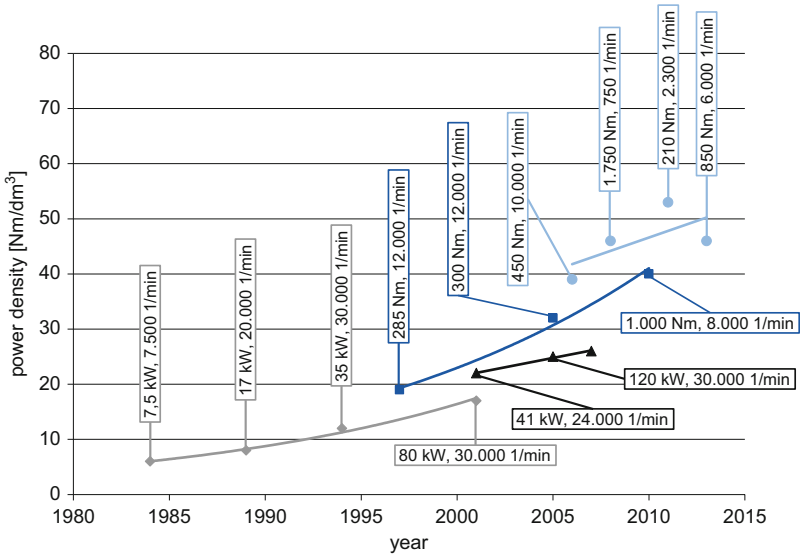
The heating of the shaft and the bearings represents another limitation. This results from electrical losses in the motor and from friction energy generated in the bearings and in the cutting process itself. If the components become too hot, the positioning accuracy is impaired due to the thermal expansion, and stresses as well as unallowable loads and changes in material properties can occur. Efforts are hence aimed at maximizing the efficiency of the electrical machine and minimizing the bearing friction.

Along with the cutting performance, the load on many spindle components is also increased, and the life of these components is accordingly decreased. The increased productivity is hence always weighed against this effect to determine the cost-effectiveness.

Because of the multitude of interactions between individual components, bearing design and prediction of the operating behavior of a spindle are no trivial tasks. For example, centrifugal effects, thermal effects, stiffness and damping characteristics, and application of force must be taken into account. This can be accomplished using a large number of



Spindle, Fig. 3 Chatter stability map



Spindle,
Fig. 4 Development of main spindle performance

software tools, the majority of which make use of finite element methods and take into account different influencing factors (Altintas and Cao 2005; Kreis 2008). A comprehensive description of spindle calculation methods can be found in the literature (Abele et al. 2010).

Recent Developments and Outlook

Over the last two decades, improvements such as those made in cutting materials have led to a continuous increase in the performance of main spindles (Fig. 4, Development of main spindle).

Development efforts are currently focused on optimization of the bearing systems, e.g., through modification of groove geometries in spindle bearings or new lubrication concepts (Brecher et al. 2007 and Spachtholz 2008), integration of sensors and actuators for monitoring and influencing of the spindle behavior in the process, and improvement of auxiliary systems such as the clamping system or the rotary feed-through. Other current topics include optimization of drives for high speeds and simultaneously high torques and energy efficiency (Damm 2015 and Heyers 2013).

Cross-References

► Machine Tool

References

Abele E, Altintas Y, Brecher C (2010) Machine tool spindle units. CIRP Ann Manuf Technol 59(2):781–802

Altintas Y, Cao Y (2005) Virtual design and optimization of machine tool spindles. CIRP Ann Manuf Technol 54(1):379–382

Brecher C, Spachtholz G, Paepenmüller F (2007) Developments for high performance machine tool spindles. CIRP Ann Manuf Technol 56(1):395–399

Butz F (2007) Lösungen zur Gestaltung der Loslagerung von Werkzeugmaschinenspindeln [Examples for the floating bearing-design of machine tool spindles]. Dissertation, RWTH Aachen (in German)

Damm H (2015) 160 Liter Späne pro Minute. WB Werkstatt+Betrieb 5/2015; S.38–41, Carl Hanser, München (in German)

Heyers C (2013) Energieeffizienter Betrieb von Asynchron-Hauptspindelantrieben in Werkzeugmaschinen. Disserstation, RWTH Aachen (in German)

Koch A (1996) Steigerung der Höchstdrehzahl von Schrägkugellagern bei Ölminimalmengenschmierung [Improvement of the maximum speed of angular contact bearings using minimum oil quality lubrication]. Ph.D. thesis, RWTH Aachen. (in German)

Kreis M (2008) Zum Eigenverhalten von Motorspindeln unter Betriebsbedingungen [On the behavior of motorized spindles under operational conditions]. Ph.D. thesis, TU Darmstadt. (in German)

Schaeffler KG (2014) Super precision bearings: spindle bearings, super precision cylindrical roller bearings, axial angular contact ball bearings. Schaeffler Technologies GmbH & co. KG, Herzogenaurach. http://www.schaeffler.de/content.schaeffler.de/de/mediathek/library/library-detail-language.jsp?id=114232

Spachtholz G (2008) Erweiterung des Leistungsbereichs von Spindellagern [Power-range extension of spindle bearings]. Ph.D. thesis, RWTH Aachen (in German)
Tüllmann U (1999) Das Verhalten axial verspannter schnelldrehender Schrägkugellager [The behavior of axially tensioned high-speed angular contact bearings]. Ph.D. thesis, RWTH Aachen. (in German)
Weck M, Brecher C (2006) Werkzeugmaschinen – Konstruktion und Berechnung. [Machine tools – design and calculation], 8th edn. Springer, Berlin. (in German)

Spindle Bearing

▶ Bearing

Spinning Tool

▶ Turning with Rotary Tools

Springback

Z. Cedric Xia[1] and Jian Cao[2]
[1]Research and Innovation Center, Ford Motor Company, Dearborn, MI, USA
[2]Department of Mechanical Engineering, McCormick School of Engineering and Applied Science, Northwestern University, Evanston, IL, USA

Synonyms

Dimensional control; Shape distortion

Definition

Springback is generally referred to as the change of part shape that occurs upon removal of constraints after forming.

In many cases, such geometric change is undesirable as the final obtained part shape geometry deviates from the desired tooling geometry. Therefore, often people use "shape distortion" to describe springback. It is closely related to *dimensional control* for sheet metal parts.

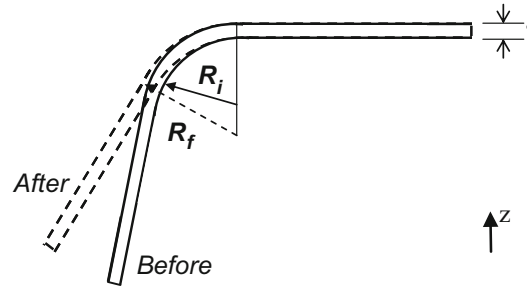
Theory: Mechanics of Springback

Springback can be considered as a dimensional change which happens during unloading, due to the occurrence of primarily elastic recovery of the part. After springback the part reaches an internal equilibrium in the absence of external forces. Residual stresses still exist within the part; however, they are self-balanced. Take a simple bending case in Fig. 1, for example. Under the loaded condition of bending, the top surface is under tension, while the inner surface is under compression (Fig. 2a). During the unloading stage, one can consider that an equivalent reverse moment is applied to the sheet (Fig. 2b), resulting in a residual through-the-thickness stress distribution as shown in Fig. 2c. Correspondingly, the radius of the bent sheet changes from R_i (before unloading) to R_f (after unloading). The change of the radius in a pure bending case can be roughly estimated by Eq. 1,

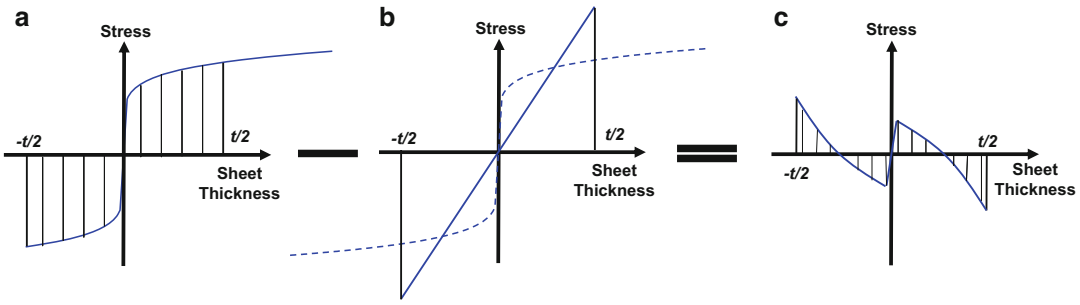
$$\frac{R_i}{R_f} \cong 4\left(\frac{R_i Y}{Et}\right)^3 - 3\left(\frac{R_i Y}{Et}\right) + 1 \qquad (1)$$

where E is the material's Young modulus, Y is the yield stress, and t is the sheet thickness.

The above equation works reasonably well for a simple bending case and provides the first-order



Springback, Fig. 1 Illustration of springback in a simple bending case



Springback, Fig. 2 Illustration of stress distributions in the sheet before and after springback. (**a**) Before springback, (**b**) elastic unloading, and (**c**) after springback

insights about the magnitude of springback with respect to some basic material properties. However, springback in a general three-dimensional situation is much more complicated as the stress distribution in a formed part is highly nonuniform and dependent on deformation history. Finite element methods are commonly adopted in the effort of accurately capturing the springback amount. One can find advancements and lessons learned in springback prediction from the benchmark activities organized along with the NUMISHEET conferences, in which both laboratory and industrial parts were examined.

Methods and Applications

Springback is usually the root cause when the geometry of formed parts deviates from the desired shape. A general representation can be described as:

$$\mathbf{U}_{\text{part}} = \mathbf{U}_{\text{die}} + \mathbf{U}_{\text{sb}} \tag{2}$$

where $\mathbf{U}_{\text{die}}$ is the die geometry which the sheet blank assumes at the end of a forming process, but before the tools are removed, $\mathbf{U}_{\text{sb}}$ is the dimensional recovery of the part from springback, and $\mathbf{U}_{\text{part}}$ is the final part shape after springback. In some cases where matched tools are not used in the forming process, $\mathbf{U}_{\text{die}}$ should be interpreted as the part shape just before tool removal.

Unloading is mostly a linear elastic process, and it is the dominant factor in the consideration of springback. However, if the sheet metal such as some advanced high-strength steels exhibits strong Bauschinger behavior, or the formed part has high stiffness to resist significant shape changes, reverse plastic deformation can occur during the unloading stage, which will normally contribute to larger springback than pure elastic recovery.

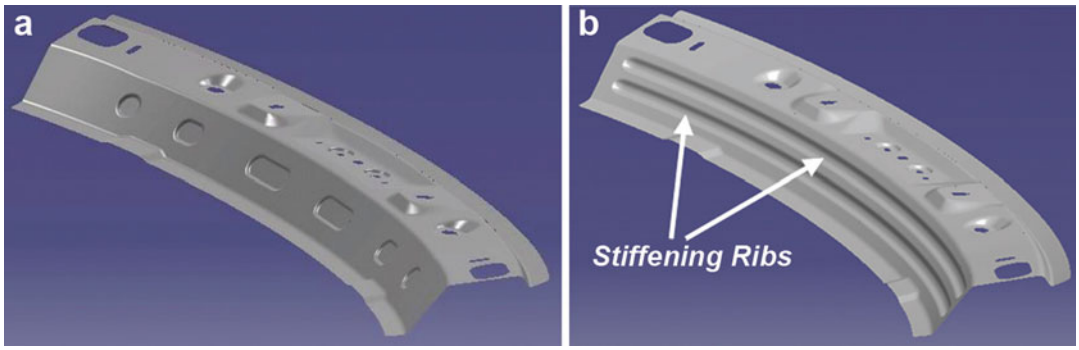
The objective for a forming process designer is to obtain a final part shape $\mathbf{U}_{\text{part}}$ close enough to the desired part shape $\mathbf{U}_{\text{design}}$. If the tolerance of part shape deviation is ε, we want to have the following relationship satisfied:

$$\left\|\mathbf{U}_{\text{die}} + \mathbf{U}_{\text{sb}} - \mathbf{U}_{\text{design}}\right\| \leq \varepsilon \tag{3}$$

In the process of tooling design, $\mathbf{U}_{\text{design}}$ and ε are considered as given. Two types of approaches have been developed to satisfy Eq. 3. The first approach is often referred as *springback control* where the objective is to minimize or even eliminate springback $\mathbf{U}_{sb}$ while keeping the die geometry $\mathbf{U}_{\text{die}}$ identical to the design shape $\mathbf{U}_{\text{design}}$. This is generally the preferred approach if it is possible or economically feasible. The second approach is to recognize the fact that there will always be a certain amount of springback. The focus is instead to adjust the tooling geometry to an appropriate shape that satisfies Eq. 3. In such cases $\mathbf{U}_{\text{die}}$ is no longer identical to $\mathbf{U}_{\text{design}}$. This approach is called *springback compensation* or *die compensation*.

Springback Control

The methods for springback control fall broadly into two categories. The first method attempts to minimize springback through part design. Since



Springback, Fig. 3 An example of springback control through part design (Reproduced from Z.C. Xia "Recent Advances in Springback Technology", NUMIFORM'2013, the 11th International Conference on Numerical Methods in Industrial Forming Processes, Shenyang, China, July 8, 2013. In public domain). (**a**) Original design and (**b**) modified design for springback control

the magnitude of springback is inversely proportional to the stiffness of a part, springback can be reduced if a part can be made stiffer either locally or as a whole. This is often achieved by incorporating stiffening ribs, beads, and other features to reinforce local or global stiffness without altering the functionality of the part (Fig. 3). Maximal springback reduction effects are accomplished by placing these reinforcements in target locations in accordance to the anticipated springback modes. Tightening radii and eliminating open sections are also effective methods for springback control.

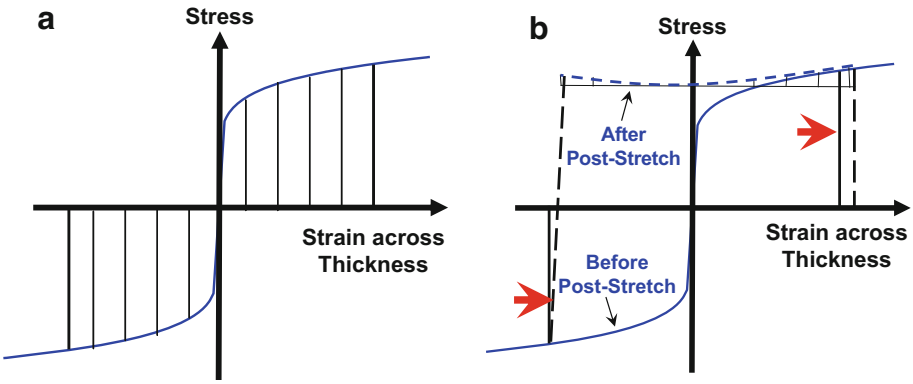
The second category for springback control is by optimizing the forming process to reduce springback. One of the most popular methods is the so-called post-stretch or ShapeSet method as originally proposed by Ayres (1984). The concept can be illustrated as follows:

For a part under simple bending as illustrated in Fig. 1, the amount of springback is determined by the bending moment before unloading,

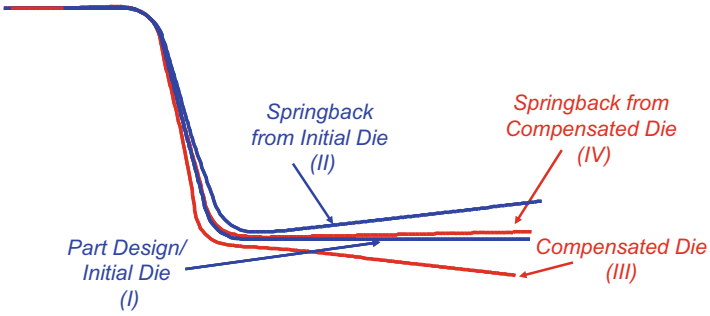
$$M = \int_{-\frac{t}{2}}^{\frac{t}{2}} \sigma z \; dz \tag{4}$$

Now suppose a small axial tensile strain is applied to the part. The top surface will continue its plastic loading in tension, and its tensile stress will increase by a small amount due to material hardening. However, the bottom surface will unload from its compressive condition and continue to load into tensile regime. The deformation is initially elastic and then enters into reverse plastic loading in tension. The change in stress distribution across sheet thickness due to post-stretching is illustrated in Fig. 4. The resulting bending moment, according to Eq. 4, is reduced significantly (in Fig. 4b). Two to 4% of post-stretch strain is often effective in minimizing springback.

Several mechanisms have been proposed by automakers to implement the concept of *ShapeSet*. Some of them are patented processes. We will use the process developed by Owens (2001) here as an example. In the first step, drawbeads are not used, and the punch penetration is less than that of the desired part depth. The metal is allowed to flow easily into the die to obtain a preliminary shape. Springback is severe at the end of this step. In the second step, the partially deformed part is placed in a die of the desired final component shape. In this step, the metal is tightly clamped by drawbeads or lockbeads, and stretching is now the predominant mode of deformation. The resulting springback from the second step is much lower due to the stretching. As seen, different restraining force histories during punch stroke impose different distributions and ratios of membrane stretching forces and bending moments. Bending will dominate when material flowing toward the die cavity is allowed; conversely, stretching will dominate when material flow into the die cavity is restrained.



Springback, Fig. 4 Stress distributions through sheet thickness. (**a**) After bending and (**b**) after post-stretch



Springback, Fig. 5 Illustration of the concept of springback compensation

The ShapeSet process, although providing good results, is a "passive" system where the restraining force is altered during the process due to the two-step variable die nature of the forming process and may not be applied in general.

There are several other methods developed over the years to control springback by optimizing the forming process. Liu (1988) has proposed a method to reduce springback using different histories of restraining force during a forming cycle through the imposition of binder force paths. He considered springback as a problem in relation to the restraining force and developed an "intermediate restraining" process to form high-quality flanged steel channel in a single operation. The in-process variation of binder force provided tensile pre-loading or post-loading on the formed part in order to significantly reduce springback. However, a tight control during production is required, making this process sensitive to any variations in manufacturing conditions such as friction coefficient. Sunseri et al. (1996) developed and implemented a closed-loop algorithm for binder force control to make the forming process more robust and repeatable. In their strategy, a punch force trajectory is introduced as the target curve instead of using a binder force trajectory. The approach is becoming more promising with the wide adoption of servo press in recent years.

Springback Compensation

The second approach for achieving dimensional accuracy is the so-called springback compensation or die compensation. Instead of trying to minimizing springback itself, the focus is to adjust the tooling shape in such a way that the part shape after springback meets the design target. The concept is best explained with the illustration in Fig. 5. Suppose the geometry of the part design is used as the surface of the initial forming die (surface I in Fig. 5), the part will be at position II after forming and springback. Now if we modify the die surface to position III, the formed part after springback will end up at position IV instead, very close to the desired geometry I.

Two key components are required for the method of springback compensation to succeed: the magnitude of springback as an input and an effective algorithm to compute the compensated tool surface. There are two ways to obtain the magnitude of springback. One is through physical tryout where forming trials are conducted and springback is measured. It requires the physical build of forming tools, which are often costly and time-consuming. The alternative way is to predict springback amount with numerical simulation of forming processes. Although recent advances in simulation technology made this approach more feasible, the accuracy of springback prediction is highly dependent on the complexities of the formed part and the forming process and on the abilities to accurately model the friction behavior and the mechanical behavior of the sheet material.

The development of algorithms for effective springback compensation requires a comprehensive understanding of springback mechanics and having the capability to compensate springback completely for even complex parts. Three approaches are generally employed.

1. Displacement method. In this method, the tool geometry is modified in the opposite direction of the springback of the formed part by either the same magnitude or in predetermined proportion (Wu 1997). This has been the most traditional approach on tryout floors since the beginnings of sheet forming. One-to-one compensation is usually adopted unless the operators have enough experience to believe otherwise. The underlying assumptions for this approach are that (1) springback response is linear to the geometric changes of the forming die and (2) geometric changes in the die only affect springback in the local area. This is not necessarily true if the magnitude of springback becomes larger or the part geometry is complex. As a result, the displacement method works best if springback is relatively small.
2. Mechanics method. In this method, the required compensation of the die surface is computed from mechanics principles, usually through the manipulation of residual stresses. One of the popular approaches is the "spring forward" concept, where the direction of residual stresses in the part after forming is first reversed, letting it go free, and "unloading" elastically (Karafillis and Boyce 1992). Because of the change of the direction, the part will "spring forward" instead of "springback" and will reach a "spring forward" position in self-equilibrium. This new position will be taken as the geometry for the compensated die in the hope that a part formed to this new position will "springback" to the design intent. This method works well even for complex parts with relatively large springback.
3. Optimization method. If a tool surface can be parameterized, and the springback can be reliably predicted, an iterative optimization algorithm can be developed with the tool surface parameters as design variables and the part design geometry as the target. This method works best for relatively simple geometries and for theoretical studies.

In practice, an iterative process is often needed for springback compensation. Each modified die will bring the formed part closer to its design target.

References

Ayres RA (1984) Shape set: a process to reduce sidewall curl Springback in high strength steel rails. J Appl Metalwork 3:127–134

Karafillis AP, Boyce MC (1992) Tooling design in sheet metal forming using springback calculations. Int J Mech Sci 34:113–131

Liu YC (1988) The effect of restraining force on shape deviations in flanged channels. J Eng Mater Tech ASME Trans 110:389–394

Owens HW (2001) Flow lock bead control apparatus and methods for drawing high strength steels. US Patent 6276185 B1

Sunseri M, Cao J, Karafillis AP, Boyce MC (1996) Accommodation of springback in channel forming using active binder control. Trans ASME J Eng Mater Technol 118:426–435

Wu L (1997) Tooling mesh generation technique for iterative FEM die surface design algorithm to compensate for springback in sheet metal stamping. Eng Comput 14:630–618

SPRT

► Self-Propelled Rotary Tool

Sputter Deposition

► Ion Beam Machining

Sputtering

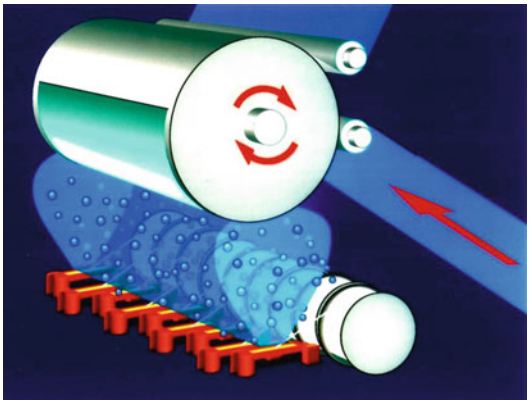
Günter Bräuer
Fraunhofer-Institute for Surface Engineering and Thin Films IST, Braunschweig, Germany

Definition

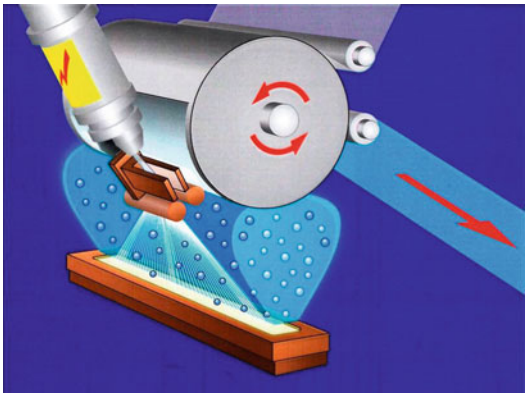
Coating technologies are used to modify mechanical, chemical, electrical, or optical properties of surfaces. A distinction can be made between "thick" and "thin" film technology. "Thin films" in general have thicknesses less than 10 µm, but the demarcation line is not so clear-cut.

The most important processes for thin-film formation are physical vapor deposition (PVD) processes. Basic PVD processes are evaporation and sputtering. The coating material is solid; thin films are deposited by solid-liquid-vapor-solid phase transition in the case of evaporation and solid-vapor-solid transition in the case of sputtering.

Figures 1 and 2 show the principles of thermal and electron beam evaporation. Materials with a low melting point (many metals) are evaporated from graphite boats by resistive heating. For materials with a higher melting point (e.g., metal oxides), electron beam evaporation is commonly used. Evaporation is a fast process with deposition rates of 500–5000 nm/s for metals and 100–1,000 nm/s for oxides. Film quality and adhesion to the substrate may suffer from the low energy of evaporated particles (0.2–0.5 eV). Modern evaporation processes therefore make use of an additional plasma. The evaporated particles



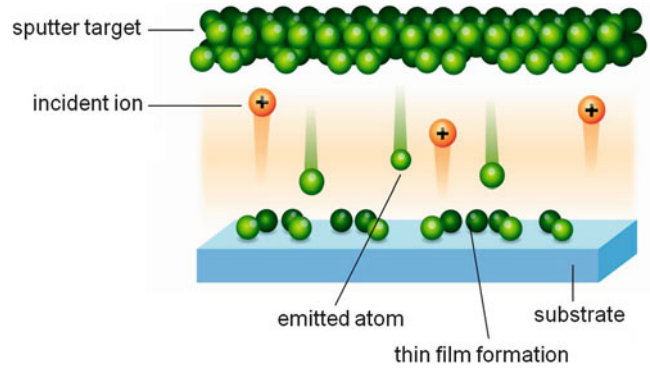
Sputtering, Fig. 1 The principle of web coating by thermal evaporation



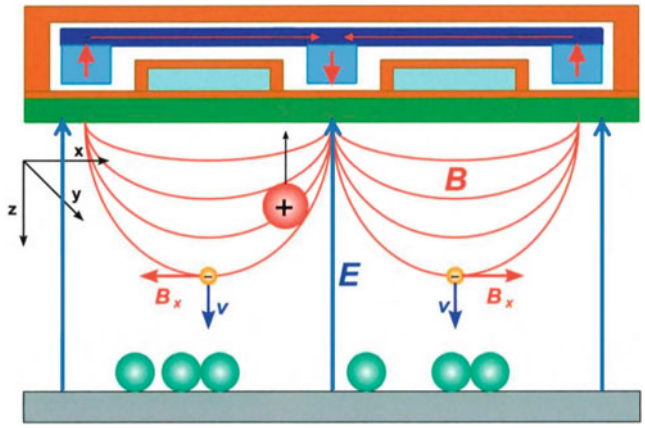
Sputtering, Fig. 2 The principle of web coating by electron beam evaporation

crossing the plasma zone are activated and ionized and consequently can form a much denser film.

Plasma is the key for modern thin-film deposition processes. It is an ionized gas where basic physical and chemical reactions are used to etch or coat surfaces. In sputter processes a plasma is ignited by applying an electric field between two electrodes in a high vacuum (glow discharge). Ions (preferably argon) generated by collisions with fast electrons transfer their momentum to the atoms of a target. Atoms close to the surface are ejected and form a thin film on an opposite substrate (see Fig. 3). Furthermore, energetic ions also cause the ejection of secondary electrons from the target surface. These are necessary to sustain the plasma discharge. If a reactive gas (e.g., oxygen or nitrogen) is added to the argon,



Sputtering, Fig. 3 The principle of sputtering



Sputtering, Fig. 4 Planar magnetron cathode

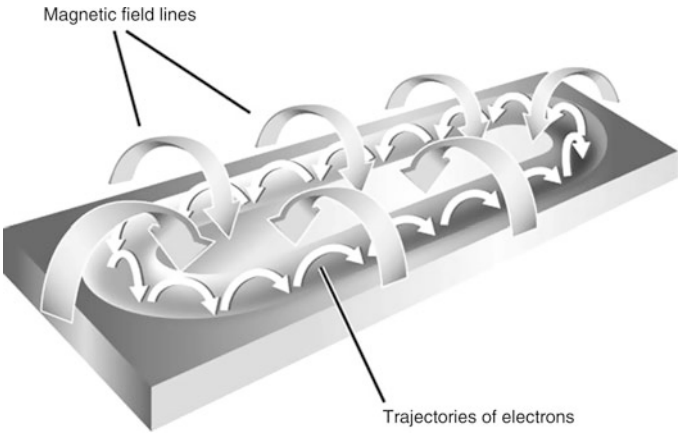
metal oxides, nitrides, or other compound films can be deposited. Compared with other deposition techniques, sputtering is a rather slow process (deposition rate in the range of several nm/s), but as the energy of sputtered atoms is around ten times higher than that of evaporated particles, dense and smooth films of high quality can be obtained.

Theory and Application

The sputter process is described by the sputter yield Y, which is the number of emitted target atoms divided by the number of incident ions. Important parameters determining Y are the ion energy, the mass numbers of ion and target atom, the surface binding energy of the target material, the angle of ion incidence, and the target temperature (Sigmund 1969; Oechsner 1975).

The industrial use of sputtering for thin-film deposition until about 1980 was limited to so-called sputter diodes with low plasma (ion) density and discharge voltages around 1,000–3,000 V. The film growth rates (deposition rates) of such a system are in general lower than 0.1 nm/s.

The planar magnetron cathode was introduced in 1974. Magnetron sputtering today is the key process for manufacturing of innovative products like all kinds of discs for data storage and entertainment, flat displays, or thin-film solar cells. The sputter cathode is equipped with a permanent magnetic field according to Figs. 4 and 5. Electrons are trapped in front of the target by the Lorentz force (magnetic bottle). Compared to diode sputtering, they generate much more ions, consequently deposition rates could be increased by a factor of 10 or more, and the discharge voltage decreases to

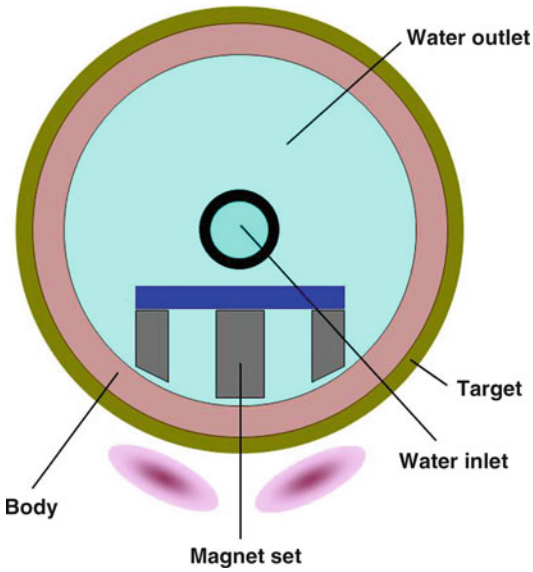


Sputtering, Fig. 5 Magnetic field lines and electron trajectories for a planar magnetron

several hundreds of volts. The main advantages of magnetron sputtering are as follows:

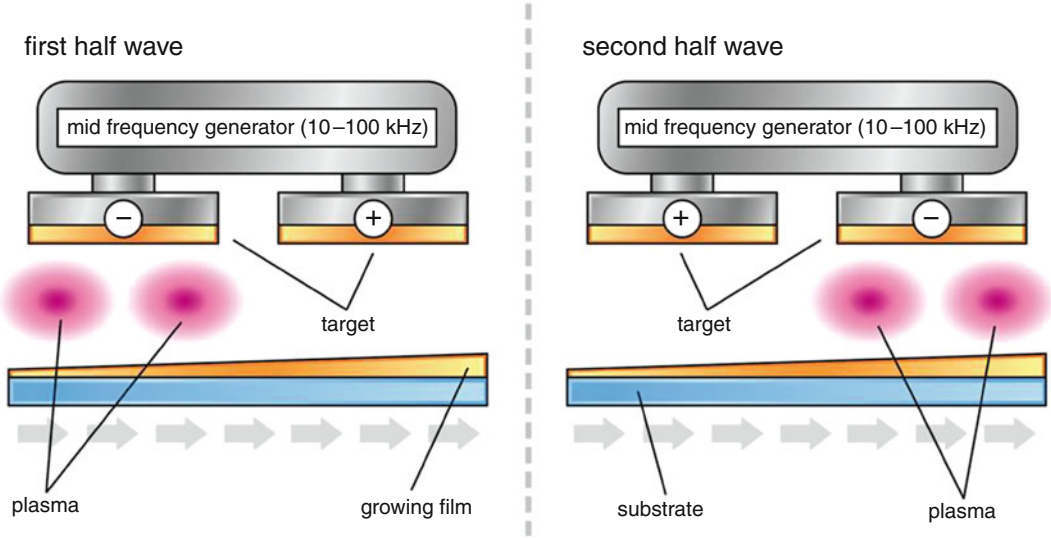
- Low plasma impedance and thus high discharge currents from 1 to 100 A (depending on cathode length) at typical voltages around 500 V
- Deposition rates in the range from 1 to 10 nm/s
- Low thermal load to the substrate
- Coating uniformity in the range of a few percent even for several meters long cathodes
- Easy to scale up
- Dense and well-adherent coatings
- Large variety of film materials available (nearly all metals and compounds)
- Broadly tunable film properties

Magnetron sputter processes using planar targets and driven by direct current plasma generation for many years have suffered from serious drawbacks. First of all, a non-homogeneous ion current distribution across the target surface determined by the geometry of the magnetic field causes a poor target material utilization (often less than 30%). Furthermore, various applications require high insulating oxides (SiO_2, Al_2O_3, TiO_2) or nitrides (AlN, SiN). During the reactive sputter process, insulating films are then formed on all inner surfaces of the deposition chamber, in particular in the non-eroding target areas. Arcing occurs, and the process parameters may drift due to the loss of an effective anode (Scholl 1993).



Sputtering, Fig. 6 Rotatable magnetron

Figure 6 shows the principle of a cylindrical magnetron ("C-MAG"), where a tubular target is rotating around a fixed magnet array. Redeposition zones on the target surface can almost be minimized, and the material utilization is around 90% (McKelvey 1982; Wright and Beardow 1986). Another milestone in reactive sputtering was the development of pulsed plasmas. In Fig. 7 a double magnetron arrangement is outlined. Such systems are typically powered by mid frequency in the range of 10–100 kHz. The plasma is switched between the two electrodes; each of them alternatively



Sputtering, Fig. 7 Pulse magnetron sputtering

serves as a sputter cathode or an anode. The process stability for deposition of dielectric films is substantially increased. Pulse magnetron sputtering also results in denser and smoother coatings (Jäger et al. 1996; Bräuer et al. 2010).

High-power impulse magnetron sputtering (HIPIMS) is a rather novel technique where very short energetic pulses (typical 1,000 A at 1,000 V) are applied to the target. The power density may be about 50 times higher than in standard sputter processes. In such HIPIMS discharges, 50–90% of the sputtered atoms are ionized (about 1% for conventional sputtering); they may form layers with outstanding properties (Kouznetsov et al. 1999). Potential applications are:

- Hard coatings for tribology
- Coatings for precision optics
- Coatings for thin-film solar cells, heating, and antifog

Today magnetron sputtering serves various branches. Tools and components for vehicles or general mechanical engineering need coatings with low wear and low friction. Energy saving by architectural glazings or solar energy conversion would not work without thin films, and each innovative component for data handling or entertainment like hard disc, optical discs for storage of music or movies, and flat displays is based on the physical properties of sputtered films. Besides the race for better coatings, of course the improvement of process efficiency is an ongoing challenge. In a sputter process, only around 2% of the ion energy are converted to sputtering. Much higher sputter yields are desirable, and a promising attempt is sputtering from hot targets.

References

Bräuer G, Szyszka B, Vergöhl M, Bandorf R (2010) Magnetron sputtering – milestones of 30 years. Vacuum 84(12):1354–1359

Jäger S, Szyszka B, Szczyrbowski J, Bräuer G (1996) Comparison of transparent conductive oxide thin films prepared by AC and DC reactive magnetron sputtering. Surf Coat Technol 98(1–3):1304–1314

Kouznetsov V, Macák K, Schneider JM, Helmersson U, Petrov I (1999) A novel pulsed magnetron sputter technique utilizing very high target power densities. Surf Coat Technol 122(2–3):290–293

McKelvey HE (1982) Magnetron cathode sputtering apparatus. US patent 4,356,073 A

Oechsner H (1975) Sputtering – a review of some recent experimental and theoretical aspects. Appl Phys 8(3):185–198

Scholl RA (1993) A new method of handling arcs and reducing particulates in DC plasma processing. In: 36th annual technical conference proceedings of the society of vacuum coaters, pp 405–408

Sigmund P (1969) Theory of sputtering. I. Sputtering yield of amorphous and polycrystalline targets. Phys Rev 184(2):383–416
Wright M, Beardow T (1986) Design advances and applications of the rotatable cylindrical magnetron. J Vac Sci Technol A 4(3):388–392

Stability

Dirk Biermann[1] and Tobias Surmann[2,3]
[1]Institut für Spanende Fertigung, Technische Universität Dortmund, Dortmund, Germany
[2]Premium AEROTEC GmbH, Varel, Germany
[3]Mechanical Engineering, Technical University of Dortmund ISF, Dortmund, Germany

Definition

The term *stability* as used in this entry is a property of an equilibrium of a dynamic system. There are various definitions of stability which describe how a dynamic system being at an equilibrium reacts on a small disturbance (Plaschko and Brod 1995). In machining science, the Lyapunov stability and the asymptotic stability are of high interest. Consider the continuous dynamic system:

$$\dot{x} = f(x(t)),\ x(0) = x_0. \tag{1}$$

We suppose that the system has an equilibrium at

$$x(t) = x_e. \tag{2}$$

Lyapunov Stability

The above equilibrium is called Lyapunov stable, if

$$\forall \epsilon > 0 \quad \exists \delta = \delta(\epsilon) > 0 \tag{3}$$

such that

$$|x_0 - x_e| < \delta \Rightarrow \forall t \geq 0\, |x(t) - x_e| < \epsilon. \tag{4}$$

This means that solutions starting within a range of δ from the equilibrium will forever stay within a range of ϵ to the equilibrium.

Asymptotic Stability

An equilibrium is called asymptotically stable if it is Lyapunov stable and if $\exists\ \delta$ such that

$$|x_0 - x_e| < \delta \Rightarrow \lim_{t \to \infty} (x(t) - x_e) = 0. \tag{5}$$

Asymptotic stability means that solutions which start close enough to the equilibrium not only remain close enough but also converge to the equilibrium.

Theory and Application

Discrete Dynamic Systems

For simplicity, we will explain the fundamentals of stability analysis by usage of discrete dynamic systems. A general discrete dynamic system can be written as an iteration.

$$\vec{x}_{i+1} = G\left(\vec{x}_i\right), \tag{6}$$

where $\vec{x}_i$ is the state vector, which can be of an arbitrary dimension. The Function G maps the state vector of the i-th point in time on the state vector of the following point in time. For instance, the state of a freely vibrating harmonic oscillator is well defined by its current elongation and velocity (Ewins 2000; He and Fu 2001).

Fixed Points

Of high importance for machining processes are fixed points. A fixed point $\vec{x}_f$ fulfills the fixed point equation

$$\vec{x}_f = G\left(\vec{x}_f\right). \tag{7}$$

Hereby, a fixed point takes the role of an equilibrium.

Stability Analysis

The stability of a fixed point can be determined by using the linearization approach. By adding a distortion $\vec{y}_i$ to the fixed point $\vec{x}_f$ and by definition of

$$\vec{x}_i = \vec{x}_f + \vec{y}_i \tag{8}$$

we get the dynamic system

$$\vec{x}_f + \vec{y}_{i+1} = G\left(\vec{x}_f + \vec{y}_i\right). \tag{9}$$

The linearizaton of the dynamic system G with the first two terms of the Tailor series results in

$$G\left(\vec{x}\right) = G\left(\vec{x}_f\right) + J_{\vec{x}_f}(G) \cdot \left(\vec{x} - \vec{x}_f\right), \tag{10}$$

where $J_{\vec{x}_f}(G)$ is the Jacobian matrix of G evaluated at the fixed point. This linearization of our dynamic system leads to

$$\vec{x}_f + \vec{y}_{i+1} = G\left(\vec{x}_f\right) + J_{\vec{x}_f}(G) \cdot \left(\vec{x}_f + \vec{y}_i - \vec{x}_f\right). \tag{11}$$

Because of the fixed point equation, $\vec{x}_f$ vanishes from the equation, and we get a new dynamic system describing the progression of the distortion $\vec{y}_i$:

$$\vec{y}_{i+1} = J_{\vec{x}_f}(G) \cdot \vec{y}_i. \tag{12}$$

Since for an asymptotically stable fixed point any distortion (which is small enough) must vanish over time, the fixed point $\vec{x}_f$ is stable if the absolute values of all Eigenvalues of the Jacobian matrix of G are less than 1.

Simple Example

A simple example is given by the discrete dynamic system:

$$\vec{x}_{j+1} = G\left(\vec{x}_j\right) = \begin{pmatrix} k\cos(\alpha) & -k\sin(\alpha) \\ k\sin(\alpha) & k\cos(\alpha) \end{pmatrix} \vec{x}_j \tag{13}$$

The matrix is a rotation matrix which rotates the two-dimensional vector $\vec{x}_j$ by the angle α around the origin. Simultaneously, the vector becomes scaled by the factor k. Obviously, this system has a fixed point at $\vec{x}_f = (0;0)$.

In order to analyze the stability of the fixed point dependently on the parameters k and α, we need the Jacobian matrix of the system at the fixed point $\vec{x}_f$, which in this specific case is the same matrix as already used in Eq. 13:

$$\mathbf{J}_{\vec{x}_f}(G) = \begin{pmatrix} k\cos(\alpha) & -k\sin(\alpha) \\ k\sin(\alpha) & k\cos(\alpha) \end{pmatrix}. \tag{14}$$

The eigenvalues of this matrix are

$$\sigma_0 = k\cos(\alpha) - ik\sin(\alpha) \tag{15}$$

$$\sigma_1 = k\cos(\alpha) + ik\sin(\alpha) \tag{16}$$

from which we get the absolute values

$$|\sigma_0| = \sqrt{(k\cos(\alpha))^2 + (-k\sin(\alpha))^2} = \sqrt{k^2\,cos(\alpha)^2 + k^2\,sin(\alpha)^2} = k \tag{17}$$

and

$$|\sigma_1| = \sqrt{(k\cos(\alpha))^2 + (-k\sin(\alpha))^2} = \sqrt{k^2\,cos(\alpha)^2 + k^2\,sin(\alpha)^2} = k. \tag{18}$$

For $k < 1$ the fixed point $\vec{x}_f$ is stable, and for $k > 1$ it is unstable.

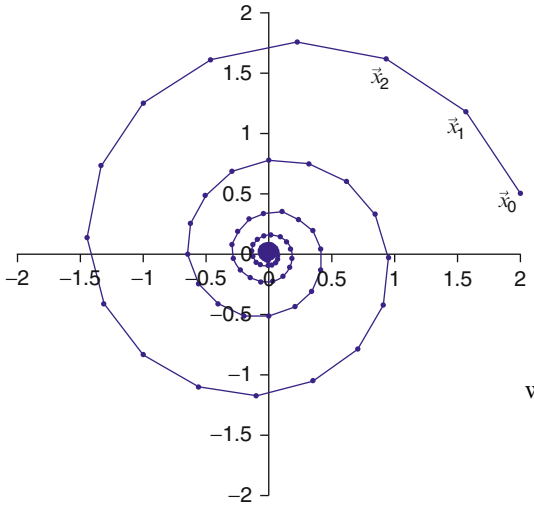
For $k = 1$ the eigenvalues lie on the unit circle. Since the parameter k determines whether the fixed point is stable or not, it is called the critical parameter. Analog to that, the value $k = 1$ is called the critical (or bifurcation) value of the parameter k. Figures 1 and 2 show the orbits of the dynamic system for a stable and an unstable case, respectively. Nevertheless, in both cases $\vec{x}_f = (0;0)$ is a fixed point.

Cycles and Hopf Bifurcation

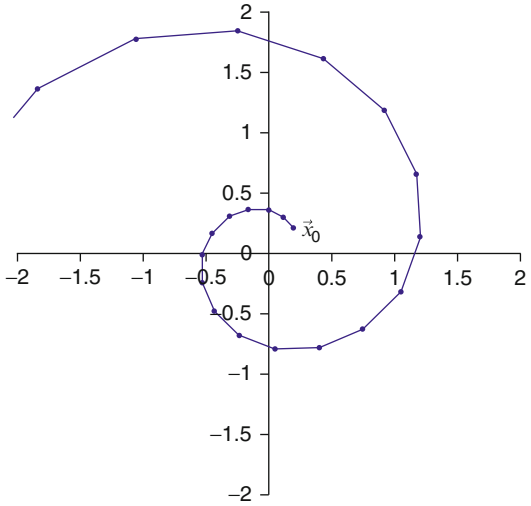
The orbit (i.e., the trajectory) of a mapping

$$\vec{x}_{j+1} = G\left(\vec{x}_j\right) \tag{19}$$

is defined as the sequence of points $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$. A k-periodic orbit (k-cycle) exists, if there is a sequence of k (k ¿ 1) different points having the property



Stability, Fig. 1 For $k = 0.95$ and $a = 0.4$, the orbit of the dynamic system starting at the point $\vec{x}_0 = (2;0.5)$ converges to the stable fixed point which lies in the origin



Stability, Fig. 2 For $k = 1.11$ and $a = 0.4$, the orbit of the dynamic system starting at the point $\vec{x}_0 = (0.2;0.2)$ diverges from the now unstable fixed point

$$\vec{x}_j = G^j\left(\vec{x}_0\right) \neq \vec{x}_0 \text{ for } 1 \leq j < k \text{ and } \vec{x}_0 = G^k\left(\vec{x}_0\right). \tag{20}$$

The stability of such a cycle again depends on the eigenvalues of the Jacobian matrix of G.

As an example for different types of orbits and for a Hopf bifurcation, we consider the dynamic system

$$\vec{x}_{j+1} = \begin{pmatrix} x_{j+1} \\ y_{j+1} \end{pmatrix} = \begin{pmatrix} (1 - hD)x_j - h\theta y_j \\ h\theta x_j + (1 - hD)y_i \end{pmatrix} = G\left(\vec{x}_j\right) = \begin{pmatrix} G_x\left(\vec{x}_j\right) \\ G_y\left(\vec{x}_j\right) \end{pmatrix} \tag{21}$$

with

$$D = x_j^2 + y_j^2 - \rho. \tag{22}$$

This system has a parameter vector of three components, (h, ρ, θ). Figure 3 shows the orbits of the system for two different values of θ while the other parameters have been kept constant. For (h, ρ, θ) = (0.07, − 0.4, 2.5), the system has a stable fixed point in the origin (Fig. 3, left), whereas for (0.07, − 0.4, 4), the stable fixed point has turned into an unstable fixed point with a surrounding stable cycle (Fig. 3, right). By deriving the Jacobian matrix of the system and computing the eigenvalues of it, it turns out that the critical value of θ, at which this happens, is approximately at $\theta \approx 3.357$. Such a sudden change in the behavior of dynamic systems is called bifurcation. In this specific case where a stable equilibrium turns unstable with a surrounding stable cycle is a kind of Hopf bifurcation.

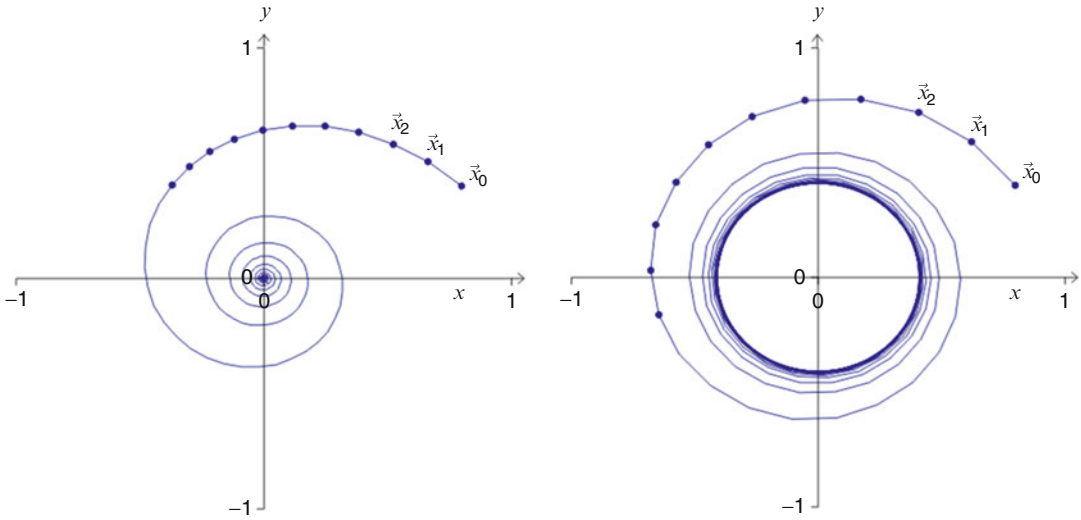
Application on Machining Processes

A Simple Turning Process

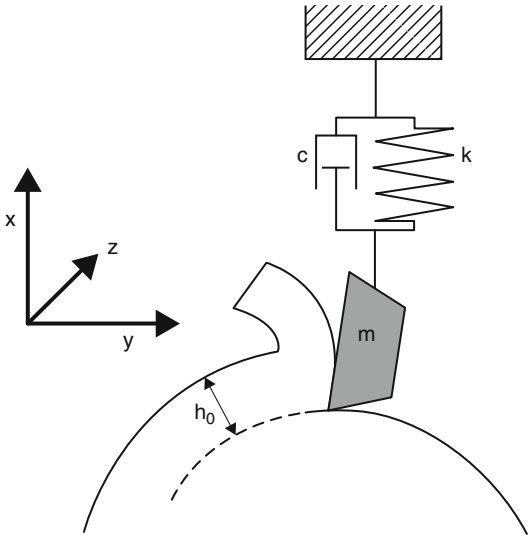
The differential equation of a simple turning process, as shown in Fig. 4, reads as

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = b \cdot k_n \cdot (h_0 - x(t) + x(t - T)). \tag{23}$$

Here we consider that the cutting tool can only vibrate in the normal direction of the cut, where $x(t)$ is the current deflection of the tool. The variable b is the width of cut, k_n is the specific normal force of cut, and h_0 is the intended uncut chip thickness. Finally, T is the time needed for one revolution of the workpiece.



Stability, Fig. 3 When changing the parameter θ from 2.5 to 4, the fixed point of the dynamic system in the origin changes from stable (left) to an unstable fixed point with a surrounding stable cycle (right)



Stability, Fig. 4 Model of a simple turning process

This equation is a delayed differential equation (DDE) which can be found in most of the machining processes where a tool moves over the same surface more than once (Insperger and Stpn 2011). The dynamic system underlying the above DDE has a fixed point solution or an equilibrium at

$$x_f = x(t) = \frac{b \cdot k_n \cdot h_0}{k} = const. \tag{24}$$

This equilibrium becomes unstable at a certain b, where a sinusoidal vibration starts growing up, which is called *chatter vibration*. The stability limit, i.e., the maximum width of cut, where the equilibrium is stable, can also be found by a stability analysis.

Therefore, we assume that any distortion results in a damped or excited sinusoidal vibration with a starting amplitude of a and with an angular velocity of ω:

$$x(t) = x_f + a \cdot e^{(-\gamma + i\omega)t}. \tag{25}$$

Inserting this into the original DDE results in

$$\begin{aligned} &m \cdot (-\gamma + i\omega)^2 + c \cdot (-\gamma + i\omega) + k \\ &\quad + (b \cdot k_n) \\ &\quad = (b \cdot k_n) \cdot e^{\gamma T} \cdot e^{-i\omega T}. \end{aligned} \tag{26}$$

Since at the stability limit, the vibration has neither to vanish nor to excite we set γ to 0 and get

$$-m\omega^2 + i\omega + k = b \cdot k_n \cdot \left(e^{-i\omega T} - 1\right). \tag{27}$$

This is a complex equation which can be resolved for b, which is then the limiting width of cut b_{lim}, below which the equilibrium is stable:

$$b = b_{lim} = -\frac{m\sqrt{\left(\omega^2 - \omega_0^2\right)^2 + 4\gamma_0^2\omega^2}}{2k_n \cos(\varphi)}. \tag{28}$$

Hereby,

$$\omega_0 = \sqrt{\frac{k}{m}}, \tag{29}$$

$$\gamma_0 = \frac{c}{2m} \tag{30}$$

and

$$\varphi = \frac{\omega T - 3\pi}{2} \tag{31}$$

apply.

Since $\varphi(\omega)$ also is the phase shift between the tool vibration and the cutting force, additionally

$$\varphi(\omega) = \frac{\omega T - 3\pi}{2} = \arctan\left(\frac{2\gamma_0\omega}{\omega_0^2 - \omega^2}\right). \tag{32}$$

applies. From this we can express the time of one revolution $T(\omega)$ in dependency of the chatter frequency ω

$$T(\omega) = \frac{2\varphi(\omega) + 3\pi + j \cdot 2\pi}{\omega} \tag{33}$$

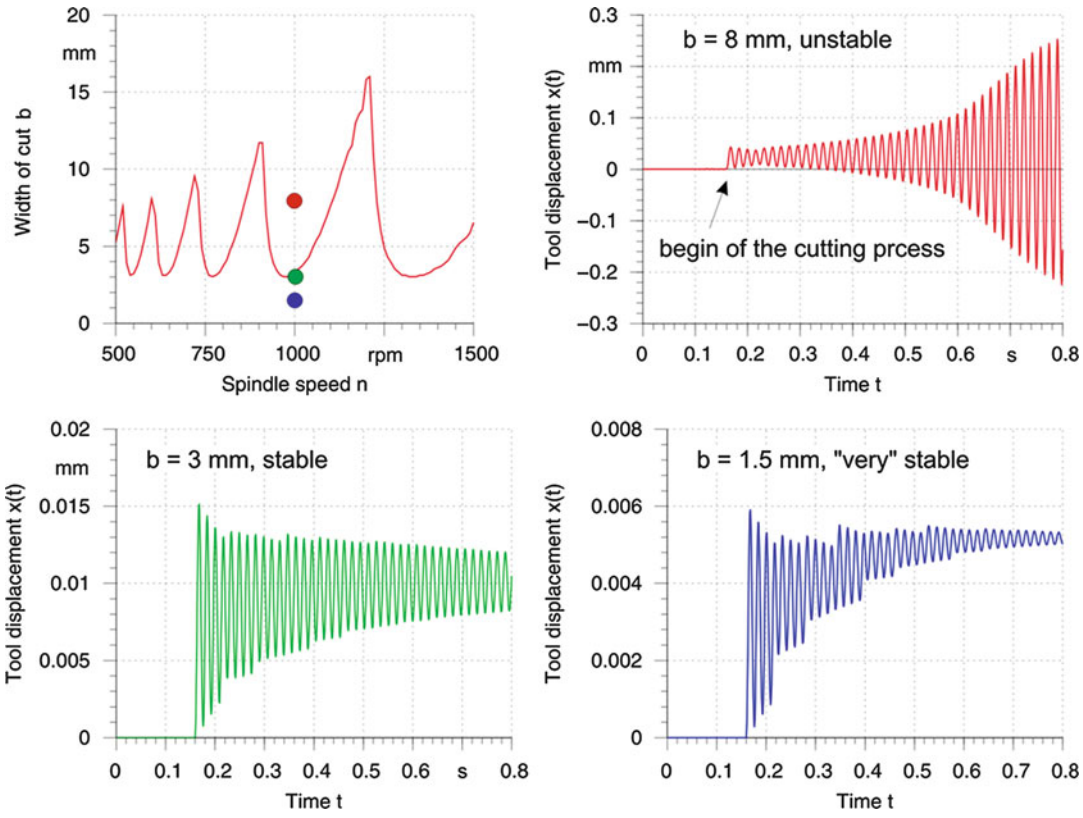
in order to obtain the spindle speed $n(\omega)$

$$n(\omega) = \frac{1}{T} \cdot 60\ \mathrm{min}^{-1} \tag{34}$$

related to the vibration (chatter) frequency ω.

We see that although the differential equation has an equilibrium independent of the width of cut, this equilibrium is only stable for $b < b_{lim}$. Since real machining processes are never free of distortions, the unstable equilibriums of these processes have unfortunately no practical meaning.

Figure 5 shows a stability diagram (Altintas 2000) and three different tool vibrations in x-direction (i.e., in direction of the normal of the cut). The first one (red) is from an unstable process,



Stability, Fig. 5 Trajectories of the tool in a simple turning process. In case of a width of cut above the stability limit, the static solution of the differential equation is unstable, and thus, the vibration amplitude forced by the begin of the cut grows over time (red). In case of a width of cut lower than the stability limit, the vibration vanishes over time and converges to an equilibrium (green and blue graphs)

where the vibration, which is invoked by the first contact between the tool and the work, grows over time. The green and blue vibrations depict stable turning processes where the vibration vanishes over time.

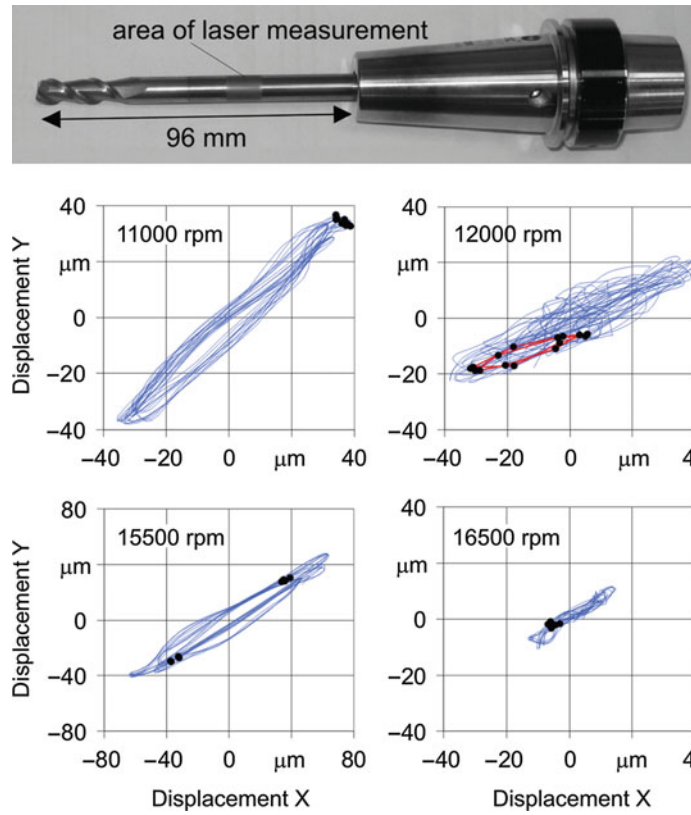
Milling

The milling process is a little bit more difficult because of the intermittent cut and the rotating coordinate system of the cutting edges. Figure 6 shows four different vibration plots measured at different spindle speeds in the x/y-plane of a peripheral milling process with low radial immersion. It appears that in the so-called stable milling at 11,000 rpm, the tool describes an elliptic cyclic orbit whereby the cycle is exactly repeated in each tooth feed. A stroboscopic sampling with the tooth-feed frequency results in a sharp defined point cloud (black points) which is called the Poincaré-section. Theoretically, this Poincaré-section should be exactly one point which applies to stable simulated tool vibrations. At 12,000 rpm, the vibration is no longer a cycle and the Poincaré-section changed to a cycle itself.

In milling with low radial immersions, a period-doubling, which is a particular type of tool vibration, may occur. Hereby, the Poincaré-section splits up into two sharp defined point clouds (here: 15,500 rpm).

An interesting feature of the Poincaré-section is that its location changes with the spindle speed (Insperger et al. 2006, Bachrathy et al. 2009). Here, at stable milling at 16,500 rpm, it has almost no displacement in the y-direction which denotes a very low surface error.

In simulations of the milling process, mostly a dynamic system which describes the Poincaré-section is constructed. Thus, a stability analysis of the Poincaré-section is conducted in order to find out whether a milling process is stable or not (Hartung et al. 2006). Actually, a milling process is called stable if the Poincaré-section is a stable fixed point. Since in case of a transition from



Stability,
Fig. 6 Measured vibrations of a milling tool at different spindle speeds (Al7075, $D = 8$ mm, $z = 2$, $a_e = 0.4$ mm, $a_p = 0.5$ mm)

stable to unstable milling the Poincaré-section changes from a stable fixed point to a stable cycle, the bifurcation from stable to unstable milling is also a kind of Hopf bifurcation.

Cross-References

- ▶ Chatter
- ▶ Damping
- ▶ Dynamics
- ▶ Structural Analysis
- ▶ Vibration

References

Altintas Y (2000) Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design. Cambridge University Press, New York

Bachrathy D, Insperger T, Stpn G (2009) Surface properties of the machined workpiece for helical mills. Mach Sci Technol 13(2):227245

Budak E, Altintas Y (1998) Analytical prediction of chatter stability in milling, part I: general formulation. J Dyn Sys Meas Control T ASME 120(1):2230

Ewins DJ (2000) Modal testing: theory, practice and application, 2nd edn. Research Studies Press, Baldock

Hartung F, Insperger T, Stpn G, Turi J (2006) Approximate stability charts for milling processes using semi-discretization. Appl Math Comput 174(1):5173. https://doi.org/10.1016/j.amc.2005.05.008

He J, Fu ZF (2001) Modal analysis. Butterworth Heinemann, Oxford

Insperger T, Stpn G (2011) Semi-discretization for time-delay systems: stability and engineering applications. Springer, New York

Insperger T, Gradisek J, Kalveram M, Stepan G, Weinert K, Govekar E (2006) Machine tool chatter and surface location error in milling processes. J Manuf Sci Eng 128(4):913920. https://doi.org/10.1115/1.2280634

Plaschko P, Brod K (1995) Nichtlineare dynamik, bifurkation und chaotische systeme [non-linear dynamics, bifurcation and chaotic systems]. Vieweg, Braunschweig (in German)

Schmitz TL, Smith KS (2009) Machining dynamics frequency response to improved productivity. Springer, New York

Stamping

- ▶ Embossing
- ▶ Deep Drawing
- ▶ Hot Forging

Statistical Mechanics

- ▶ Molecular Dynamics for Cutting Processes

Statistical Process Control

Marcello Colledani
Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy

Synonyms

Statistical Process Control (SPC); Statistical Quality Control (SQC)

Definition

An analytical approach developed to continuously improve the quality of the output of manufacturing processes and systems.

Extended Definition

SPC helps in detecting, identifying, and eliminating unpredictable sources of variability in the process. Moreover, it helps monitoring the process by issuing signals whenever deviations from in-control conditions are detected. Statistical control charts are the basic tools to implement SPC in manufacturing processes.

SPC plays a key role in Six Sigma quality control implementations. It has been successfully applied both to continuous manufacturing (e.g., chemical processes) and to discrete part manufacturing (e.g., automotive, semiconductor, mechanical component production). Their applications have now moved far beyond manufacturing into engineering, environmental science, biology, genetics, epidemiology, medicine, and finance.

SPC only deals with the monitoring of the process and the identification of nonrandom variability sources in the process output. SPC does not deal directly with the adjustment of process parameters to compensate process deteriorations

due to out-of-control conditions. Process adjustment (Del Castillo 2002) is instead the primary activity of engineering or automatic process control (EPC-APC). Recently researchers have advocated integrating SPC with EPC tools, which regularly change process inputs to improve process performance.

Theory and Application

History

Statistical process control's foundation was laid by Dr. Walter Shewhart, an engineer working in the Bell Telephone Laboratories in the 1920s who was conducting research on methods to improve quality and lower production costs. He developed the concept of control with regard to variation and proposed statistical process control charts as a simple way to determine whether the process is in or out of control. Dr. W. Edwards Deming built upon Shewhart's work and, during the World War II, focused on methods to improve quality in the manufacture of munitions and other strategically important products in the USA. After World War II, he took the concepts to Japan. There, Japanese industry adopted the concepts wholeheartedly. Dr. Deming is famous throughout Japan as a "God of Quality."

Dr. Joseph M. Juran later emphasized a more strategy- and planning-oriented approach to quality. He proposed quality planning, control, and improvement as three highly interrelated processes. The Juran Institute is still an active organization promoting the Juran philosophy and quality improvement practices. In the 1980s, the theories of Deming and Juran became the basis for the total quality management strategy which aims at implementing a quality improvement strategy involving all company levels. Ishikawa proposed the integrated use of seven basic tools, including statistical control charts, for quality monitoring and improvement in manufacturing processes. Today, SPC is used in manufacturing facilities worldwide.

Product Quality Characteristics

Every product possesses a number of elements that jointly describe what the customer thinks of as quality. These parameters are called *key product quality features or characteristics*. Product quality characteristics can be physical (e.g., the diameter of a hole, the roughness of a surface) or sensorial (e.g., the appearance of the product or the color). The subsequent process stages in the process chain typically modify the values of these quality characteristics determining the feature values of the final product delivered in output by the manufacturing system. For each key product quality feature, *specification limits* (Tolerances) are imposed by design to guarantee the functionality of the product during the use phase. In particular, for each key product quality feature, the lower specification limit (LSL) and the upper specification limit (USL) are defined. If the observed value of the quality feature is outside the specification limits, the product is nonconform and of poor quality.

Variability in Manufacturing Processes

Manufacturing processes are transformations that ideally provide perfectly predictable and stable results, but in practice are affected by variability. If a key quality characteristic (e.g., the diameter of a hole) is observed in multiple processed items, its value will not be identical, but will be a random variable following a statistical distribution. High variability in production processes leads to quality defects, lack of stability, and product inconsistency. Two different causes for variability can be defined in manufacturing processes, namely, *random and assignable causes*.

Random or common causes of variations are inherent in the process. These sources of variations are unavoidable and are due to slight differences in processing, such as differences in materials, workers, machines, tools, and other factors. A stable process exhibits only common causes of variation. The behavior of a stable process is predictable and consistent. The variability of the observed sample results can be attributed to a set of unidentifiable random causes, which do not seem to change over time. This process state is defined as *statistical control state*. A process is "in control" if each relevant quality measure (e.g., the mean, the variance, the fraction of nonconformity of the product) is in a state of statistical control.

Assignable or special causes of variations are not part of the process. These can be traced, identified, and eliminated. Examples of this type of variation are poor quality in raw materials, an employee who needs more training, or a deteriorating machine in need of repair. In each of these examples, the problem can be identified and corrected. Also, if the problem is allowed to persist, it will continue to affect the quality of the product. In the presence of special or assignable causes of variation, a process is unpredictable or inconsistent, and the process is said to be out of statistical control. In general, "out of control" may consist of deviations of process output mean, of process output variance, or both. An out-of-control condition generates an increase in the production of nonconforming products.

The main goal of statistical control charts is to identify the presence or the occurrence of special causes of variability in order to remove them and keep the process in control, according to the continuous improvement paradigm.

Remark: Manufacturing processes are endowed with advanced sensor systems that make it possible to gather and collect the specific values of *process variables* over time. For example, the temperature in a furnace, the cutting force in a material removal process, and the material flow velocity in a rolling process are data that can be nowadays collected directly from the process. Since a consistent deviation from in-control process conditions may reflect into a signal variation, observing the process characteristics by SPC tools is an effective way to identify assignable causes of variations. Therefore, statistical control charts can be applied to both product- and process-related data.

Statistical Control Charts

A *statistical control chart* is a statistical procedure that identifies out-of-control conditions as effected by special variability causes through the systematic analysis of the output of a process (Montgomery 2013; Alwan 2000).

A control chart is a graphical tool characterized by a central line (CL), a lower control limit (LCL), and an upper control limit (UCL). The basic procedure adopted in traditional control charts is the following:

- Observe a sample of $n \geq 1$ process output data (n is the sample size; if $n = 1$, the control chart is for individuals; if $n > 1$, it is for subgroups) at regular intervals in time (h is the sampling frequency, expressed in time between samples or parts between samples).
- For each sample, compute the interesting sample statistics V (e.g., the sample mean, the sample standard deviation, the sample range).
- Report the obtained sample value in the control chart and apply out-of-control identification rules.

Out-of-control identification rules can be simple, e.g., an alarm is issued if one point is outside the control limits, or more complex, e.g., an alarm is issued when more points in sequence are out of the control limits or when there are trends or cyclic patterns.

In general, the central limit of the control chart is located at the expected value of the sample statistics V, μ_V, and the control limits are located according to the characteristic parameter of the control chart K. For example, for an Xbar control chart:

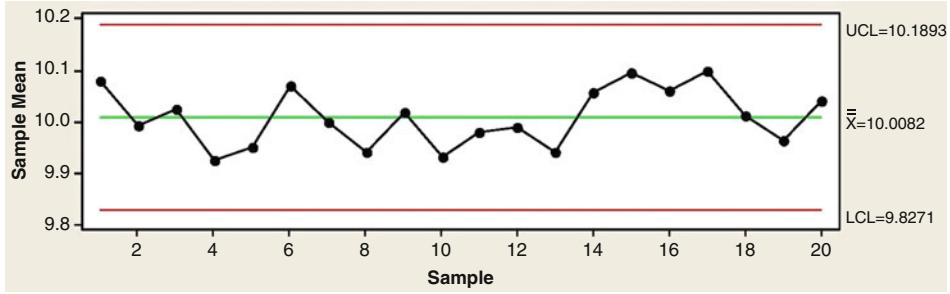
$$\mathrm{LCL} = \mu_V - K\sigma_V \tag{1}$$

$$\mathrm{CL} = \mu_V \tag{2}$$

$$\mathrm{UCL} = \mu_V + K\sigma_V \tag{3}$$

An example of an Xbar control chart is reported in Fig. 1. If K is set to 3, the Xbar chart is the traditional Shewhart chart.

Remark: There is a fundamental difference between the control limits of a control chart and the product specification limits. Product specifications are preset ranges of acceptable quality characteristics. For a product to be considered acceptable, its characteristics must fall within this preset range. Product specifications, or tolerance limits, are usually established by design engineers or product design specialists. Therefore, they are directly tied to the specific value of each single observation of the quality characteristics for every single item produced. On the other hand, control limits are defined by quality managers for the purpose of monitoring the process in-control conditions and are related to



Statistical Process Control, Fig. 1 Example of an Xbar control chart

the sample statistics monitored by the control chart. Therefore, in case of monitoring based on subgroups, specification limits should not be confused with control limits.

Conceptually, a control chart repeatedly performs a test of hypothesis on the processed data, with competing hypotheses:

- H0: the process is in control.
- H1: the process is out of control.

This test of hypothesis is subject to statistical errors, namely, type I and type II errors.

- Type I error happens when the alarm is issued while the process is in control. This leads to a false alarm that requires a search for an assignable cause that is not actually happening. This phenomenon is undesirable in practice since it increases the process downtime, requires human intervention, and generates user disaffection towards the quality control tool.
- Type II error happens when the alarm is not issued while the process is out of control. This leads to a prolonged out-of-control state and, ultimately, an increased production of nonconforming parts.

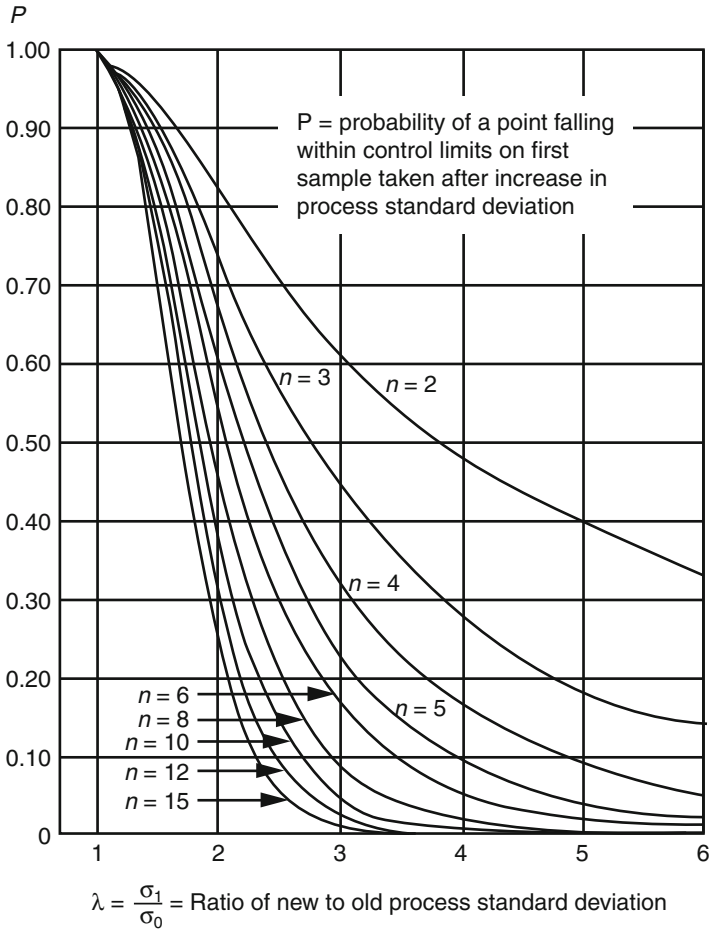
Type I error happens with probability α at any collected sample, while type II error happens with probability β at any collected sample; these probabilities are functions of the type of control chart adopted, of its parameters, and of the process characteristics (process output distribution and process deviations). Since they are very important for determining the performance of the control chart while monitoring the process, these probabilities have been extensively studied; for example, the so-called characteristic operating curve for a control chart for monitoring the sample range R is reported in Fig. 2. It can be noticed, given a process deviation to be detected (e.g., $\lambda = 2$), by increasing the sample size, the type II error decreases. Therefore, the out-of-control condition will be identified earlier. On the contrary, given a sample size (e.g., $n = 5$), a bigger process shift can be more easily identified by the control chart than a small shift.

Performance Measures

Since a control chart is affected by statistical error, its performance while monitoring a process output has to be defined in order to properly select the control chart's parameters for specific process conditions. The main performance measure of interest is the average run length (ARL), i.e., the average number of samples to be observed before the control chart issues an alarm. If a false alarm is considered, then the ARL_0 is computed, which is equal to the inverse of the type I probability, α. On the contrary, if the correct out-of-control alarm is considered, then the ARL_1 is considered, which is equal to the inverse of $(1-\beta)$. By multiplying the ARL by the time between samples h, the so-called average time to signal (ATS) is obtained. This performance measure is representative of the responsiveness of the control chart while identifying out of controls.

Control Chart Types

In the literature, different control chart types have been developed for monitoring processes featuring different characteristics.



Statistical Process Control,
Fig. 2 Characteristic operating curve for a control chart monitoring the range (Adapted from Montgomery 2013)

Control Charts for *Variables and Attributes*

A control chart for variables is used to monitor characteristics that can be measured and have a continuum of values, such as height, weight, or volume. A soft drink bottling operation is an example of a variable measure, since the amount of liquid in the bottles is measured and can take on a number of different values. A control chart for attributes, on the other hand, is used to monitor characteristics that have discrete values and can be counted. An attribute requires only a single decision, such as yes or no, good or bad, acceptable or unacceptable, or counting the number of defects. Both of them find wide application in the manufacturing industry.

Univariate and Multivariate Control Charts

Univariate charts monitor a single process output, while multivariate charts are developed to monitor simultaneously dependent or independent multiple process outputs. Examples of multivariate control charts include the χ^2 control chart, for known process variances, and the T^2 control chart, for unknown process variances.

Control Charts for Time *Auto-correlated Process Data*

Two approaches are typically followed in this case. The first approach involves the analysis of the underlying statistical model of the autocorrelation (e.g., ARIMA model) and the monitoring of the residuals with traditional control charts for independent observations. The second approach develops monitoring schemes directly on the correlated data (e.g., EWMA control charts).

Control Charts for *Small Shifts*

Since traditional control charts based on variables perform well only for consistent shifts of the

observed sample statistics, control charts for detecting small shifts have been developed. Different from traditional control charts, they do not use only the last sample process output observation but instead integrate either all the previous sample observations in the analysis (CUSUM charts) or the most recent observations (EWMA chart).

Bayesian Control Charts

The term "Bayesian" is used to denote any approach that adopts Bayes' theorem within the monitoring procedure. In particular, the literature on Bayesian control charting is characterized by two streams of research (Colosimo and Del Castillo 2006). The first assumes that all the information required to detect the degraded state is enclosed in the posterior probability that the process is out of control. The second stream of research adopts a more classical perspective, where Bayes' theorem is used to compute the posterior distribution of the quality characteristics of interest for the monitoring purpose.

Design of Control Chart Parameters and Sampling Plans

Due to the impact control charts have on the process and system effectiveness and profitability, considerable effort has been spent towards the proposal of optimal design of control chart parameters, including the sample size, the sampling frequency, and the control limit position. Among these approaches, economical design tries to select the best possible control chart parameters that minimize the long-term expected cost, including fixed and variable inspection costs, false alarm costs, lost production costs, and machine repair costs (Lorenzen and Vance 1986). Other design approaches are based on the adoption of Taguchi loss functions.

More recently, *adaptive control charts* have been proposed. Adaptive control charts feature design parameters that are not fixed but are allowed to vary over time. Depending on the current process state, a more reactive monitoring or a more relaxed monitoring can be set up by modifying the control chart and underlying sampling parameters. For example, the sampling frequency can be decreased if the process is shown to be in control for a long observation time, while it can be increased in the process ramp-up phase, where out-of-control conditions are more likely to occur or if the process shows an unknown out-of-control cause. This sampling schema is also called variable sampling plan.

Another family of adaptive sampling plan is *continuous sampling plans* (*CSP*). They allow switching between 100% screening inspections and sampling inspections when the process does not exhibit undesired behavior while restoring screening inspection if out-of-control behaviors are identified.

More recently, methods to design control chart parameters jointly with system logistics control parameters have been developed (Colledani and Tolio 2011). In particular, methods to jointly design control charts and buffer sizes in multi-stage production lines with remote inspections have been developed. They showed that improved overall system performance could be achieved with respect to isolated approaches.

Process Capability

A key aspect of statistical quality control is evaluating the ability of a production process to meet or exceed preset specifications. This is called *process capability*. Simply setting up control charts to monitor whether a process is in control does not guarantee a high process capability. To produce an acceptable product, the process must be capable and in control before production begins. The process capability refers to the ability of the process to produce an output characterized by key product quality features that are within the imposed specification limits. This determines the level of nonconformities produced by the process (Ishikawa 1976).

Process capability is measured by process capability indices. The most common index is the Cp which is computed as the ratio of the specification interval width to the width of the process variability, i.e., $(USL\text{-}LSL)/6\sigma$, where s is the standard deviation of the quality feature produced by the process, also called population. A process with small Cp, i.e., $Cp < 1$, is poorly capable, meaning that it will produce a consistent

amount of nonconforming items, while a process with $Cp > 1$ exceeds a minimal capability, although it can still deliver a fairly large fraction of nonconformities. Indeed, if $Cp = 1$, the fraction of conforming items produced is 99.74%, while the fraction of nonconforming items is 0.26%. Although this percentage seems small, in terms of parts per million (ppm), it can still result in an unacceptable fraction of defects. Assessing the process capability for a manufacturing process is of paramount importance for quality assurance purposes. It is worth remembering that the process capability can be assessed only once the in-control condition of the process has been observed. Therefore, the process capability analysis in SPC typically follows the in-control condition assessment performed by the use of control charts.

Critical Aspects of Traditional SPC

Critical reviews of SPC tools have been proposed in the literature (Woodall 2000), pointing out the main barriers while implementing SPC in real manufacturing contexts. The major criticalities that have been raised in the recent years are summarized in the following:

- Robustness to data gathering: A first detailed analysis of process output data is recommended before selecting the correct control chart type to be adopted. Indeed, the selection of the control chart highly affects the monitoring effectiveness. This is not always easy, especially when data collection is affected by measurement uncertainty and other disturbances.
- Dependency on the process mean estimation: SPC tools performance is highly affected by the process mean estimation accuracy. As a matter of fact, the theoretical analysis of the control chart performance can be very far from reality. Indeed, since the a and b probabilities refer to the tails of the statistical distribution of the sample statistics, the empirical values can be far from the theoretical ones. A preliminary sensitivity analysis of the performance to a and b is suggested before any implementation to address this problem.
- Normality assumption: Most traditional SPC tools are based on the assumption that the process output characteristic is normally distributed, among which Shewhart control charts and multivariate control charts. In some cases, the central limit theorem can be used to justify approximate normality when monitoring means, but in numerous cases normality is an untenable assumption, and one is unwilling to use another parametric model. A number of nonparametric methods are available in these cases. As data availability increases, nonparametric methods seem especially useful in multivariate applications where most methods proposed thus far rely on normality.
- Economic design: Economic control chart design requires the availability of cost coefficients that are not easy to estimate in practice. Furthermore, there are costs that can only be assessed by taking a system level perspective. For example, the false alarm costs of a machine may vary depending on the effect of the machine on the entire system. If a bottleneck process is monitored, false alarm costs may be very high since the effect will be directly propagated to the entire system. If the machine is not a bottleneck, the impact on the system performance and consequently the cost may be negligible.

New Frontiers of SPC Approaches

Advances in automated manufacturing systems coupled with advances in sensing and automatic inspection technology will continue to increase the volume of data available for drawing inferences about many processes. In some applications, this will change the inference problem from one dependent on scarce data to one based on plentiful data. In this context, the following opportunities and challenges raise:

- Monitoring of geometric features and profile monitoring: The increasing product complexity and the diffusion of geometrical tolerances have generated interest in statistical approaches for monitoring functional data with respect to traditional dimensional data. New SPC methods based on advanced statistical approaches such as principal component analysis (PCA) have been shown as viable solutions to this problem.

- SPC in multistage process chains: In real and complex manufacturing systems, the output of a process can affect the quality of downstream processes. For example, in automotive assembly, the output of sheet metal-forming processes affects the output of spot welding processes. SPC methods that are effective to capture these correlations have only recently been proposed and are of interest for the future.
- SPC for small batch, "one of a kind" processes: With the development of customized productions, there is a growing need to develop SPC tools for processes producing in small batches where few clones of the same item are produced. In this case, only few data are available, and it is a challenge to define in-control conditions and process capability indices. Ad hoc SPC tools should be developed for this purpose.
- Process capability indices for micromanufacturing processes: Micromanufacturing processes have found wide applications in the last decade. However, the problem of controlling the quality of microproducts and monitoring microprocesses is still far from being solved. The main challenge regards the consideration of the measurement uncertainty that is typically comparable to the product tolerances in micromanufacturing.
- Integrated SPC, production logistics, and maintenance (Colledani and Tolio 2006): Although quality control, production logistics control, and maintenance control are highly interrelated aspects determining the overall profitability of manufacturing systems, they have been traditionally considered almost in isolation both by researchers and practitioners. However, there is evidence that the production system design affects the product quality and that the maintenance policies could be improved if used in conjunction with quality control tools. Recent contributions have shown that consistent benefits could be achieved by a joint consideration of these production functions.
- Inspection effort allocation in multistage systems: Although several rules of thumb have been developed to allocate inspection and measurement stations (where and what to inspect) in complex multistage systems to simultaneously guarantee the right monitoring and the right productivity levels, practitioners still massively use end-of-line inspection due to the lack of scientific methods supporting this design activity. However, end-of-line inspections do not allow a reactive control of early production stages and only allow detecting nonconformities when several defect causes have been accumulated throughout the process stages. Typical considerations that have been suggested regard the processing and inspection time, the lot sizes, the interference with logistics performance (throughput time), the operators' availability, and economical figures. However, a comprehensive model including these features would represent a useful tool for manufacturing companies while solving this challenging issue.

These new challenges in statistical process control have motivated the launch of quality control-oriented topics within several international research funding programs. For example, within the European Community 7th Framework Program, the topic "zero defect manufacturing" has generated calls for proposals in recent years.

Cross-References

- ▶ Inspection (Assembly)
- ▶ Inspection (Precision Engineering and Metrology)
- ▶ Micromachining
- ▶ System
- ▶ Tolerancing

References

Alwan LC (2000) Statistical process analysis. Int edn. Irwin/McGraw-Hill, Boston/Singapore

Colledani M, Tolio T (2006) Impact of quality control on production systems performance. Ann CIRP 55(1):453–456

Colledani M, Tolio T (2011) Joint design of quality and production control in manufacturing systems. CIRP J Manuf Sci Technol 4(3):281–289

Colosimo BM, del Castillo E (2006) Bayesian process monitoring, control and optimization. Chapman and Hall/CRC, Boca Raton

Del Castillo E (2002) Statistical process adjustment for quality control, Wiley series in probability and statistics. Wiley, New York
Ishikawa K (1976) Guide to quality control. Asian Productivity Organization, Tokyo
Lorenzen TJ, Vance LC (1986) The economic design of control charts: a unified approach. Technometrics 28(1):3–10
Montgomery DC (2013) Introduction to statistical quality control, 7th edn. Wiley, Hoboken
Woodall WH (2000) Controversies and contradictions in statistical process control. J Qual Technol 32(4):341–378

Statistical Process Control (SPC)

► Statistical Process Control

Statistical Quality Control (SQC)

► Statistical Process Control

Storage

► Microwave Radiation

Straightening (for Reduced Number of Rolls Having Bigger Diameter)

► Roll Levelling

Straightness

Robert Schmitt
Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen, Germany

Definition

Straightness is a form tolerance that prescribes the largest possible deviation of a line from its geometrical ideal form (Pfeifer and Schmitt 2010). For an axis, the straightness tolerance specifies that the derived median line must lie within some cylindrical zone whose diameter is the specified tolerance. For the line elements of a feature, the straightness tolerance specifies that each line element must lie in a zone bounded by two parallel lines that are separated by the specified tolerance and that are in the cutting plane defining the line element.

Theory and Application

Figure 1 shows an example for a tolerance of spatial straightness. The reference arrow is shown in the extension of the dimension line of the largest cylinder, and therefore the tolerance of straightness is related to the axis of this cylinder. The real axis must be over the whole length within the cylinder with a diameter of the tolerance value, in the example shown 0.06 mm.

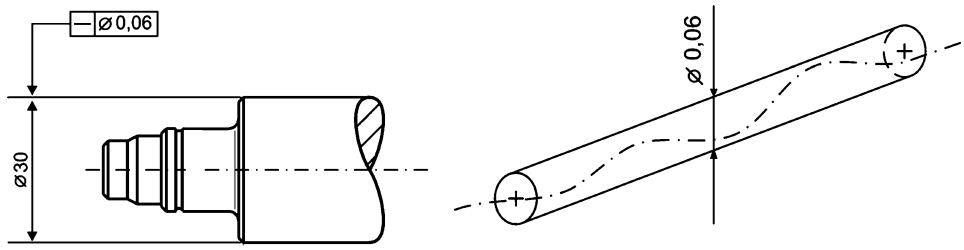
Straightness tolerances are necessary if the straightness is not sufficiently tolerated by another form tolerance, e.g., flatness or cylinder form, or by a location tolerance, e.g., parallelism, perpendicularity, angularity, symmetry, or position (Pfeifer and Schmitt 2010).

Assessment

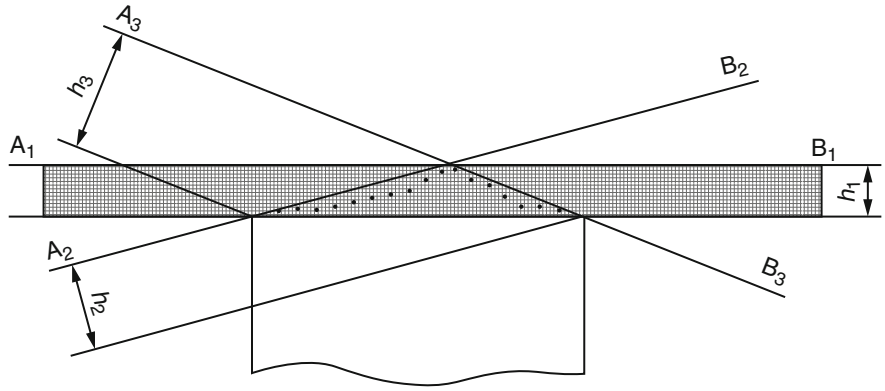
The straightness of a single toleranced feature is deemed to be correct when the feature is confined between two straight lines, and the distance between both is equal to or less than the value of the specified tolerance. The orientation of the straight lines shall be chosen so that the maximum distance between them is the least possible value (DIN EN ISO 1101 2014).

An example for a particular cross section is given in Fig. 2.

Therefore the correct orientation of the straight lines is A1-B1. The distance h1 is to be equal to or less than the specified tolerance (DIN EN ISO 1101 2014).



Straightness, Fig. 1 Straightness



Possible orientations of the straight lines:	A_1-B_1		A_2-B_2		A_3-B_3
Corresponding distances:	h_1		h_2		h_3
In the case of Figure B.1:	h_1	<	h_2	<	h_3

Straightness, Fig. 2 Straightness tolerance

Cross-References

► Tolerancing

References

DIN EN ISO 1101 (2014) Geometrical product specifications (GPS): geometrical tolerancing – tolerances of form, orientation, location and run-out. Beuth, Berlin

Pfeifer T, Schmitt R (2010) Fertigungsmesstechnik [Metrology in production], 3rd edn. Oldenbourg, München (in German)

Stratified Surface

François Blateyron
Digital Surf, Besançon, France

Synonyms

Multi-process surface; Plateau-like surface

Definition

Surface obtained after two sequential machining processes that transfer two types of texture signature on the material, usually inscribed at two different depth layers. Some processes may have more than two stages. The term stratified comes from *strata*, which means layers, and that underlines the fact that the surface has several functional behaviors depending on the layer in depth.

Theory and Application

The term *stratified surface* is often used in conjunction with the term *functional* in the scope of engineering surfaces, as the machining process intends to produce functional surfaces that achieve a programmed performance index.

Honed Cylinder Liners

One of the most frequent mechanical components to exhibit stratified surfaces is the cylinder liner whose surface is machined in two or three sequences. The aim is to produce a *plateau-like surface* with a relatively flat and smooth surface on the top and a series of deep grooves. The top surface guarantees a correct sealing between the liner surface and piston rings and reduces wear and friction by avoiding hills and peaks above the surface. Grooves allow lubrication and storage of fine particles. Modern manufacturing nowadays uses a three-stage process known as "plateau-honing" or "slide-honing," starting with a "rough-honing" aimed at creating the cross-hatched grooves and correcting cylindricity deviations, followed by a second intermediate honing aimed at producing the core roughness, and a third stage honing with fine grit ceramic sticks aimed at refining the plateau surface (see Whitehouse 1983; Stout et al. 1990; Pawlus 2008; Dimkovski et al. 2013).

Honed surfaces can be recognized by their nonsymmetrical shape and their height distribution skewed toward the top surface (see Fig. 1).

The relatively flat top surface (plateau-like) provides strong contact areas that are supposed to resist to the strong pressure during function without too many elastic deformation and if possible without any plastic deformation. This surface should minimize friction with the piston and reduce wear in order to avoid creating fine particles and residues that may obstruct oil retention grooves.

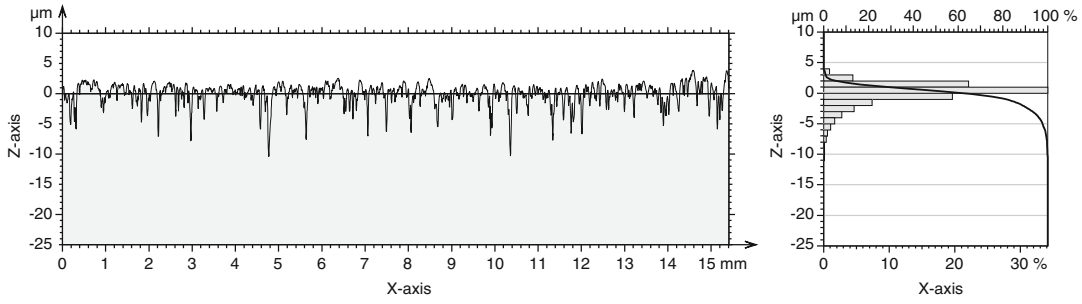
The grooves provide paths for the lubrication fluid under hydrodynamic pressure, and their density, depth, and volume need to be controlled in order to reduce oil consumption and air pollution.

Mass Finish

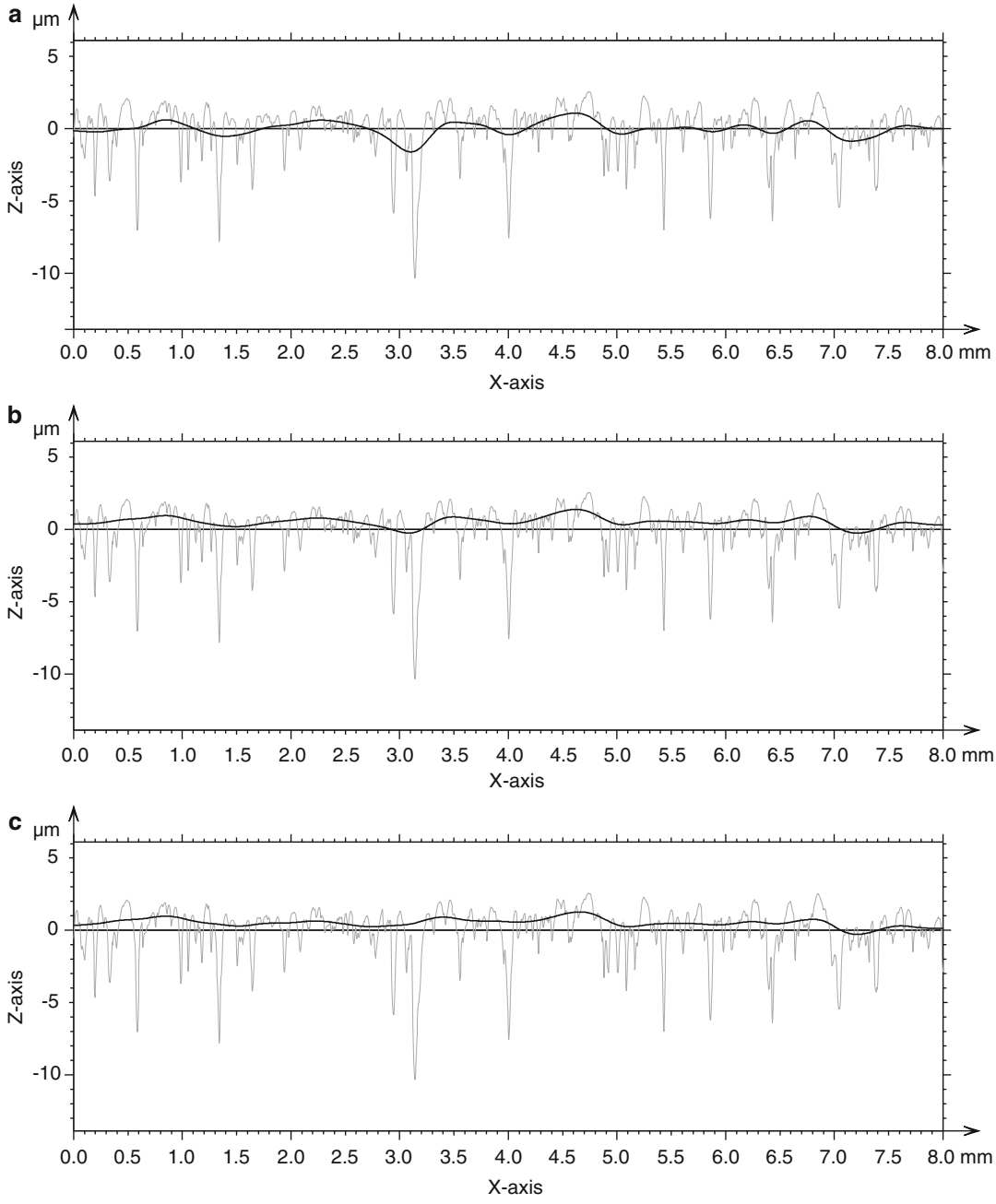
Many mechanical components are finished by mass finish in order to reduce the fine roughness on the surface or create a shiny aspect, which is obtained by reducing roughness amplitudes. Depending on the first machining operation, these surfaces after mass finish have a texture morphology close to that of stratified surfaces, which is characterized by a height distribution shifted toward the top surface.

Surface Texture Characterization

Stratified surfaces cannot be characterized with classical height parameters such as *Ra* and *Rq* (or *Sa*, *Sq*) because these parameters average all surface heights without segmenting points that are in grooves and those that are close to the top surface. Moreover, the skew of the height distribution, due to deep grooves, makes it difficult for the usual Gaussian filter to produce a mean line (or mean plane) that follow the top surface without being influenced by these grooves (Fig. 2a). That is why robust filters are preferred on stratified surfaces.



Stratified Surface, Fig. 1 Left: profile of a plateau-honed surface showing deep valleys. Z-axis: heights from the mean line. X-axis: evaluation length. Right: height distribution and height density curve (Abbott-Firestone curve). Z-axis: heights from the mean line. Upper x-axis: material ratio (on the Abbott curve). Lower x-axis: population of each class in the distribution



Stratified Surface, Fig. 2 (**a**) Gaussian filter, ISO 16610–21. The mean line created by the waviness (dark curve) is influenced by the deep grooves, so it creates artificial overshoot on the roughness profile by lifting grooves upper parts on both sides. (**b**) Double-Gaussian, ISO 13565–1. The mean line is less influenced by deep grooves and better follows the general trend of the upper surface. (**c**) Robust Gaussian, ISO 16610–31. The mean line is even more robust against deep grooves and would also be robust for high peaks (contrary to the double-Gaussian)

Filtration

Double–Gaussian: ISO 13565–1

The double-Gaussian filter was a first attempt to produce a robust filter. It is robust against deep grooves but not against peaks; this is why it is dedicated at stratified surfaces. It was initially defined in the DIN 4777 German standard and dedicated to the filtration of automotive surfaces, especially cylinder liners. It is calculated via an interim waviness profile that is used to threshold deep grooves and produce a modified roughness profile called "valley-cut profile," on which the final waviness profile is calculated (Fig. 2b). That second waviness is not (or much less) influenced by deep grooves.

Robust Gaussian: ISO 16610–31

The part 31 of ISO 16610 introduces a robust Gaussian filter for profiles. Its areal version is described in ISO 16610–71. The aim of the robust filter is to provide a mean line (mean plane) that is not influenced by local irregularities such as deep grooves and high peaks, in order to flatten the roughness profile without altering these irregularities that are of great importance for stratified surfaces (Fig. 2c).

The robust Gaussian filter is implemented as an iterative algorithm where a polynomial is fitted locally to the profile, taking into account the previous iteration to weight the tension between the mean line and the primary profile (see Brinkmann et al. 2000; Li et al. 2004; Dobrzanski and Pawlus 2013). After several iterations, the mean line is "relaxed" from the influence of irregularities and better fits the general trend of the main profile.

Robust Spline: ISO 16610–32

The part 32 of this standard series initially proposed a robust version of the spline filter, but the nonlinear resolution of its equation, required for an implementation, made it quite difficult to realize. And its benefits, compared to the robust Gaussian, were too tiny. When the standard series was about to be converted from TS (technical specification) to IS (international standard), it was decided to cancel the robust spline filter from the list.

Upper Envelope: CNOMO/ISO 12085

Contrary to the previous filters that calculate a mean line and are part of the so-called M-system, the R&W motifs method described in ISO 12085 – formerly CNOMO recommendation, *Comité de NOrmalisation des MOyens de production* (fr.) or *Standardisation Committee for Manufacturing and Equipment* (Eng.), formed mainly by Renault and Peugeot-Citroën – proposes to calculate an envelope that more or less corresponds to the upper contact surface. It is part of the "E-system." Besides the parameters that calculate mean height and mean width of texture motifs, this standard proposes to calculate a levelled profile using the upper envelope and then calculate functional parameters derived from the *Rk* parameters. These variations, called *Rke*, *Rpke*, and *Rvke*, were used mainly within the French automotive industry (Peugeot-Citroën and Renault) during the 1980s and 1990s, but they are well suited to characterize stratified surfaces.

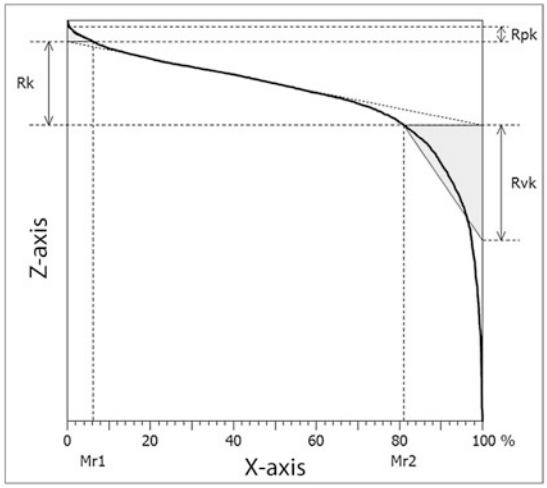
Parameters

The first parameters introduced specifically for the characterization of stratified surfaces were defined in a German standard DIN 4776, later included in ISO 13565–2 and ISO 13565–3. They are based on the Abbott-Firestone curve for parameters of part 2 and on a probability curve for parameters of part 3 (see Pawlus et al. 2013).

In the usual list of parameters, the skewness parameter *Rsk* (or *Ssk*) is an interesting indicator as it should provide a negative value on almost all stratified surfaces.

Rk Parameters: ISO 13565–2

The construction starts with finding the lowest slope on the Abbott-Firestone curve, within a sliding window of 40% width, and then to find the intersection points on both sides (0% and 100%) of the slope segment. The vertical distance between these points defines the *Rk* parameter, *core roughness depth*. The horizontal intersection from these points to the curve gives two material ratio values, *Mr1* (on the left) and *Mr2* (on the right). The vertical distance from the left intersection point and the upper-left corner is *Rpk**, and the other from the right intersection point and the lower-right corner *is Rvk**. The void area enclosed between the curve, the horizontal line



ISO 13565		
ISO 13565-2		
Rk	2.541	µm
Rpk	0.4384	µm
Rvk	3.496	µm
Mr1	6.173	%
Mr2	81.03	%
A1	13.53	µm²/mm
A2	331.6	µm²/mm
*Rpk**	0.6409	µm
*Rvk**	8.958	µm

Stratified Surface, Fig. 3 Left: the graphical construction of Rk parameters on the Abbott-Firestone curve. Z-axis: heights (not labelled). X-axis: material ratio. Right: the parameter set of ISO 13565–2

defining *Mr2* and the vertical axis 100%, defines the parameter *A2*, usually given in µm²/mm. By construction, the triangle built from the horizontal line defining *Mr*2, that have the same area as *A2*, will give the height *Rvk, reduced valley depth*. A similar construction can be done from the other horizontal intersection, to define *A1* and *Rpk* (Fig. 3).

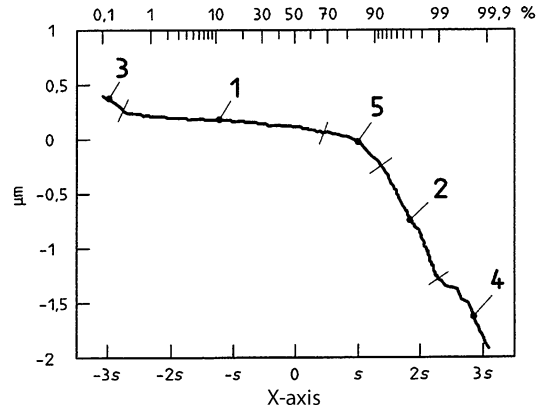
Initially associated to the double-Gaussian filter, these parameters are nowadays calculated after a robust Gaussian filter (VDA 2006).

When measuring a surface, similar areal parameters can be used, as defined in ISO 25178–2: *Spk*, *Sk*, *Svk*, *Sr1*, and *Sr2*. No official equivalents have been defined for *A1*, *A2*, *Rpk**, and *Rvk**, but they may be used as well and could have been called *Sa1*, *Sa2*, *Spk**, and *Svk**.

ISO 13565–3

This standard defines a set of parameters calculated on the material probability curve, which is a transformation of the Abbott curve in order to draw it in function of the standard deviation of a normal distribution. In case of stratified surfaces, the two process stages correspond to two normal distributions mixed together. The material probability then shows two straight lines (asymptotes) that can be characterized as shown in Fig. 4.

Details on this method can be found in Malburg and Raja (1993).



Stratified Surface, Fig. 4 The main regions of interest are marked with (1) corresponding to the plateau part of the profile and (2) corresponding to the valley part Sections (3), (4), and (5) are not analyzed per se Vertical axis represents heights from the mean line. The upper horizontal axis is the probability to have a point at the corresponding height. The lower horizontal axis represents deviations from the Gaussian distribution, labelled in standard deviation

Parameters can be calculated either on the primary profile or the roughness profile (after a double-Gaussian filtration):

- *Ppq/Rpq*: Slope of the plateau area (can be seen as the Rq value of the process used to create plateaus)

- *Pvq/Rvq*: Slope of the valley area (can be seen as the Rq value of the process used to create valleys)
- *Pmq/Rmq*: Material ratio at the intersection of the two process areas

These parameters are usually used to control process parameters during the machining of stratified surfaces.

SurfStand: EUR 15178 EN

The so-called "SurfStand" project, financed by the European Commission between 1993 and 2000 under the program called Applied Metrology and Chemical Analysis, had a section that proposed functional indices, aimed at characterizing stratified surfaces.

Three indices were defined:

- *Sbi*, surface bearing index
- *Sci*, core fluid retention index
- *Svi*, valley fluid retention index

These parameters have barely been used in the industry because they have been quickly replaced by volume parameters.

Volume Parameters: ISO 25178–2

Volume parameters are defined from material volume or void volume:

- *Vmp*: Material volume of peaks (can be seen as a sibling of *Spk*)
- *Vmc*: Material volume of the core (can be seen as a sibling of *Sk*)
- *Vvc*: Void volume of the core
- *Vvv*: Void volume of the valleys (can be seen as a sibling of *Svk*)

Volume parameters are given in volume units, either ml/m^2 or $\mu m^3/mm^2$. While these parameters are quite generic, they can be used to characterize stratified surfaces.

Conclusion

Stratified surfaces are an important part of functional surfaces used in the automotive industry, mainly for cylinder liners. The specific nature of their surface texture has led to the creation of dedicated filters and parameters in order to overcome drawbacks of traditional methods.

Cross-References

- ▶ Roughness
- ▶ Superfinishing
- ▶ Surface Parameter
- ▶ Surface Texture

References

Brinkmann S, Bodschwinna H, Lemke H-W (2000) Development of a robust Gaussian regression filter for three-dimensional surface analysis. In: Xth international colloquium on surfaces, Chemnitz, Germany, pp 122–131

Dimkovski Z, Anderberg C, Ohlsson R, Rosén B-G (2013) Characterisation of cylinder liner honing textures for production control. In: Leach R (ed) Characterisation of areal surface texture. Springer, London

Dobrzanski P, Pawlus P (2013) Modification of robust filtering of stratified surface topography. Metrol Meas Syst 20(1):107–118

Li H, Jiang X, Li Z (2004) Robust estimation in Gaussian filtering for engineering surface characterization. Prec Eng 28:186–193

Malburg MC, Raja J (1993) Characterization of surface texture generated by plateau-honing process. CIRP Ann 42:637–640

Pawlus P (2008) Simulation of stratified surface topographies. Wear 264(5–6):457–463

Pawlus P, Reizer R, Lenart A (2013) Comparison of parameters describing stratified surface topography, Journal of physics, conference series, vol 483, conference 1

Stout KJ, Davis EJ, Sullivan PJ (1990) Plateau honed surfaces. In: Atlas of machined surfaces. Springer, Dordrecht

Whitehouse DJ (1983) Some theoretical aspects of a practical measurement problem in plateau honing. Int J Prod Res 21(2):215–221

Further Readings

Deepak Lawrence K, Ramamoorthy B (2016) Multi-surface topography targeted plateau honing for the processing of cylinder liner surfaces of automotive engines. Appl Surf Sci 365:19–30

Grabon W, Pawlus P (2014) Description of two-process surface topography. Surf Topogr Metro Prop 2:025007

Hu S et al (2017) Truncated separation method for characterizing and reconstructing bi-Gaussian stratified surfaces. Friction 5(1):32–44

Santochi M, Vignale M, Giusti F (1982) A study on the functional properties of a honed surface. Cirp Annals, Manuf Tech 31(1):431–434
Walton K, Blunt L, Fleming L (2015) The topographic development and areal parametric characterization of a stratified surface polished by mass finishing. Surf Topog Metrol Prop 3(3):035003
Yousfi M, Mezghani S, Demirci I, El Mansori M (2013) Study on the relevance of some description methods for plateau honed surfaces. In: Fourteenth international conference on metrology and properties of engineering surfaces, Taipei

ISO Standards (International Organisation of Standards, Geneva)

ISO 12085:1996, GPS – surface texture: profile method – motif parameters
ISO 13565-1:1996, GPS – surface texture: profile method – surfaces having stratified functional properties – part 1: filtering and general measurement conditions
ISO 13565-2:1996, GPS – surface texture: profile method – surfaces having stratified functional properties – part 2: height characterization using the linear material ratio curve
ISO 13565-3:1998, GPS – surface texture: profile method – Surfaces having stratified functional properties – part 3: height characterization using the material probability curve
ISO/TS 16610-32:2009, GPS – filtration – robust profile filters: spline filters
ISO 16610-21:2011, GPS – filtration – linear profile filters: Gaussian filters
ISO 25178-2:2012, GPS – surface texture: Areal – Terms, definitions and surface texture parameters
ISO 16610-71:2014, GPS – filtration – robust areal filters: Gaussian regression filters
ISO 16610-31:2016, GPS – filtration – robust profile filters: Gaussian regression filters

Other Standards

DIN 4776:1990, Determination of surface roughness parameters Rk, Rpk, Rvk, Mr1, Mr2 serving to describe the material component of roughness profile, German Institute for Standardisation, *Deutsches Institut für Normung*
DIN 4777:1990, Metrology of surfaces, Profile filters for electrical contact stylus instruments, Phase-corrected filters, German Institute for Standardisation, *Deutsches Institut für Normung*
EUR 15178 EN:1993, Stout K, The development of methods for the characterisation of roughness in three dimensions, EC Brussels, ISBN 0704413132
VDA 2006:2003, GPS – surface texture – rules and procedures for the assessment of surface texture, *Verband der Automobilindustrie*

Stress, Strain

Alan Bramley
Mechanical Engineering, University of Bath, Bath, UK

Synonyms

Deformation

Definition

A stress is a force F distributed over a flat supporting surface.

Strain is a parameter which describes the relative displacement of particles within a deforming body.

Theory and Application

Stress

A stress is a force F distributed over a flat supporting surface of Area A.

Stress at a point, identified with σ, is given by

$$\text{limit as } \delta A \to 0 \text{ of } \delta F/\delta A = dF/dA$$

Stress may vary over surface.

Stress may act as an arbitrary angle.

It is appropriate to transform into components in x, y, and z system.

Choosing x as normal to surface with y and z being tangential:

$$\sigma_{xx} = dF_x/dA$$

$$\sigma_{xy} = dF_y/dA$$

$$\sigma_{xz} = dF_z/dA$$

Suffix notation: first suffix – direction of normal to area. Second suffix – direction in which component acts.

Sign convention: normal stress – Tensile +ve. Compressive −ve. Shear stress – follows coordinates.

Stress tensor: Within a cartesian coordinate system, the stress at any point can be defined by nine stress components:

$$\sigma_{ij} = \begin{matrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{matrix}$$

Similarly in cylindrical coordinates,

$$\sigma_{ij} = \begin{matrix} \sigma_{rr} & \sigma_{r\theta} & \sigma_{rz} \\ \sigma_{\theta r} & \sigma_{\theta\theta} & \sigma_{\theta z} \\ \sigma_{zr} & \sigma_{z\theta} & \sigma_{zz} \end{matrix}$$

it is usual to write

$$\sigma_{xx} = \sigma_x \text{ etc.}$$
$$\sigma_{xy} = \tau_{xy} \text{ etc.}$$
$$\text{thus } \sigma_{ii} = \sigma_x + \sigma_y + \sigma_z$$
$$\text{and } \sigma_{ij}\ \delta\varepsilon_{ij} = \sigma_x\delta\varepsilon_x + \sigma_y\delta\varepsilon_y + \sigma_z\delta\varepsilon_z + 2\left(\sigma_{yz}\delta\varepsilon_{yz} + \sigma_{zx}\delta\varepsilon_{zx} + \sigma_{xy}\delta\varepsilon_{xy}\right)$$

Strain

Strain is a parameter which describes the relative displacement of particles within a deforming body. In the general case of a body subjected to stress, normal (or linear) and shear strains will occur together with translational and rotational displacements of elements within the body.

Since only those displacements corresponding to strain can be related directly to the applied stress through the mechanical properties of the continuum, it is necessary to develop formulae which will describe strains in terms of measured displacements which themselves may include components of translational and rotational movement.

The relationships between strains and displacements in Cartesian coordinate system are usually developed by ascribing u = u(x, y, z); v = v(x, y, z); and w = w(x, y, z) to displacements in the x, y, and z directions, respectively, and considering a generalized movement and strain of an elemental cube of dimensions δx, δy, and δz such as ABCD in Fig. 1a. The cube is considered to be sufficiently small that strains will be constant across it. A consequence of this assumption is that parallelism between opposing faces of the cube will be maintained.

Suppose applied stresses transform the cube to A′B′C′D′. Consider now the movement of AC to A′C′ (Fig. 1b).

Elimination of translational movement by constructing A″C″ parallel to AC with A′ and A″ coinciding enables the length of A′C′ to be determined in terms of u, v, and w:

$$(A'C')^2 = (\delta x + \partial u/\partial x \cdot \delta x)^2 + (\partial v/\partial x \cdot \delta x)^2 + (\partial w/\partial x \cdot \delta x)^2$$
$$= (\delta x)^2 \left[1 + 2\partial u/\partial x + (\partial u/\partial x)^2 + (\partial v/\partial x)^2 + (\partial w/\partial x)^2\right]$$

The linear strain of AC is defined by the ratio of its change in length to its original length,

$$\text{i.e., } (A'C' - AC)/AC = (A'C'/AC) - 1$$

Now AC = δx.

Therefore linear strain of AC, $e_{xx} = [1 + 2(\partial u/\partial x) + (\partial u/\partial x)^2 + (\partial v/\partial x)^2 + (\partial w/\partial x)^2]^{1/2} - 1$.

If any application this equation can be restricted to situations involving displacements sufficiently small that second-order terms become negligible and the linear strain of AC can be assumed to correspond to a strain in the x-direction, then Eq. 1, after binomial expansion, reduces to

$$e_{xx} = \partial u/\partial x.$$

Similar expressions for e_{yy} and e_{zz} can be derived by considering the transformations of AD and AB.

The use of the letter e is generally synonymous with the designation "engineering strain."

A description of the strain is completed by considering changes in the angles BAD, DAC,

and BAC. For example, in the x-y plane (Fig. 1c), the right angle DAC transforms to θ and $(\pi/2 - \theta)$ is a measure of the amount of this transformation:

$$
\begin{aligned}
&(\pi/2 - \theta) = \alpha_1 + \alpha_2 \tan\ \alpha_2 \\
&= \left[(\partial v/\partial x \cdot \delta x)^2 + (\partial w/\partial x \cdot \delta x)^2\right]^{1/2} \\
&/[\delta x + \partial u/\partial x \cdot \delta x]^{1/2} \\
&= \left[(\delta v/\delta x)^2 + (\partial w/\partial x \cdot \delta x)^2\right]^{1/2} \\
&/[1 + \partial u/\partial x]^{-1}
\end{aligned}
$$

The second-order term $(\delta w/\delta x)^2$ can be neglected and also $(1 + \partial u/\partial x) \approx 1$ thus,

$$
\begin{aligned}
\tan \alpha_2 \quad &\approx \quad \partial v/\partial x \\
\text{i.e.,}\ \ \alpha_2 \quad &+ \quad \partial v/\partial x
\end{aligned}
$$

Similarly, $\alpha_1 \approx \partial u/\partial y$

Thus $(\pi/2 - \theta)$ can be expressed as

$$\phi_{xy} = \partial u/\partial y + \partial v/\partial x$$

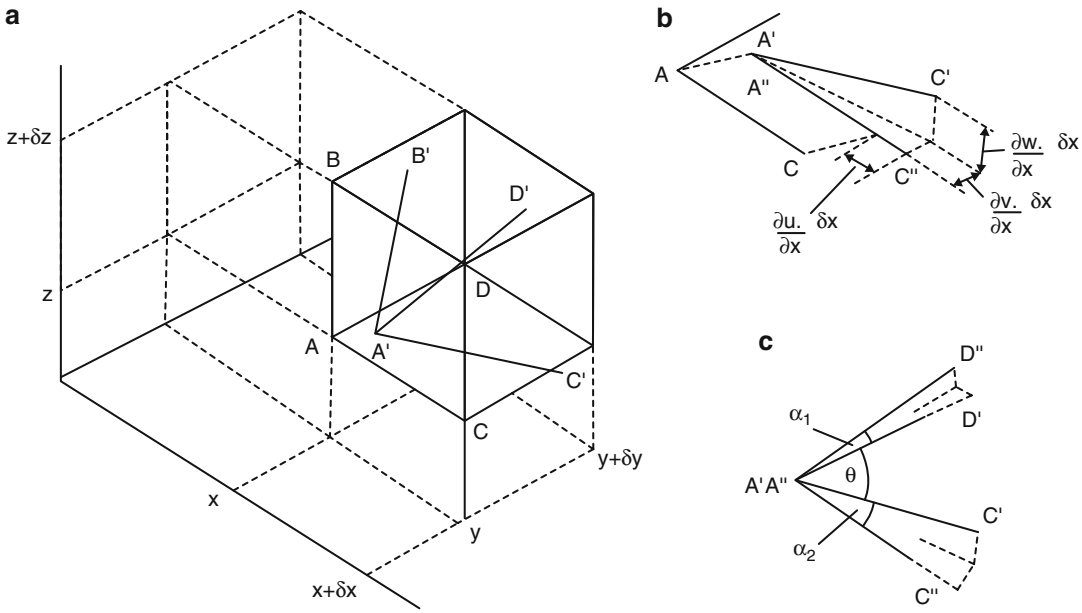
where ϕ_{xy} is known as the engineering shear strain and is the sum of two quantities usually referred to as e_{xy} and e_{yx}. It would thus appear that ϕ_{xy} does not determine uniquely the shear distortion since it can correspond to an infinity of e_{xy} and e_{yx} values. However, reference to Fig. 2 shows that the differences between e_{xy} and e_{yx} corresponds solely to the rotational displacement of the element thus contributing nothing to the shear distortion which is therefore uniquely defined by $\phi_{xy.}$ By considering the other planes and direction in the Cartesian coordinate system, the complete strain–displacement equations can be formulated:

$$e_x = \partial u/\delta x \quad e_y = \partial v/\partial y \quad e_z = \partial w/\partial z$$

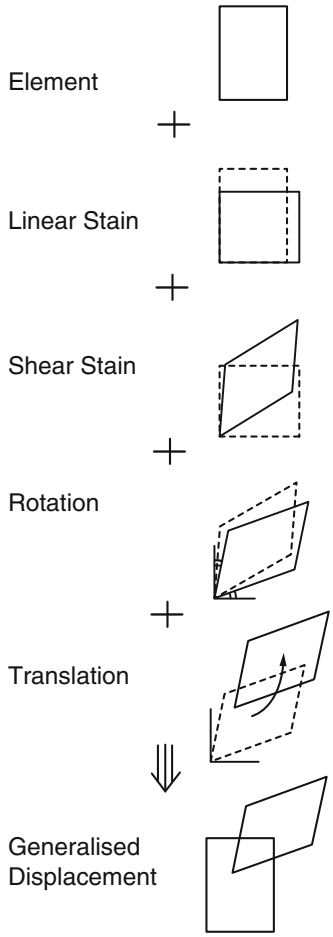
and

$$
\begin{aligned}
\phi_{xy} &= \partial u/\partial y + \partial v/\partial x \\
\phi_{yz} &= \partial v/\partial z + \partial w/\partial y \\
\phi_{xz} &= \partial u/\partial z + \partial w/\partial x
\end{aligned}
$$

Problems in the cylindrical coordinate system are also frequently encountered, and with similar assumptions regarding second-order terms, the corresponding strain–displacement equations appear as follows:



Stress, Strain, Fig. 1 (**a**) Generalized displacement in 2-D deformation, (**b**) linear strain, and (**c**) shear strain



Stress, Strain, Fig. 2 Illustrates the breakdown of general displacement for a 2-D case of strain

$$e_r = \partial u/\partial r \quad e_\theta = 1/r(u + \partial v/\partial\theta)$$
$$e_z = \partial w/\partial z$$
$$\phi_{\theta z} = 1/r.\partial w/\partial\theta + \partial v/\partial z$$
$$\phi_{rz} = \partial u/\partial z + \partial w/\partial r$$
$$\phi_r\theta = 1/r(\partial u/\partial\theta - v) + \partial v/\partial r$$

Essentially, the foregoing can be considered as the theory of infinitesimal strain and to be valid for problems involving the elasticity of metals. It is apparent that the use of these relations for finite strain would lead to considerable error whether or not the strain was uniform. The difficulty can be overcome by realizing that since the foregoing equations are valid for infinitesimal total strains, then they must also be valid for infinitesimal increments of finite strains. The value of the total finite strain can then be determined from an integration of these increments. The symbol ε is normally used to denote finite linear strain, the strain–displacement equations henceforth being valid with δε replacing e:

$$\text{i.e.,}\ \ \delta\varepsilon_x = \partial u/\partial x \text{ etc.}$$

Lack of universal agreement for an appropriate symbol to denote finite shear strain can lead to confusion when referring to textbooks; δϕ, δγ, δεij are variously encountered.

It is useful and instructive to consider conditions of uniform finite straining where integration of strain increments can be performed analytically. Uniform straining is taken to mean that the principal axes of strain do not rotate. Plastic flow in a simple tensile test (Fig. 3a) prior to the onset of necking fulfils this condition.

Initially, the original gauge length l_o increases by an infinitesimal increment δl. This increment is represented graphically in Fig. 3b, whence strain increment can be formed as

$$\delta\varepsilon = \partial u/\partial x = \delta 1/1_0$$

For subsequent equal displacements δl (Fig. 3c), corresponding strain increments can be formed:

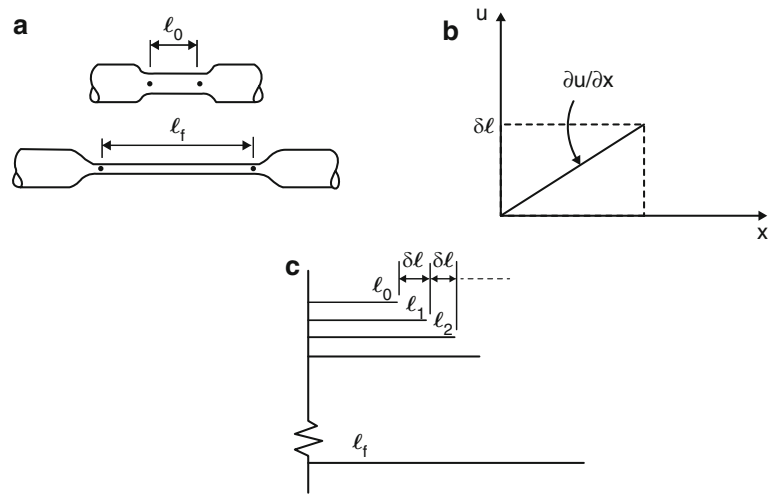
$$\delta\varepsilon_1 = \delta 1/1_\sigma \quad \delta\varepsilon_2 = \delta 1_1 \quad \delta\varepsilon_3 = \delta 1/1_2 \text{etc.}$$

i.e., in general $\delta\varepsilon = \delta 1/1$

The total strain can then be obtained by summing these increments from the original length 1_0 to the final length l_f say

$$\varepsilon = \int d\varepsilon = \int_{1_0}^{1_f} \delta 1/1 = 1n\ 1_f/1_0$$

Such total finite strains are known as natural or logarithmic strains, the use of which was first suggested by Ludwig (1909). The engineering strain is frequently used as a measure of finite linear strain (e.g., the percentage elongation in a simple tensile test is usually quoted). The relationship between these two measures can be derived as follows:



Stress, Strain, Fig. 3 (**a**) Plastic flow in simple tension, (**b**) a strain increment, and (**c**) successive increments of displacement

In simple tension $e_x = (l_f - l_0)/l_0 = (l_f/l_0) - 1$

$$\varepsilon = 1n(1 + e)$$

With very small strains ε and e become equal. However, e is not a realistic measure of large finite strain; thus the use of ε is favored under such conditions. For example, in simple compression of a billet from ho to h,

$$\varepsilon = \ln\ (h_0/h)$$
$$\text{and}\, e = h_0 - h/h_0$$

as h tends to zero, $\varepsilon \to \infty$, and $e \to 1$. ε is obviously a more realistic measure of the strain. Analytical work may involve the addition of strains during the finite steps of a finite strain program and the additive property of natural strains has obvious advantages. Thus, in simple tension where

$$l_0 \to l_1 \to l_2$$

in terms of e, these steps are

$$(l_1 - l_0)/l_0 \text{ and } (l_2 - 1_1/1_1)$$

$$\text{with total strain} = (l_2 - l_0)/l_0$$

$$\text{but } e_{01} + e_{12} \neq e_{02}$$

However, in terms of ε, the steps are $1n\ l_1/l_0$ and $1n\ l_2/l_1$ and total strain $= 1n\ l_2/l_0$

i.e., $\varepsilon_{01} + \varepsilon_{12} = \varepsilon_{02}$

Natural strains are thus better to use in the analysis of plastic flow.

Cross-References

▶ Flow Stress, Flow Curve

Reference

Ludwig P (1909) Elemente der technologischen Mechanik (Elements of technological mechanics). Springer, Berlin (in German)

Stretching

Andreas Sterzing
Fraunhofer Institute for Machine Tools and Forming Technology (IWU), Chemnitz, Germany

Synonyms

Elongating; Extension; Tensile forming; Tensile stress; Tension

Definition

Stretching means tensile forming of a workpiece (sheet metal part, bulk metal part) by a tensile force which is externally applied and acting in the longitudinal axis of the workpiece.

Theory and Application

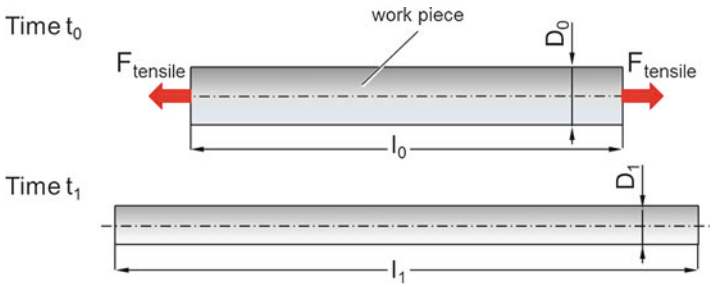
Generally, three different technological processes with various objectives can be differentiated in stretching (DIN 8585-2 2003).

1. Elongating (*stretching*) to increase the dimensions of the workpiece in the direction of force, e.g., in order to adapt to specified dimensions
2. Elongating (*stretch leveling*) in order to remove distortions from bars, profiles, and pipes and to remove dents from sheets
3. Elongating (*straightening by stretch bending*) in order to extend the rolling stock during (after) rolling, expressed by the degree of stretching

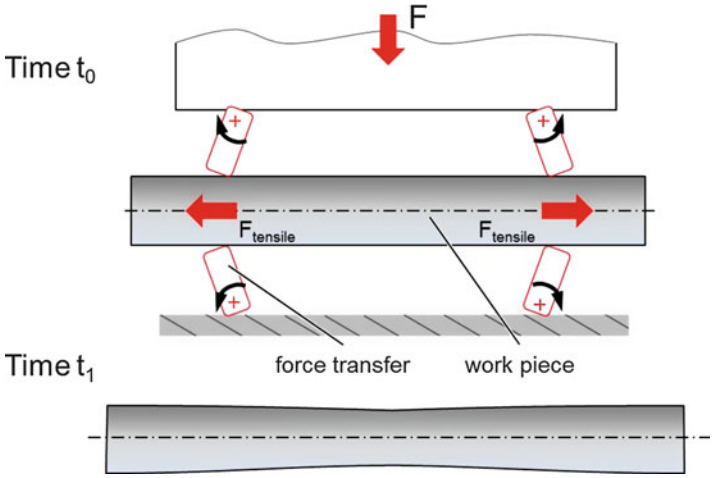
According to DIN 8585 and DIN 8582, elongating is part of the subgroup tensile forming within the main group 2 – forming (DIN 8585-1 2003). During tensile forming of a solid body, the plastic state is induced by uniaxial or multiaxial tensile load. A comparable homogeneous distribution of tensile stress develops in the complete forming zone.

Stretching

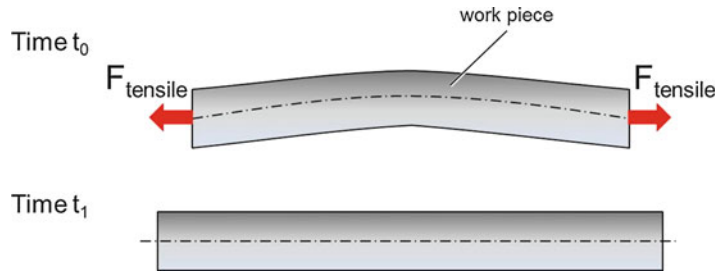
In stretching, the workpiece is subject to expansion in the direction of force. The tensile forces acting in the longitudinal axis of the workpiece can be applied to the workpiece in different ways (see Figs. 1 and 2).



Stretching,
Fig. 1 Stretching (with reference to DIN 8585-2 2003)



Stretching,
Fig. 2 Orthogonal operating directions of applied and tensile forces (with reference to DIN 8585-2 2003)



Stretching, Fig. 3 Stretch leveling (with reference to DIN 8585-2 2003)

This also includes the elimination of unknown residual stress (and internal stress) by a plate extensor. A plate extensor consists of units of movable and stationary clamping devices. Both clamping devices are equipped with gripping devices for holding the ends of the plates.

Elongating also includes a stretching process of sheet blanks close to or up to the flow limit of the material as a preliminary stage to tangential stretch forming.

Due to the occurring stress, this also covers the uniaxial tensile test on sheets and bars to determine the flow curve and the stress strain curve.

Stretch Leveling

Stretch leveling is intended to remove three-dimensional bends, roughness, and oddness at pipes and bars as well as buckles on sheets by tensile force (see Fig. 3). In this context, internal stress is removed by longitudinal strain.

The workpiece is fixed in a clamping device (chucks). The stretch leveling process is mainly carried out by hydraulic cylinders on a rack. The cross section is plasticized along the total length. Longitudinal strains from 1% to 3% are realized, depending on the workpiece dimensions.

When removing sheet buckles, a special tool, consisting of a rectangular upper and lower part with swivel-mounted plates, is used. As a consequence of the plate force acting on the sheet, the buckles are leveled (Fritz and Schulze 2006).

Straightening by Stretch Bending

Straightening by stretch bending is a strip straightening process to remove unevenness such as waves (waviness at edge and center) and saber shapes (transverse and longitudinal bends) from hot-rolled and cold-rolled strips (Spur and Stöferle 1985). Length adjustment in the strip is realized by plastic straining (straining degrees from 1% to 2%) of the shorter strip parts on stretch-bending devices. The latter consists of several pairs of upstream and downstream tensioning pulleys and a variable amount of bending rolls with a small diameter (e.g., from 5 to 23 rolls). The loading is characterized by a combination of bending and tensile stresses (Lange 1990).

Applications

Elongating (leveling) is applied in many branches of industry, including mechanical engineering, apparatus engineering, automotive industry, consumer goods industry, and the manufacture of semifinished products. Application examples are stretch leveling of coverings, adjusting the evenness of metal strips, achieving the desired straightness in flat bars, and leveling of extruded profiles (Bräutigam and Becker 2009).

An innovative application is represented by stretch leveling of complete coils for balancing strains along the total width and thickness. New systems are designed so that the gripper system does not leave impressions in the material. The result is an extremely flat, low-tension strip material. Strip thicknesses of up to 8 mm and strip width of up to 2100 mm can be leveled (Henrich 2004).

References

Bräutigam H, Becker S (2009) Richten mit Walzenrichtmaschinen [Straightening using roller straighteners]. Vogel Business Media, Würzburg. (in German)

DIN 8585-1 (2003) Fertigungsverfahren Zugumformen, Teil 1: Allgemeines; Einordnung, Unterteilung,

Begriffe [Manufacturing processes forming under tensile conditions, part 1: general; Classification, subdivision, terms and definitions]. Deutsches Institut für Normung e.V, Berlin. (in German)
DIN 8585-2 (2003) Fertigungsverfahren Zugumformen, Teil 2: Längen; Einordnung, Unterteilung, Begriffe [Manufacturing processes under tensile conditions, part 2: stretch reducing; classification, subdivision, terms and definitions]. Deutsches Institut für Normung e.V, Berlin. (in German)
Fritz H, Schulze G (2006) Fertigungstechnik [Production engineering]. Springer, Berlin/Heidelberg. (in German)
Henrich Publikationen (2004) Mit "SCS" zur rostfreien Stahloberfläche? Streckrichten von Coils mit Beize- und Ölfreiem Fertigungsverfahren [Using SCS (Stretch Cold Roll Surface) for stainless surfaces? Pickle- and oil-free straightening of coils]. bbr Bänder Bleche Rohre 45 (2004) 1/2:32–34. (in German)
Lange K (1990) Umformtechnik: Handbuch für Industrie und Wissenschaft, Band 3: Blechbearbeitung [Metal forming: handbook for research and industry, vol 3: sheet metal forming]. Springer, Berlin/Heidelberg. (in German)
Spur G, Stöferle T (1985) Handbuch der Fertigungstechnik, Band 2/1–3: Umformen und Zerteilen [Production engineering handbook, vol 2/1–3: forming and separating]. Carl Hanser, München/Wien. (in German)

Stripping-Down

► Disassembly

Structural Analysis

Mikel Zatarain
IK4-Ideko, Elgoibar, Gipuzkoa, Spain

Synonyms

Finite element analysis; Modal analysis

Definition

The process of evaluation of the stresses and of the distortion of a mechanical structure when submitted to load constraints.

Theory and Application

In the machine building industry, the most frequently used materials have an isotropic behavior. For these materials, some geometrically simple structures can be analyzed by analytical methods, what means to integrate the Hooke's law for isotropic materials in the structure domain. Hooke's law for these materials can be written as (Young and Budynas 2001):

$$\begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \\ \varepsilon_{xy} \end{bmatrix} = \frac{1}{E} \begin{bmatrix} 1 & -v & -v & 0 & 0 & 0 \\ -v & 1 & -v & 0 & 0 & 0 \\ -v & -v & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1+v & 0 & 0 \\ 0 & 0 & 0 & 0 & 1+v & 0 \\ 0 & 0 & 0 & 0 & 0 & 1+v \end{bmatrix} \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{zx} \\ \sigma_{xy} \end{bmatrix} \quad (1)$$

where ε represents the unitary deformations and σ the stresses. x, y, z are the Cartesian directions. E is Young modulus, and v Poison ratio.

For more detailed explanations and for the definitions of the terms, the website http://www.efunda.com/formulae/solid_mechanics/mat_mechanics/stress.cfm can be accessed.

Nonisotropic materials are more difficult to be analyzed, especially when using analytical methods. Composite materials, which are very seldomly used in machines, have nonisotropic behavior.

In most applications, the structure is considered as having linear behavior, but nonlinearity of several types can be found: nonlinear deformation models (rubber, etc.), nonlinear geometry (large deformations), plasticity (nonlinearity between strain and stress above the yield point), etc. Nonlinear structural analysis must almost inevitably be performed by numerical methods.

For a structure to be functional, it is required for it to be submitted to efforts giving rise to sufficiently low stress distribution, and at the same time, it is required that its static and dynamic deformations be smaller than some allowable value. In the case of chip removing machine tools, the deformation tends to be the limiting factor, while in cutting and shaping machines, strength is also an issue.

Since the advent of computers, numerical methods have found an impressive development. These methods are able to analyze linear and nonlinear material behavior, isotropic or anisotropic, large deformations, etc.

Methods for Structural Analysis

Structural analysis can be performed by analytical methods and by numerical methods. Since the generalization of the use of computers, the latter methods are the most common.

Analytical Methods

Only geometrically simple systems with linear material behavior tend to be calculable by analytical methods. Roark' formula for stress and strain (Young and Budynas 2001) is one classical source of information for these structures. When an analytical method can be applied, the calculation can be done in a much simpler and faster way than when numerical methods are applied. Furthermore, the use of computers is not required, although this is not an issue nowadays.

Numerical Methods

Numerical methods perform calculations of structures by considering the structural space divided in sufficiently small elements, where the material constitutive laws are applied in simplified ways. In some cases, numerical methods are able to exactly represent the equations, while in other cases, they do not; in these latter cases, the size of the elements into which the structure was divided must be analyzed for the required calculation accuracy to be obtained (Clough and Penzien 1975; Hurty and Rubinstein 1964).

Computer Methods for Large Problems

There are several numerical methods that can be implemented in computers. The oldest method is probably the Finite Difference Method, which currently has been almost completely substituted by the Finite Element Method (Zienkiewicz and Taylor 2000). An alternative to the latter can be the Boundary Element Method.

The Finite Element Method is based on modelization of the displacement field of the structural points by a reduced set of displacements at a certain number of nodes. The structure is divided in elements, which are geometrically defined by the coordinates of some nodes. The coordinates of the internal points of the element, as well as their displacements, are interpolated from the coordinates and the displacement of the nodes. Then, by minimization of the potential energy at the element, a stiffness matrix of the element, relating the forces applied at the nodes and the displacements, is obtained.

The complete structural stiffness matrix is obtained by assembling the matrices of each of the elements. Then, the displacement field, as well as the stresses, can be calculated.

For dynamic problems, besides the stiffness matrix, it is necessary to calculate the inertia matrix and the damping matrix. The inertia matrix is calculated in a similar way to the stiffness matrix, while the damping matrix cannot be obtained in an accurate way at the current state of the art.

Boundary elements are based on the use of interpolating functions whose volume integration reduces to the integration at the surfaces. In this way, the number of elements and of nodes/degrees of freedom is considerably reduced. One of the drawbacks of the boundary element method is that it gives rise to fully populated matrices. Real size of the matrix tends to rise with the square of the size of the problem, while in finite element, the size tends to rise linearly with the size of the problem.

Currently, machines structural analysis is almost always based on the Finite Element Method.

Matricial Methods, Variable Change

Numerical methods are usually considered by using a matricial representation

$$\{f\} = [K]\{\delta\} \tag{2}$$

or an energy representation

$$\frac{1}{2}\{\delta\}^T\{f\} = \frac{1}{2}\{\delta\}^T[K]\{\delta\}, \tag{3}$$

where f represents the vector of externally applied forces, and δ the vector of displacements at the nodes. K is the stiffness matrix.

This means that the work done by the external forces equals the deformation energy.

Matrix K can be obtained by direct methods for simple parts, for example, for rods or beams, and typically by the Finite Element Method for more complex shaped parts.

Once the equation is worked out, the displacements at the nodes considered are known, but the calculation of the stresses needs further work, as it is necessary to come back to the formulation of each of the parts (elements) considered in the discretization.

When dynamics have to be considered, for example, when variable forces are applied into the structure, the inertia and damping forces have to be included in the formulation:

$$\{f\} = [K]\{\delta\} + [C]\left\{\dot{\delta}\right\} + [M]\left\{\ddot{\delta}\right\} \quad (4)$$

where C is the damping matrix and M the inertia or mass matrix.

Here the mass and damping matrices are added to the equation. Mass matrices are obtained in a straightforward manner by the Finite Element Method or by other techniques, but in the current state of the art, the way of obtaining approaches for the damping matrix is not sufficiently established.

A very useful operation when dealing with structural analysis is the variable change. This operation is used when different coordinate systems are required, or when size reduction is needed. Variable change can be represented in matricial form as

$$\{\delta\} = [T]\{u\}, \quad (5)$$

where u are the new variables, and T is the transformation matrix. The new matrices are obtained by pre- and postmultiplication by the transformation matrix in the form:

$$[Kn] = [T]^T[K][T] \quad (6)$$

$$[Cn] = [T]^T[C][T] \quad (7)$$

$$[Mn] = [T]^T[M][T] \quad (8)$$

The force vector has also to be projected into the new coordinate system:

$$\{fn\} = [T]^T\{f\} \quad (9)$$

The transformed system becomes:

$$[T]^T\{f\} = [T]^T[K][T]\{u\} + [T]^T[C][T]\{\dot{u}\} + [T]^T[M][T]\{\ddot{u}\} \quad (10)$$

Modeling of Structures

Simplifications

The power of computers and the development of Finite Element Method Systems allow dealing with very detailed discretization of the structures. Nevertheless, there is always the need to simplify the modeling of some components or the constraints. Good engineering criterion is required to perform a good structure model and to correctly interpret the results obtained. The quality of the model cannot be measured at all by the size of the elements when using the Finite Element Method, or by the number of degrees of freedom, as the most important errors will come from the inadequate simplifications done.

In machine tools, the representation of bearings, guiding pads, floor behavior, etc. is the critical aspect. Also, the structural parts which are not included in the model have to be considered in a wise way.

Considerations on Damping

Damping is a very important factor for the dynamic behavior of any structure. Although statically, or at excitation frequencies far from the natural frequencies, the rigidity and the inertia terms dominate the equilibrium equation, at frequencies close to natural frequencies, these terms cancel each other, and damping is the only factor able to withstand the external forces applied to the structure.

There are many phenomena giving rise to loads that are a function of the delayed structural vibration. These loads can produce self-excited vibration of the structure, like in the flutter of airplane wings or the chatter of machine tools. Damping is the most important structural property to impede that kind of vibration.

Currently, there are no established methods for estimation of damping at the design stage. The engineers have to make an experience-based

guess, or measure it when the prototype is built. This is a very undesirable situation, as the size of the structure required for a particular application can well be mainly dependent on the required stability against self-excited vibration, directly dependent on the damping that the structure will have.

Material damping tends to be very low, damping of a complete system being produced mainly at the mechanical interfaces. It is known that the interface surface's roughness and preload have an influence on the damping, but there is no method to predict the amount of damping that will be obtained as a function of these parameters. It seems also that damping has a nonlinear behavior.

Methods for Size Reduction

Structural models reach easily very large dimensions, many times counted by millions of degrees of freedom. In many circumstances, it is required to reduce the size of the matrices used for the calculation, especially when dealing with dynamic problems.

A common way of dealing with very large problems is substructuring, where the complete structure is divided into substructures. The matrices modeling each of the substructures can be reduced (condensed) before the matrices representing the complete structure are assembled.

For static analysis, a straightforward technique is the static condensation of the stiffness matrix. For dynamic analysis, several approaches can be used, none of them being mathematically exact.

Static Condensation

Condensation (reduction) of the stiffness matrix for the resolution of a static problem is a mathematically exact procedure. Let us assume that the degrees of freedom are ordered in such a way the first degrees of freedom are to be maintained and the last to be deleted. If subindex "c" denotes the degrees of freedom to be maintained, and "e" those to be deleted, the static problem can be stated as:

$$\begin{bmatrix} K_{cc} & K_{ce} \\ K_{ec} & K_{ee} \end{bmatrix} \begin{Bmatrix} \delta_c \\ \delta_e \end{Bmatrix} = \begin{Bmatrix} f_c \\ f_e \end{Bmatrix} \tag{11}$$

The displacements to be deleted from equation can be calculated after the degrees of freedom to be maintained:

$$\delta_e = k_{ee}^{-1}(f_e - K_{ec}\delta_c) \tag{12}$$

After substitution, we get:

$$\left[K_{cc} - K_{ce}K_{ee}^{-1}K_{ec}\right]\delta_c = f_c - K_{ce}K_{ee}^{-1}f_e \tag{13}$$

Dynamic Condensation

Whenever a dynamic calculation is required, matrix size reduction is usually necessary. The reasons are the high number of calculations to be performed for any time integration procedure, and the very short time integration step requested by a complete representation of the system (very high natural frequencies are present).

Dynamic condensation cannot be performed in a mathematically exact way. The approximation is done by establishing a linear relation between the complete variable set and the reduced variable set. The accuracy of the approximation depends on the capability of the reduced variable set to represent the complete structure behavior.

One of the most used methods to reduce the number of unknowns is the Guyan condensation (Guyan 1965; Irons 1965), in which it is assumed that, if the maintained displacements are known, the condensed displacements are calculated as the static response to the maintained displacements. The dynamic equilibrium equation can be stated as:

$$\begin{bmatrix} K_{cc} & K_{ce} \\ K_{ee} & K_{ec} \end{bmatrix} \begin{Bmatrix} \delta_c \\ \delta_e \end{Bmatrix} + \begin{bmatrix} C_{cc} & C_{ce} \\ C_{ec} & C_{ee} \end{bmatrix} \begin{Bmatrix} \dot{\delta}_c \\ \dot{\delta}_e \end{Bmatrix} + \begin{bmatrix} M_{cc} & M_{ce} \\ M_{ec} & M_{ee} \end{bmatrix} \begin{Bmatrix} \ddot{\delta}_c \\ \ddot{\delta}_e \end{Bmatrix} = \begin{Bmatrix} f_c \\ f_e \end{Bmatrix} \tag{14}$$

Taking the rows corresponding to the degrees of freedom to delete ("e"), and considering that there are no external forces on these degrees of freedom:

$$[K_{ec}\ \ K_{ee}]\begin{Bmatrix}\delta_c\\ \delta_e\end{Bmatrix} = \{0\} \tag{15}$$

Then

$$\delta_e = -K_{ee}^{-1}K_{ec}\delta_c, \tag{16}$$

and

$$\begin{Bmatrix}\delta_c\\ \delta_e\end{Bmatrix} = \begin{bmatrix} I \\ -K_{ee}^{-1}K_{ec}\end{bmatrix}\{\delta_c\}. \tag{17}$$

Therefore, by applying the variable change

$$\begin{aligned}&\left\lfloor K_{cc} - K_{ce}K_{ee}^{-1}K_{ec}\right\rfloor\{\delta_c\} + \\ &\left\lfloor C_{cc} - C_{ce}K_{ee}^{-1}K_{ec} - K_{ee}^{-1}K_{ce}C_{ec} + K_{ee}^{-1}K_{ce}C_{ee}K_{ee}^{-1}K_{ec}\right\rfloor\{\delta_c\} + \\ &+\left\lfloor M_{cc} - M_{ce}K_{ee}^{-1}K_{ec} - K_{ee}^{-1}K_{ce}M_{ec} + K_{ee}^{-1}K_{ce}M_{ee}K_{ee}^{-1}K_{ec}\right\rfloor \\ &\{\delta_c\} = \left\{f_c - K_{ee}^{-1}K_{ec}f_e\right\}\end{aligned} \tag{18}$$

again a reduced size matricial system is obtained. In this case, the reduction is not mathematically exact.

Although there are very deep analyses on the accuracy of the solution and on the best way to select the reduced set of equations, for an engineer, the criterion for choosing a good set of degrees of freedom to maintain is to think of the structure as constrained at these maintained displacements; that structure should not have modes/natural frequencies in the range of frequencies that are going to be analyzed, or close to it. Then, the analysis will be accurate.

Modal Analysis

A well-known technique for dynamic analysis of structures is the Modal Analysis. It consists of obtaining the natural frequencies and their associated modeshapes. Mathematically, modal analysis is performed by obtaining the eigenvalues and eigenvectors of the homogeneous equation:

$$([K] + \lambda[M])\{\phi\} = \{0\} \tag{19}$$

The eigenvalues l are all negative. The eigenvectors diagonalize the stiffness and the mass matrices:

$$[\phi]^T[M][\phi] = \text{diagonal matrix} \tag{20}$$

$$[\phi]^T[K][\phi] = \text{diagonal matrix} \tag{21}$$

To start we will assume that damping is proportional, that is, C = a1.K + a2.M. Then, the eigenvectors also diagonalize the damping matrix:

$$[\phi]^T[C][\phi] = \text{diagonal matrix} \tag{22}$$

The sizes of the eigenvectors are not important; they would continue fulfilling the fundamental equation. Therefore, there is freedom in choosing their size. The size of the eigenvector is usually characterized by its modal mass, that is, the product

$$m_m = \{\phi\}^T[M]\{\phi\} \tag{23}$$

The most common way of selecting the size of the eigenvectors is to obtain unity modal mass. Therefore,

$$[\phi]^T[M][\phi] = \langle I\rangle \tag{24}$$

$$[\phi]^T[K][\phi] = \langle -\lambda\rangle = \langle \omega_n^2\rangle \tag{25}$$

$$[\phi]^T[C][\phi] = -2 = \langle \alpha_n\rangle \tag{26}$$

where ω are the natural frequencies of the system, α represent the damping factors, and the notation $<>$ is used to denote a diagonal matrix.

When the variable change

$$\{\delta\} = [\phi]\{q\} \tag{27}$$

is applied, we obtain

$$\begin{aligned}&[\phi]^T[K][\phi]\{q\} + [\phi]^T[C][\phi] \\ &\quad + [\phi]^T[M][\phi]\{q\} \\ &\quad = [\phi]^T\{f\}\end{aligned} \tag{28}$$

or

$$\langle \omega_n^2\rangle\{q\} - 2\langle\alpha\rangle\left\{\dot{q}\right\} + \langle I\rangle\{\ddot{q}\} = [\phi]^T\{f\}, \tag{29}$$

where q is the modal content of the modes.

Therefore, by applying this coordinate transformation, the system becomes diagonal, which is very interesting for calculation speed.

Also, it is possible to concentrate the calculation only at the frequency range at which we are interested, usually the low frequency range. The time integration can be performed with much larger time steps (linearly dependent on the highest natural frequency).

For each degree of freedom (natural coordinate):

$$\omega_m^2 q_m - 2\alpha_m \dot{q}_m + \ddot{q}_m = \phi_m^T f \tag{30}$$

Applying Fourier transform:

$$\begin{aligned} &\omega_m^2 Q_m - 2j\omega\alpha_m Q_m - \omega^2 Q_m \\ &= \left(\omega_m^2 - 2j\omega\alpha_m - \omega^2\right) Q_m = \phi_m^T F \end{aligned} \tag{31}$$

where the capital letters represent the Fourier transform of the time signal denoted by its corresponding lower case letter.

From (31), it is possible to calculate the modal content of each mode:

$$Q_m = \frac{\phi_m^T F}{\omega_m^2 - 2j\omega\alpha_m - \omega^2} \tag{32}$$

The displacement at any point due to the content of this natural coordinate is:

$$\delta_{mi} = \phi_{mi} \frac{\phi_m^T F}{\omega_m^2 - 2j\omega\alpha_m - \omega^2} \tag{33}$$

The total displacement of the point is obtained by adding up the contribution of all the modes:

$$\Delta_i = \sum_m \phi_{mi} \frac{\phi_m^T F}{\omega_m^2 - 2j\omega\alpha_m - \omega^2} \tag{34}$$

By considering the effect of the force at one degree of freedom only, the transfer function between both points is obtained:

$$\Delta_i = \sum_m \frac{\phi_{mi}\ \phi_{mj} F_j}{\omega_m^2 - 2j\omega\alpha_m - \omega^2} \tag{35}$$

or

$$\frac{\Delta_i}{F_j} = \sum_m \frac{\phi_{mi}\ \phi_{mj}}{\omega_m^2 - 2j\omega\alpha_m - \omega^2} \tag{36}$$

The fractions can be also expressed in the way

$$\frac{\Delta_i}{F_j} = \sum_m \frac{1}{m_{mij}\left(\omega_m^2 - 2j\omega\alpha_m - \omega^2\right)}, \tag{37}$$

where *m mij* is the reduced mass of the *m*'th mode between degrees of freedom *i* and *j*. Very often the reduced mass is incorrectly called "modal mass."

Use of transfer functions is a very convenient way to contrast experimental and theoretical results. Several MAC (Modal Assurance Criterion) exist to quantitatively perform this comparison.

Nonproportional Damping

The general case, where damping is nonproportional, is solved by using the equation

$$\begin{aligned} &\begin{bmatrix} K & 0 \\ 0 & -M \end{bmatrix} \begin{Bmatrix} \delta \\ \dot{\delta} \end{Bmatrix} + \begin{bmatrix} C & M \\ M & 0 \end{bmatrix} \begin{Bmatrix} \dot{\delta} \\ \ddot{\delta} \end{Bmatrix} \\ &= \begin{Bmatrix} f \\ 0 \end{Bmatrix}, \end{aligned} \tag{38}$$

which represents exactly the original system, but has a lower order at the cost of having twice the number of unknowns (displacements and velocities).

In this case, the homogeneous equation is:

$$\left(\begin{bmatrix} K & 0 \\ 0 & -M \end{bmatrix} + \lambda \begin{bmatrix} C & M \\ M & 0 \end{bmatrix} \right) \{\phi\} = \{0\} \tag{39}$$

The eigenvalues and eigenvectors λ come out to be complex. The eigenvalues have the shape $\alpha + j\upsilon$, where α is the damping of the mode (negative value), and υ is the damped natural frequency. As the matrices have real coefficients, the solutions come in pairs of complex conjugates.

The eigenvectors contain modal displacement in the first half, and modal velocities in the second. The modal velocities have a phase shift with regard to the modal displacements: Each degree of freedom reaches its maximum displacement at different times, opposite to the case of proportional damping.

Experimental Modal Analysis

Modal Analysis can be done experimentally, by applying controlled forces with high frequency content and measuring the displacement (usually by integration of the signal of accelerometers). Modal Analysis is a very convenient way of comparing the theoretical results with the actual behavior of the machine or component. Figures 1 and 2 show respectively one natural mode shape of a milling machine calculated by the finite element method and the corresponding mode shape obtained from experimentation.

Component Mode Synthesis

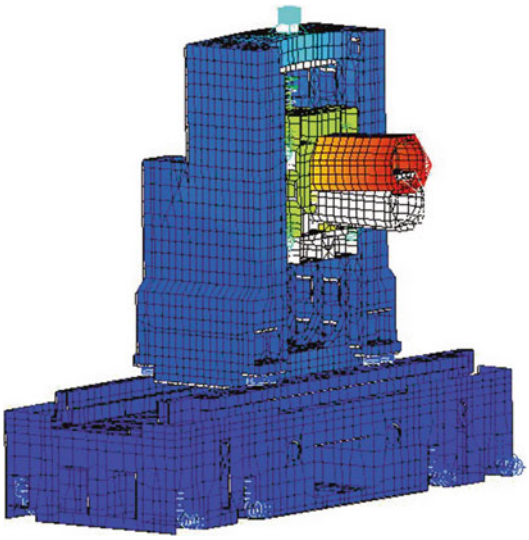
A very practical alternative to the calculation of the complete system and further size reduction of the degrees of freedom consists in the independent calculation of substructures, followed by the join calculation of the complete system.

For the join calculation to be done, a reduced coordinate set of each of the substructures is considered. There are many approaches for the selection of the reduced set, but the two most classical are the "free boundary" and the "fixed boundary" methods (Craig and Kurdila 2006).

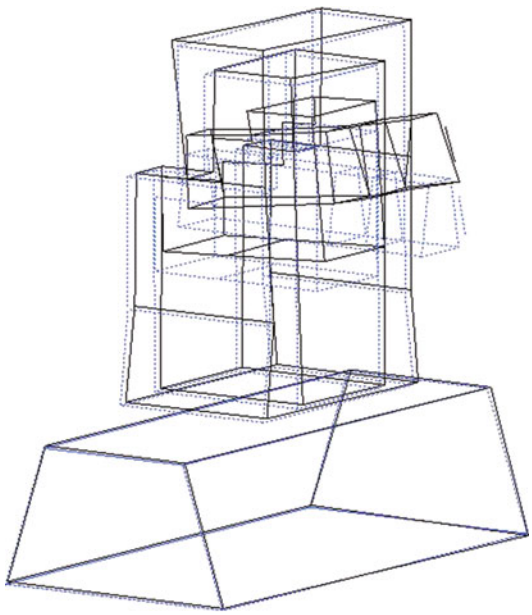
Free Boundary Component Mode Synthesis

In this method, the degrees of freedom of the structure are reduced to the 6 rigid body displacements plus the lowest natural frequency modeshapes. The great advantage of this method is that the reduced mass and stiffness matrices are diagonal (Hurty and Rubinstein 1964).

The method tends to produce results more rigid than the real value. That is because the low frequency modes do not include the deformation of the areas close to the borders. For example, a beam modal analysis would produce natural modes without bending energy at the ends for the lowest frequency modes. When that beam is



Structural Analysis, Fig. 1 Natural modeshape calculated by the finite element method



Structural Analysis, Fig. 2 Natural modeshape obtained by experimental modal analysis

attached to other components, the ends will most probably suffer the highest bending deformation energy, and therefore, the reduced coordinate set will not be able to represent that energy. As a consequence, a huge number of natural modes have to be considered to get a reasonable precision.

One way of improving the accuracy of the method is the inclusion of the "residual" flexibility, that is, the flexibility at the borders that is not considered by the modes included in the model (Rubin 1975). With this correction, the method provides much better results, at the cost of much higher computation and programming effort.

Fixed Boundary Component Mode Synthesis

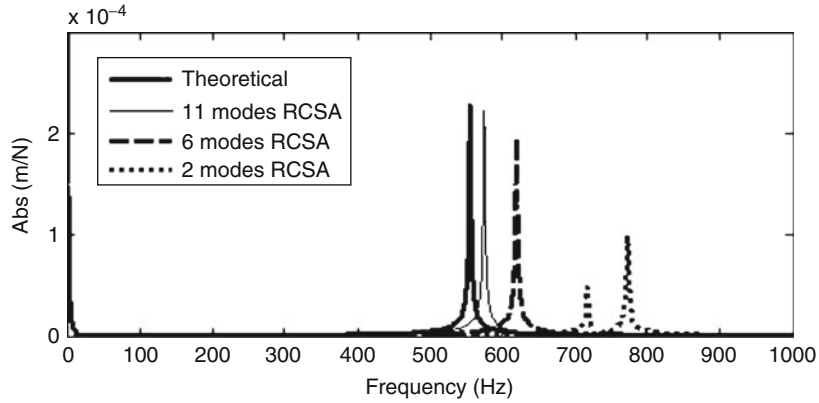
This technique uses as maintained degrees of freedom the natural modes calculated with the connection nodes blocked. Additional deformation modes that have to be included in this method consist of the displacement field obtained when each one of the connection degrees of freedom is submitted to a unitary displacement while the rest is blocked (Craig and Bampton 1968).

This method is a little more complicated to program than the free boundary method, but its precision is much better. Just a few natural modes are necessary usually, with figures as low as one or two.

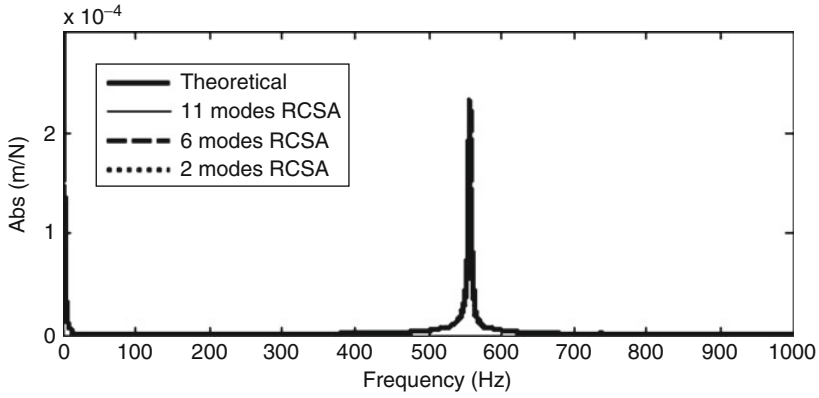
Results of Component Mode Synthesis Methods

Fixed boundary method for mode synthesis provides very good approximations with low number of modes, while free boundary method requires that a very large number of modes be included. The reason is that when the structures are modeled independently considering free boundaries, the flexibility of the areas close to the boundaries do not appear up to very high frequency modes, as they do not suffer from the inertia efforts because they do not have mass around. Opposite to that, when fixed boundaries are used, the flexibility of the regions close to the boundaries appears already since the very first modes, those with lowest natural frequency.

As an example, both mode synthesis methods were applied to calculate the joint behavior of a tool mounted to the main spindle of a machine tool (Mancisidor et al. 2011). Figure 3 shows the transfer function obtained when two, six, and 11 modes are included in the free boundary approach. It is clear that even with 11 modes considered, the approximation is not very accurate. Figure 4 shows the transfer functions for the same system with the same number of modes. Considering two modes only is sufficient to have high accuracy. In fact, for this case, fixed boundary method provided high accuracy even when one mode only was considered.



Structural Analysis, Fig. 3 Frequency response function obtained by free boundary component mode synthesis (RCSA = Receptance Coupling Substructure Analysis)



Structural Analysis, Fig. 4 Frequency response function obtained by fixed boundary component mode synthesis (RCSA = Receptance Coupling Substructure Analysis)

Cross-References

- ▶ Bearing
- ▶ Chatter
- ▶ Damping
- ▶ Dynamics
- ▶ Finite Element Method
- ▶ Stability
- ▶ Vibration

References

Clough R, Penzien J (1975) Dynamics of structures. McGraw-Hill, New York

Craig R, Bampton M (1968) Coupling of substructures for dynamic analysis. AIAA J 6:1313–1319

Craig RR, Kurdila AJ (2006) Fundamentals of structural dynamics, 2nd edn. Wiley, Hoboken

Guyan R (1965) Reduction of mass and stiffness matrices. AIAA J 3(2):380

Hurty W, Rubinstein M (1964) Dynamics of structures. Prentice Hall, Englewood Cliff

Irons B (1965) Structural eigenvalue problem: elimination of unwanted variables. AIAA J 3:961–962

Mancisidor I, Zatarain M, Munoa J, Dombovari Z (2011) Fixed boundaries receptance coupling substructure analysis for tool point dynamics prediction. Adv Mater Res 223:622–631

Rubin S (1975) Improved component-mode representation for structural dynamic analysis. AIAA J 13(8):995–1006

Young W, Budynas R (2001) Roark's formulas for stress and strain, 7th edn. McGraw-Hill, New York

Zienkiewicz O, Taylor R (2000) The finite element method. Heinemann, Butterworth, Oxford

Structural Behavior

- ▶ Thermal Error

Sublimation

- ▶ Laser Ablation

Super Precision

- ▶ Ultraprecision

S

Superabrasives

Barbara Linke
Mechanical and Aerospace Engineering,
University of California Davis, Davis, CA, USA

Synonyms

Cubic boron nitride (CBN); Diamond; High-performance abrasives

Definition

The term "superabrasives" denotes abrasive grit materials with higher hardness (commonly higher than 4,000 HK) and refers to cubic boron nitride and diamond. Sometimes the term is also used to describe the tools made of these abrasives. Besides higher hardness, superabrasives are also characterized by higher wear resistance standing out from "conventional abrasives" like corundum and silicon carbide.

Theory and Application

Superabrasives are distinguished from conventional abrasives by their higher hardness, which is accompanied by higher Young's modulus and heat conductivity but typically lower thermal stability (Table 1).

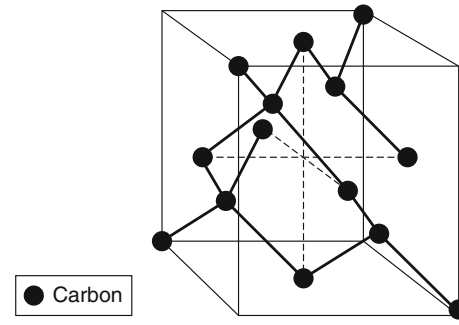
Diamond

Diamond is the hardest material in nature and highly resistant against compaction (Knoop hardness of 7,000–8,000). This is based on the dense structure of the carbon atoms, their regular, symmetrical order, and the energy-rich covalent atom bonds. Every carbon atom in diamond is surrounded uniformly by four atoms (Fig. 1). The angle between two neighboring atoms amounts consistently to 109.5°. The crystal morphology of diamond can range from a perfect cube to a perfect octahedron (O'Donovan 1976).

The genesis of natural diamond takes place at high pressures and temperatures within the Earth's mantle. The source of natural diamond is the primary influence of its chemical and physical properties. Today, natural diamond is mainly used in dressing tools and has been replaced by synthetic diamond in grinding tools.

The first artificial diamond synthesis was conducted by the Swedish company ASEA with a six-anvil press in February 1953 (Jackson and Davim 2011). In 1955, General Electric (GE) followed with the synthesis in a belt press and in 1958 De Beers (Marinescu et al. 2007). Metals of the eighth main group (nickel, cobalt, iron, etc.) enable diamond synthesis at pressures of 5–8 GPa and temperatures of 1,500–2,100 K (O'Donovan 1976).

Diamond density is about 3.52 g/cm^3 depending on pureness. Synthetic diamonds expose metal inclusions of used catalysts. All crystal defects such as substituted atoms and atoms between lattice sites or lattice vacancies are imperfections of diamond structure and enable



Superabrasives, Fig. 1 Structure of diamond (Bailey and Juchem 1998)

Superabrasives, Table 1 Physical properties of abrasive materials (Klocke 2009; Toenshoff and Denkena 2013; Rowe 2009)

		Hardness	Young's modulus	Heat conductivity	Thermal stability
Superabrasives	Diamond	5,000–8,000 HK	890 GPa	600–2,100 W/m K	Up to 900 °C
	Cubic boron nitride	4,500–5,000 HK	680 GPa	200–1,300 W/m K	Up to 1,370 °C
Conventional abrasives	Silicon carbide	2,400–3,000 HK	400 GPa	55–100 W/m K	Up to 1,500 °C
	Corundum	1,600–2,000 HK	400 GPa	6–35 W/m K	Up to 2,000 °C

micro-splintering (O'Donovan 1976). Small grits are often tougher than large diamonds because they have fewer and smaller defects and inclusions (Field 1979).

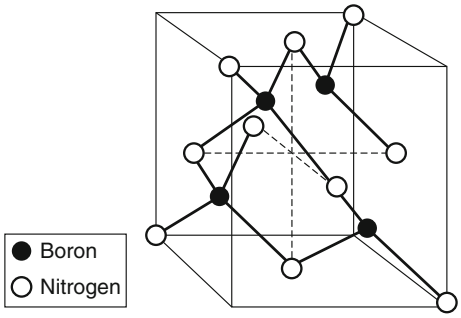
The hardness of single crystal diamond is anisotropic and depends on the crystal orientation. This results from the different distances of the carbon atoms in different crystal planes. Moreover, the cleavage behavior of diamond is defined by the density of atom bonds in the different diamond planes. Diamond has four cleavage planes (Gardinier 1988).

Diamond has a high affinity to the metals of the eighth main group as well as to Mn, Cr, Ra, and Nb. At temperatures above 800 °C, diamond burns with oxygen and becomes carbon dioxide. Depending on the grit size, i.e., the specific surface area, and the grit crystal type, reactions with oxygen can occur already at 500–700 °C (Gardinier 1988). Graphitization is a commonly known state transition of diamond to graphite. In the presence of oxygen, the diamond surface can graphitize already at temperatures of 900 K (Field 1979). The thermal stability of diamond affects how grinding tools are processed, e.g., within inert atmosphere.

At room temperature, diamond is the material with the highest known temperature conductivity with a value between 600 and 2,000 W/(m*K) depending on crystal purity (Marinescu et al. 2007). To prevent destruction of the grit/bond interface, often coatings of the diamond grits with nickel, cobalt, or composite metals are applied (Klocke 2009). Coatings enhance heat capacity and facilitate heat conducting into the tool bonding.

CBN

The invention of cubic boron nitride (CBN) is closely linked to the synthesis of artificial diamond. Cubic boron nitride synthesis was conducted first in 1957. CBN crystals are produced from boron and nitrogen at high pressures of 50–90 kbar, high temperatures between 1,800 °C and 2,700 °C, and in the presence of a catalyst (Klocke 2009). During the first years on the market, CBN was seen as a competitor to diamond. However, CBN proved to be a better material for machining of hard-to-machine ferrous materials than diamond due to the missing chemical affinity and the higher thermal stability.



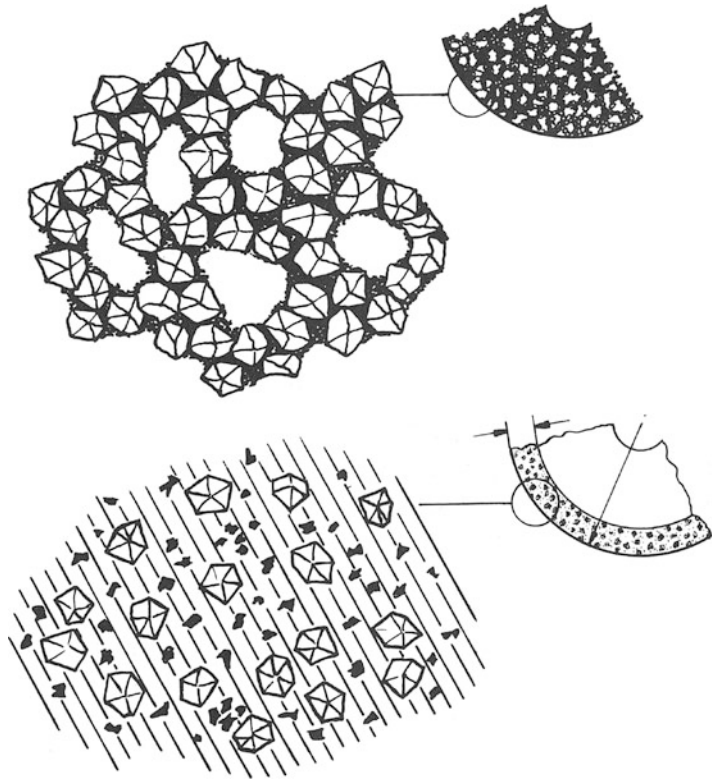
Superabrasives, Fig. 2 Structure of CBN (Bailey and Juchem 1998)

CBN and diamond have similar structures. In CBN, each nitrogen atom binds to four boron atoms and vice versa, forming the bond angle of 109.5° (Fig. 2). In CBN, covalent bonds are predominant with a small degree of ionic bonding which results from the fact that boron and nitrogen are dissimilar atoms (Bailey and Juchem 1998). Therefore, cubic boron nitride is less symmetric than diamond. Therefore, its morphology is more complex. CBN has six cleavage planes and the crystal shapes range from octahedron to tetrahedron and cubo-octahedron (Gardinier 1988).

CBN has a Knoop hardness of around 4,700 (Gardinier 1988). It is thermally stable in inert atmosphere at temperatures up to 1,500 °C. In air, CBN forms a stable layer of boron oxide preventing oxidation up to 1,300 °C (Rowe 2009). Although this layer dissolves in water under heat, CBN is successfully used with water-based cooling lubricants. It is presumed that the minimal contact times in grinding prevent the reaction with water (Klocke 2009).

Tools with Superabrasives

In contrast to conventional grinding tools, superabrasive tools are commonly built from an abrasive layer applied to a carrier; a so-called body (Fig. 3). The carrier has to provide sufficient heat conductivity, high mechanical strength, and good vibrational dampening. Common body materials are aluminum, steel, bronze, synthetic resin with



Superabrasives, Fig. 3 Conventional wheel versus superabrasive wheel (*bottom*) (Metzger 1986) (With copyright permission from publisher)

metallic or nonmetallic fillers, and fiber-reinforced synthetic resin or ceramics (Klocke 2009). Design of carriers has to regard expansion at rotation speed, damping behavior, thermal expansion, etc.

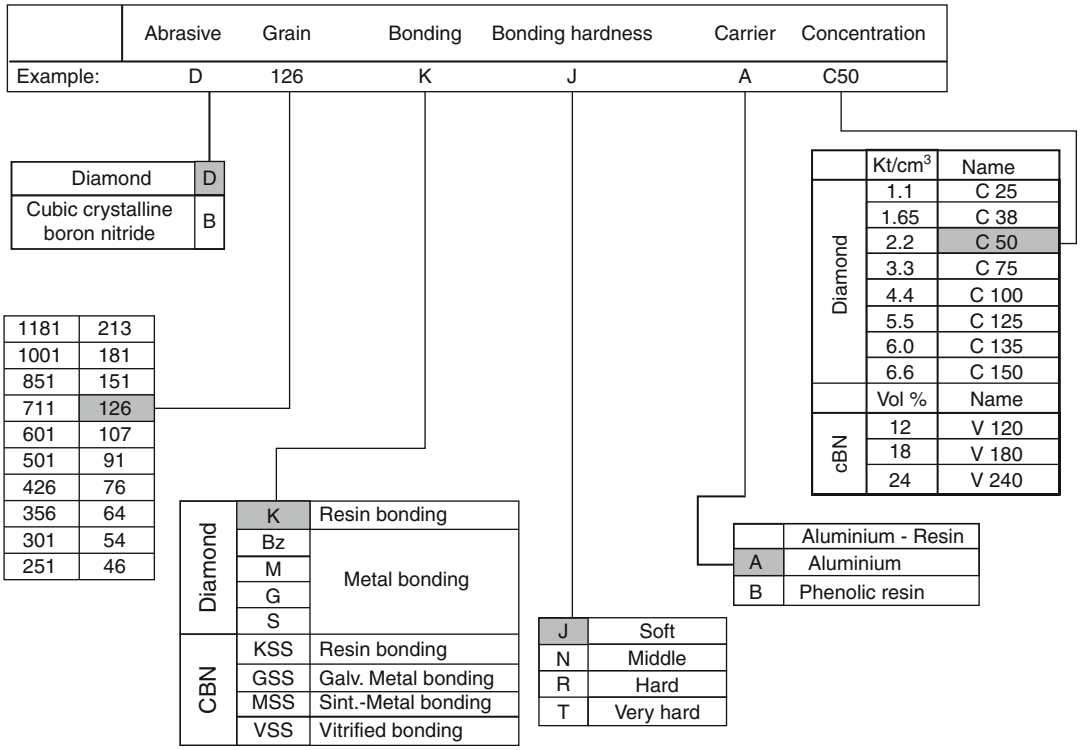
Like conventional grits, superabrasives can be held by resin, vitrified, or metallic bonds. Resin and vitrified bonds, however, have to be adapted to the chemistry and performance of superabrasives. Metallic bonds are particularly important for superabrasive grits and are nearly used exclusively for this grit type. Sintered metallic bonds, such as bronze, produce abrasive segments with several layers of grits. Electroplating or brazing affix a single layer of superabrasives to the carrier (Marinescu et al. 2007). Electroplated CBN played a major role in the development of high-efficiency deep grinding (HEDG) (Rowe 2009).

Figure 4 shows an example for the specification of bonded grinding tools with superabrasives. The common abbreviation for diamond grits is (D) and (B) for CBN (Fig. 5). The size of superabrasive grits is often given as mean diameter in micrometer instead of mesh-like conventional abrasives. Therefore, the specification of grinding wheels has to be compared carefully between conventional abrasives and superabrasives. Concentration is given in carat per volume for diamond tools or in volumetric percentage for CBN tools (Klocke 2009).

Applications and Grinding Performance

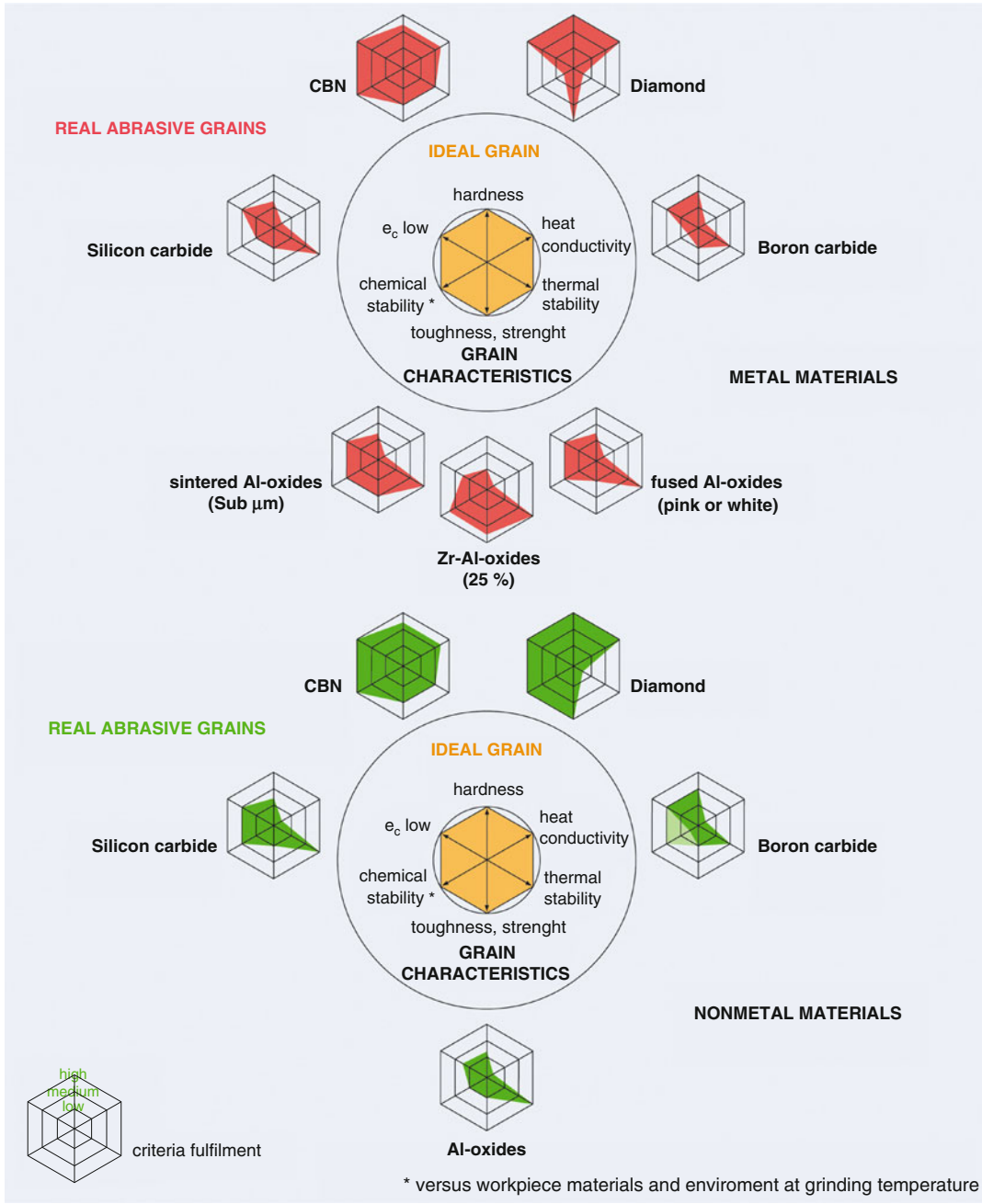
In general, corundum and CBN are used for long-chipping, ductile materials, whereas silicon carbide and diamond are used for short-chipping, brittle materials, or titanium alloys (Fig. 6; Klocke 2009). Superabrasives are chosen in particular for the higher precision or higher performance applications due to their low wear rate and ability to hold close size tolerances (Rowe 2009).

Figure 6 summarizes the most important grit characteristics in grinding technology. The reactivity of diamond with transition metals such as nickel and iron limits the use of diamond to machine these metals, especially steels. However, there are some abrasive applications with ferrous materials where diamond is the tool material of choice, e.g., honing of cast iron (Marinescu et al. 2007).



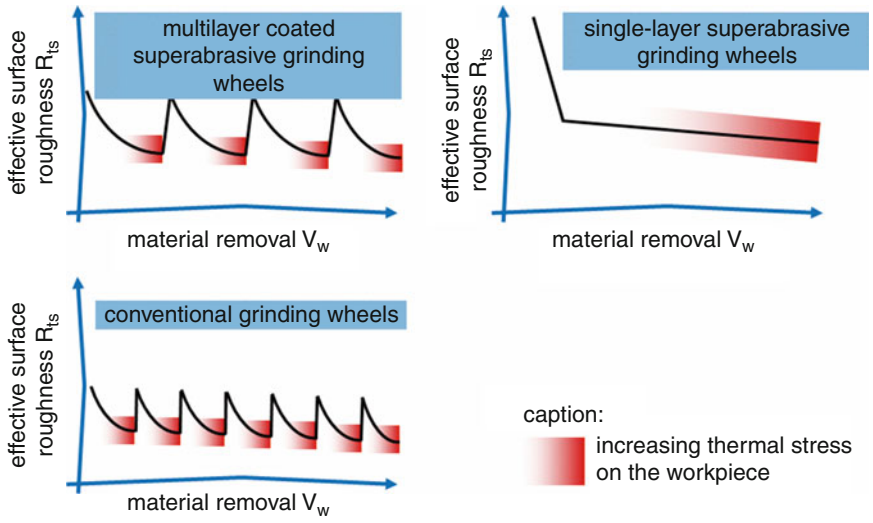
Superabrasives, Fig. 4 Example of notation of superabrasives after Saint-Gobain Diamantwerkzeuge GmbH & Co. KG, (Klocke 2009) (With copyright permission from publisher)

Superabrasives, Fig. 5 Classification of abrasive grains with standardized letter symbols and application areas, (Helletsberger et al. 2011) (With copyright permission from publisher)



Superabrasives, Fig. 6 Comparison of abrasives at machining of ductile material (*top*) or brittle material (*bottom*), (Helletsberger et al. 2011) (With copyright permission from publisher)

Superabrasives are in particular highly resistant to wear when used at high grinding wheel speeds. However, choosing superabrasives as grinding tool material should follow a thorough evaluation of the higher tool costs and the requirements on machine tool and cooling lubricant supply.



Superabrasives, Fig. 7 Schematic wear behavior of different grinding wheel types (With copyright permission from Prof. Klocke, RWTH Aachen University)

The higher thermal conductivity of superabrasives compared to conventional abrasives can reduce grinding temperatures drastically (Rowe 2009). However, there is a common understanding that the surface finish is rougher with CBN grits than with conventional wheels. The reason is believed to be the sharper and more pointed cutting edges leading to earlier chip formation at a shallower initial depth (Malkin and Guo 2008).

Multilayered superabrasive tools have longer dressing intervals than conventional tools (Fig. 7). Single-layered tools show an initial wear phase at the beginning of their use. This is followed by a quasi-stationary behavior until the tool end of life is defined by thermal damage of the workpiece (Klocke 2009). Single-layered tools are not profiled or sharpened in the traditional sense, although sometimes so-called touch-dressing is applied to level protruding edges.

Dressing

Conditioning or dressing of superabrasive tools is challenging because of the high grit wear resistance. Often, lower toughness and hardness of CBN compared to diamond allows successful dressing of CBN grinding wheels with rotating diamond tools (Marinescu et al. 2007). Nevertheless, the dressing forces for CBN are higher than for conventional wheels, which can affect the requirements on the dressing system (Jackson and Davim 2011). Dressing is also done with much smaller depth of cut than for conventional wheels (Wegener et al. 2011).

Dressing of diamond wheels with diamond dressing tools is still limited. A more common method is the use of vitrified bonded silicon carbide rollers either on brake-controlled truing devices or on driven truing spindles (Wegener et al. 2011). During this dressing procedure, the expendable and much cheaper silicon carbide wheel grinds the diamond tool (Malkin and Guo 2008).

Superabrasive grinding wheels with resin or metal bond sometimes require a subsequent sharpening process after the profiling process to generate sufficient grain protrusion (Wegener et al. 2011). The bonding can be set back with a block sharpening process.

An established method for dressing metal-bonded diamond wheels is the electrolytic in-process dressing (ELID) (Rowe 2009). Additionally, more dressing procedures for metal bonds using electrochemical and electrophysical mechanisms exist (Wegener et al. 2011).

Cross-References

- ▶ Abrasive Material
- ▶ Dressing
- ▶ Grinding Wheel

References

Bailey MW, Juchem HO (1998) The advantages of CBN grinding: low cutting forces and improved workpiece integrity. IDR 3:83–89

Field JE (1979) The properties of diamond. Academic, London

Gardinier CF (1988) Physical properties of superabrasives. Ceram Bull 67(6):1006ff

Helletsberger H, Huber W, Larch C (2011) Grindology movie GM2 – grinding stock removal. Tyrolit Grindology College, Tyrolit Schleifmittelwerke Swarovski KG, Schwaz

Jackson MJ, Davim JP (2011) Machining with abrasives. Springer, New York. https://doi.org/10.1007/978-1-4419-7302-3

Klocke F (2009) Manufacturing processes 2 – grinding, honing, lapping (RWTH Edition) (trans: Kuchle A). Springer, Berlin/Heidelberg

Malkin S, Guo C (2008) Grinding technology: theory and application of machining with abrasives, 2nd edn. Industrial Press, New York

Marinescu ID, Hitchiner M, Uhlmann E, Rowe WB, Inasaki I (2007) Handbook of machining with grinding wheels. CRC Press, Boca Raton

Metzger JL (1986) Superabrasive grinding. Butterworth, Oxford, UK

O'Donovan KH (1976) Synthetische Diamanten. Fertigung 2(76):41–48

Rowe WB (2009) Principles of modern grinding technology. William Andrew, Norwich

Toenshoff HK, Denkena B (2013) Basics of cutting and abrasive processes. Springer, Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-33257-9

Wegener K, Hoffmeister H-W, Karpuschewski B, Kuster F, Hahmann W-C, Rabiey M (2011) Conditioning and monitoring of grinding wheels. CIRP Ann Manuf Technol 60(2):757–777. https://doi.org/10.1016/j.cirp.2011.05.003

Superfinishing

Hitomi Yamaguchi
Mechanical and Aerospace Engineering,
University of Florida, Gainesville, FL, USA

Synonyms

Microhoning

Definition

Superfinishing is a process in which a rotating workpiece is finished by a relatively soft stone with fine abrasive oscillating parallel to the workpiece surface. The stone typically oscillates at an amplitude of 1–4 mm and a frequency of 10–50 Hz, and it contacts the workpiece at light pressure, typically 0.1–0.2 MPa. The superfinishing process consists of three phases: (1) the *cutting phase*, which is characterized by a high material removal rate due to sharp cutting edges; (2) the *transition phase*, which is characterized by a decrease in the material removal rate due to dulling and loading of the stone; and (3) the *finishing phase* in which dulling and loading results in only slight or no material removal. In the subsequent superfinishing process, the loaded stone contacts the rough surface of the next workpiece to initiate self-dressing of the stone – providing sharp cutting edges – and the three phases described are repeated. In comparison with honing and lapping, superfinishing exhibits higher finishing efficiency (Matsui and Nakasato 1965; Kawamura et al. 1989; Farago 1980).

Theory and Application

History

The process was originally developed to eliminate brinell marks from automobile wheel bearing surfaces that resulted from long-distance shipping. When automobile wheels were blocked to prevent rolling during shipping, only a few of the rollers or balls in the wheel bearings supported the weight of the automobile. Shipping vibration caused the hardened rollers or balls to press against the bearing components and form slight depressions: a process called *brinelling*. When the cars were put into service, the depressions caused excessive noise. It was found that the brinelling could be eliminated by removing the "fuzz" generated in the grinding of the load-carrying surfaces of the bearings. The development of fuzz-removing processes led to the development of the superfinishing process, and early in 1934, the principles of superfinishing were first conceived by

D. A. Wallace of the Chrysler Corporation (Swigert 1940).

Wallace describes his development as follows:

> *Superfinishing is the name of a method of mechanically developing on metal parts a surface finish which is optically smooth and metallurgically free of any fragmented or smear metal, such as is created by the dimensional operations of turning, grinding, honing, lapping and/or burnishing. The superimposing of this process over previous machining operations removes the defective boundary layer material and exposes, for heavy duty load-carrying contact, the unworked and undisturbed crystalline base metal. The resultant Superfinished surface is a true, geometrically developed, wear-proof bearing area, free of oil-film rupturing protuberances, and accurate to within submicroscopic range.* (Swigert 1940)

Theory

Abrasive Stone

A superfinishing abrasive stone is generally vitrified bonded and has a hardness range of HRH20–70. White aluminum oxide, silicon carbide, cubic boron nitride, or diamond abrasive is used. Abrasive sizes in the JIS#300–#500 range are used for coarse finishing, while JIS#600–#1500 abrasives are used for a fine finish (Matsui and Nakasato 1965; Onchi et al. 1995; Varghese and Malkin 1998).

Finishing Mechanism

Figure 1 shows a schematic of the superfinishing of a cylindrical workpiece. The stone follows a sinusoidal path as a result of the workpiece rotation combined with the stone oscillation in the workpiece axial direction. The cutting velocity v (m/min) and inclination angle θ at any point P are calculated with the following equations:

$$v = \sqrt{{v_r}^2 + (a\omega \cos \omega t)^2} \tag{1}$$

$$\theta = tan^{-1}(a\omega \cos \omega t / v_r) \tag{2}$$

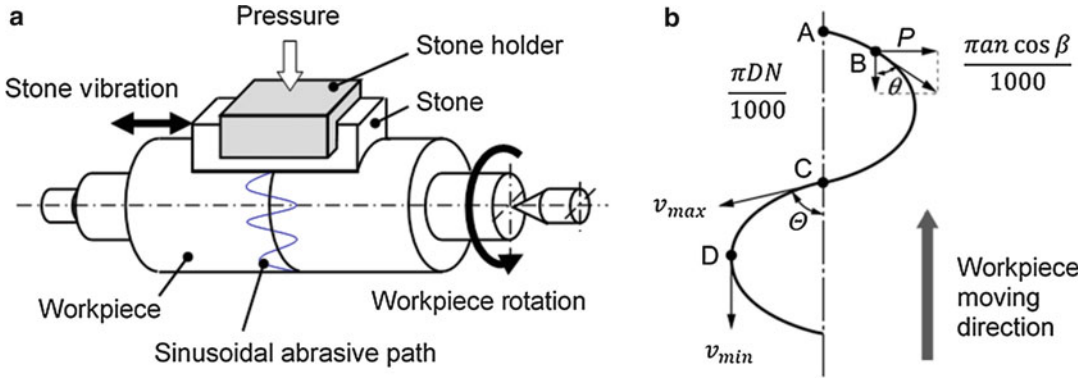
where v_r is the workpiece peripheral velocity (m/min), a is the amplitude of the stone oscillation (m), ω is the angular velocity, and t is the finishing time (min).

The cutting velocity v is maximum (v_{max}) at A and C and minimum (v_{min}) at B and D. The maximum cutting velocity and maximum inclination angle Θ are calculated as:

$$v_{max} = \sqrt{{v_r}^2 + (a\omega)^2} \tag{3}$$

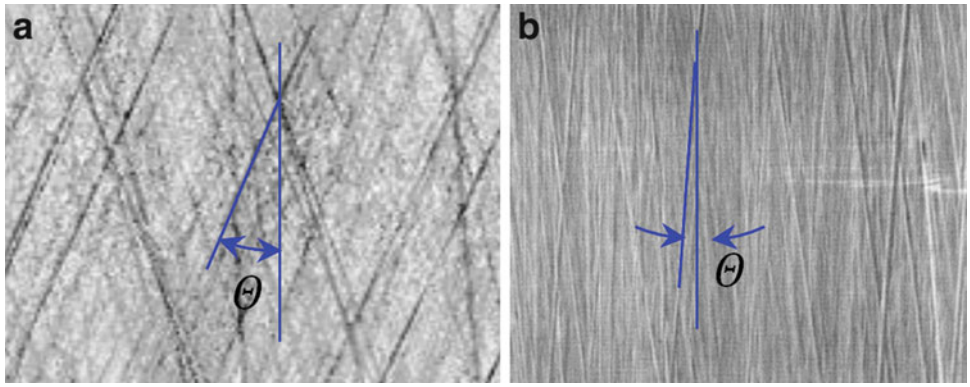
$$\Theta = tan^{-1}(a\omega / v_r) \tag{4}$$

A large maximum inclination angle Θ results in continuous and significant changes in the direction of the stone. This encourages self-dressing of the stone to maintain sharp abrasive cutting edges and contribute to the high material removal rate. Conversely, when the maximum inclination angle Θ is small, the self-dressing action is reduced, and the loading of the stone is increased, which results in a reduced material removal rate. Figure 2 shows representative hardened steel (SUJ-2) surfaces superfinished with (a) large inclination angle θ and (b) small inclination angle θ.



Superfinishing, Fig. 1 Schematics of the superfinishing process: (**a**) processing principle, (**b**) sinusoidal abrasive path

S



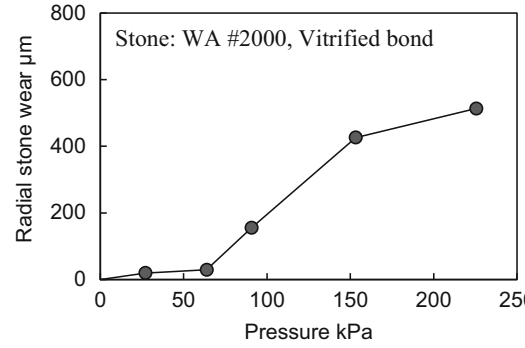
Superfinishing, Fig. 2 Surfaces processed with superfinishing: (**a**) large inclination angle, (**b**) small inclination angle (Courtesy of Dr. Hashimoto)

The wear characteristics of the stone are also affected by finishing pressure. If the finishing pressure exceeds a critical pressure, stone wear is drastically accelerated as the stone fractures or abrasive is released. Although this increases the material removal rate, it hardly achieves a fine finish. Accordingly, it is important to maintain the superfinishing pressure below the critical pressure. If two-phase finishing is needed, it is recommended that the initial pressure be slightly over the critical pressure to facilitate the self-dressing action of the stone and then reduce the finishing pressure to slightly below the critical pressure.

Figure 3 shows a case study in which the radial stone wear is plotted against the finishing pressure. In this case, the stone wear drastically changes around 60 kPa, which is determined to be the critical pressure. In this case, therefore, the finishing pressure should be set a little below 60 kPa to achieve fine finishing.

Asada's work (which was reported others) showed that the critical pressure in superfinishing is influenced as follows (Matsui and Nakasato 1965):

(a) The critical pressure decreases with finer abrasive stones.
(b) The critical pressure increases as the bond strength increases.
(c) The critical pressure decreases as the maximum inclination angle Θ increases.



Superfinishing, Fig. 3 Changes in radial stone wear with finishing pressure

(d) The critical pressure is generally not influenced by the average cutting velocity.
(e) The critical pressure decreases with decreasing workpiece material hardness.

Superfinishing improves surface roughness (producing surfaces smoother than 0.2 µm *Ra*) and forms accuracy (e.g., out of roundness). The material removal mechanisms in superfinishing are considered to be similar to those in grinding and honing (Puthanangady and Malkin 1995). As grinding and honing do, superfinishing produces an affected layer. In the superfinishing of hardened steels, the layer thickness is about 1 µm, which is about seven times thinner than the layer produced in grinding (Asaeda 1952). The specific energy is in a range of 50–700 J/mm^3 (Chang et al. 1997), and it is made up of cutting and

sliding actions. The specific energy due to cutting consists of chip formation and plowing components. The plowing component is dominant (over 80 % of the total cutting energy) when material removal rate is low. When the material removal rate is high, the ratio of the plowing component to the chip-formation component is smaller (Puthanangady and Malkin 1995).

Key Applications

Bearings
Automotive components (crankshafts, camshafts, transmission shafts, gears, pistons, cam follower rollers, etc.)
Medical devices (orthopedic implants) (Chang et al. 1997)

Cross-References

▶ Honing

References

Asaeda T (1952) Superfinishing. J Metal Finish Soc Japan 3(5):165–168 (in Japanese)
Chang SH, Balasubramhanya S, Chandrasekar S, Farris TN, Hashimoto F, Show MC (1997) Forces and specific energy in superfinishing of hardened steel. Ann CIRP 46(1):257–260
Farago FT (1980) Abrasive methods engineering. Industrial Press, New York
Kawamura S, Yano A, Higuchi M, Sugita T (1989) *Kakogaku Kiso 2, Kensakukako to Toryukako* (Fundamentals of machining technologies 2, Grinding and abrasive technologies), 5th edn. (in Japanese). Kyoritu Shuppan, Tokyo, pp 182–190
Matsui M, Nakasato S (1965) Choshiagesagyo to Sono Genri (Superfinishing processes and their principles) (in Japanese). Yokendo, Tokyo
Onchi Y, Matsumori N, Ikawa N, Shimada S (1995) Porous fine CBN stones for high removal rate superfinishing. Ann CIRP 44(1):291–294
Puthanangady TK, Malkin S (1995) Experimental investigation of the superfinishing process. Wear 185: 173–182
Swigert AM (1940) The story of superfinish. Lynn Publishing, Detroit
Varghese B, Malkin S (1998) Experimental investigation of methods to enhance stock removal for superfinishing. Ann CIRP 47(1):231–234

Superhard Cutting Material

▶ Ceramic Cutting Tools

Superhard Tools

Eckart Uhlmann[1], Fiona Sammler[3], John Barry[2], Javier Fuentes[3] and Sebastian Richarz[3]
[1]Fraunhofer Institute for Production Systems and Design Technology, Berlin, Germany
[2]Advanced Materials, Element Six Ltd, Shannon, Clare, Ireland
[3]Berlin, Germany

Synonyms

Cutting tools; Diamond; PcBN

Definition

Superhard materials for defined edge tooling may be defined as materials with a hardness exceeding 3,000 HV.

Introduction

Superhard materials also exhibit other properties which are favorable in tooling materials such as high thermal conductivity, low coefficients of thermal expansion, and low coefficient of friction on most materials. All commercial superhard materials fall within the two broad categories of diamond or cubic boron nitride (CBN). Diamond is well known as the hardest material, with single crystal diamond exhibiting hardness values of approximately 10,000 HV. Diamond materials however can be most conveniently categorized according to their microstructure/structure and corresponding method of manufacture. The five categories of "synthetic" diamond materials can thus be identified as:

- Polycrystalline diamond or "PCD" is produced by sintering micron diamond powders under ultrahigh pressure (>5 GPa) in the presence of a metal "catalyst" such as cobalt. These

materials are commercially available in a range of grain sizes and typically are manufactured as disks (50–75 mm in diameter) with a thin (0.5 mm) layer of PCD bonded to a cemented carbide substrate. A "residue" of metal in the microstructure renders the materials electrically conductive, thereby greatly facilitating cutting tool production.

- Polycrystalline thick-film diamond is produced by chemical vapor deposition. These materials are free-standing diamond layers of between 0.2 and 1 mm in thickness and comprise columnar grains of pure diamond.
- Single crystal diamond is produced by ultrahigh pressure synthesis from carbon using a metal catalyst. Commercial production methods produce cubo-octahedral crystals of several mm to 10 mm in dimension, which are sliced into regular geometric plates for the production of ultraprecision cutting tools.
- Single crystal diamond is produced by chemical vapor deposition method. Regular geometrical plates of such materials are commercially available with edge lengths up to 10 mm.
- Thin-film diamond and diamond-like coatings are deposited on finish-ground cemented carbide tools such as end-mills and drills. The main benefit of CVD diamond coatings is that there is no subsequent grinding or finishing operation on the tool necessary.

In addition to the above five categories of synthetic diamond, natural single crystal diamond is still used to some extent in ultraprecision machining applications. All single crystal applications require the diamond to be precisely orientated so as to take advantage of the extreme anisotropy in the diamond lattice. Synthetic diamonds offer the advantage of having their crystal orientation aligned to the edges of cuboids/plates.

Boron nitride is an entirely synthetic material (with no natural occurrence) and the cubic crystal structure (cBN) is synthesized from the softer hexagonal (HBN) allotrope using ultrahigh pressures as in the case of diamond. CBN crystals exhibit less than half the hardness of diamond, but there are no industrial applications for such materials. Instead, all industrially relevant materials are polycrystalline in nature, produced by sintering micron cBN powder together with a variety of ceramic phases. Thus, polycrystalline cBN (PcBN) tooling materials may be classified most generally by their cBN content, with most practitioners distinguishing between high-content (>80%) and low-content PcBN. This distinction is useful also in terms of the relevant application areas. All commercial PcBN materials are produced in the form of disks (50–95 mm in diameter) with or without a hard metal substrate. Some of the most notable properties of polycrystalline PcBN tooling materials, in comparison with other cutting materials, are the high hot hardness, up to 1,200 °C, and the high chemical resistance. These properties are not only useful for withstanding the relatively hot chip forming temperatures in the cutting zone, but also make PcBN suitable for the machining of hardened steels. The tendency of the toughness of this material to increase as the hardness decreases does not occur with the same intensity as it does in other materials, such as cemented carbides or ceramics. PcBN is therefore an effective tool material for dry machining of high-performance materials (König and Hiding 1994; Clark and Sen 1998; Reuter 2001).

Table 1 compares the properties of superhard bulk materials (excluding thin-film coatings which are not amenable to conventional property

Superhard Tools, Table 1 Indicative properties of superhard materials for defined edge tooling applications

		Hardness vickers (N/mm^2)	Transverse bending strength (MPa)	Thermal conductivity (W/m.K)
Single crystal materials	CBN	~4,000	–	400
	Diamond	5,000–10,000	1,000–3,000	1,500–2,000
Polycrystalline materials	HP PCD	5,000–8,000	1,200–2,300	400–600
	Thick-film CVD Diamond	8,500–10,000	500–1,000	1,000–1,200
	High-content PCBN	~3,500	800–1,500	80–150
	Low-content PCBN	~3,500	700–1,000	40–70

measurements). The values are intended for indication only and may vary between commercial suppliers.

Although superhard materials other than diamond and CBN have been synthesized in recent years – including boron carbon nitride, cubic silicon nitride, and boron sub-oxide, these are presently not of industrial importance as their synthesis conditions are too extreme or their properties, other than hardness, insufficient for application as cutting tools.

Theory and Application

Single Crystal Diamond for Ultraprecision Machining Applications

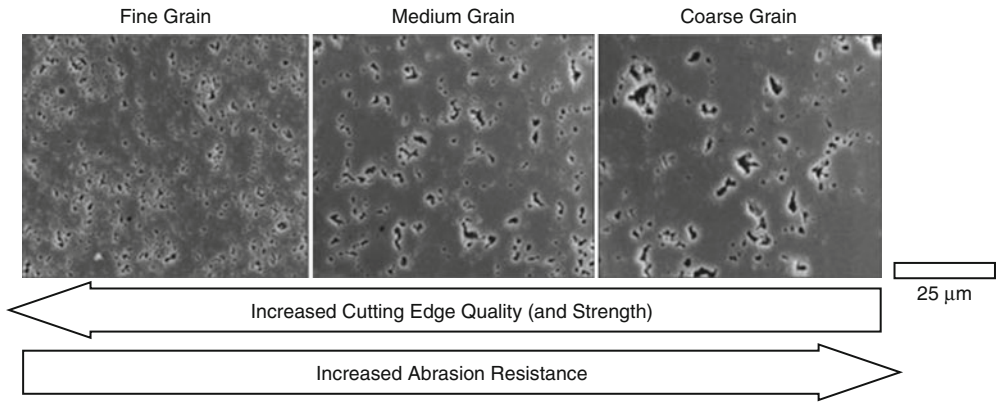
Single crystal diamond is the only practical tool material for ultraprecision machining applications. The extreme hardness of diamond provides for optimum wear resistance, but also enables the generation of near atomically sharp cutting edges (a characteristic which also draws upon the absence of grain boundaries). Diamond exhibits pronounced anisotropy, which is exploited in the production of ultraprecision tools – often, crystals are orientated so the (110) plane is presented to the polishing scaife so as maximize the rate of polishing. In addition to the crystal plane, the direction of polishing is also critical and each of the (100), (110), and (111) planes exhibits difference numbers of "soft" directions.

In addition to the material's extreme hardness, monocrystalline diamond exhibits extremely low coefficients of friction on most metals and nonmetals (of the order of 0.05), resulting in minimal chip adhesion to the cutting edge. A thermal conductivity of up to 2,000 W/m.K and a thermal expansion coefficient of less than 2 ppm/K ensure minimum thermal deflections and deviations.

Such is the relative expense of single crystal diamond tools, they tend to be reserved for applications requiring below 10 nm surface finish and submicron form accuracies. The generation of optical components such as mirrors, lenses, and lens molds from materials such as aluminum, OFHC, silicon, electroless nickel, and germanium is a relatively common application of monocrystalline diamond tools. Certain large volume manufacturing processes use MCD tools, namely, precision finishing of laminate flooring and finishing of acrylic aircraft windows.

Polycrystalline Diamond Tooling Materials

The vast majority of diamond tools in use in industry are manufactured from ultrahigh pressure sintered PCD. Grades of commercial importance vary in grain size from 1 to 30 μm, with finer grain materials exhibiting higher strength and, generally, better surface finish capabilities relative to coarse grain materials. The latter, however, tend to exhibit greater wear resistance. The general relationships are illustrated in Fig. 1. The pores evident in each micrograph are where the cobalt



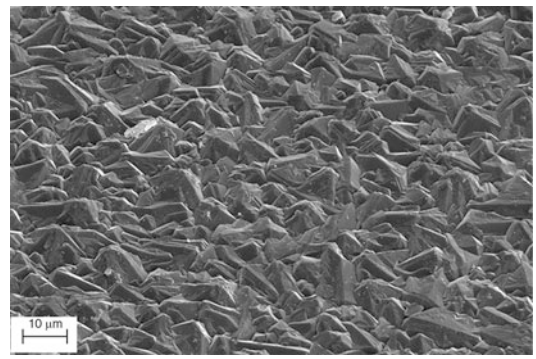
Superhard Tools, Fig. 1 General characteristics for fine, medium and coarse grained PCD materials (© element SDC)

metal used to aid sintering normally resides (it has been etched to reveal the intergrowth of diamond grains). Although in effect a residue from the manufacturing process, the cobalt phase, serves to render the material electrically conductive as well as imparting a significant degree of toughening. The ability to electrically conduct is of great advantage in commercial tool production where EDM technology is used to rough and finish-machine PCD cutting tools (Tönshoff 2011).

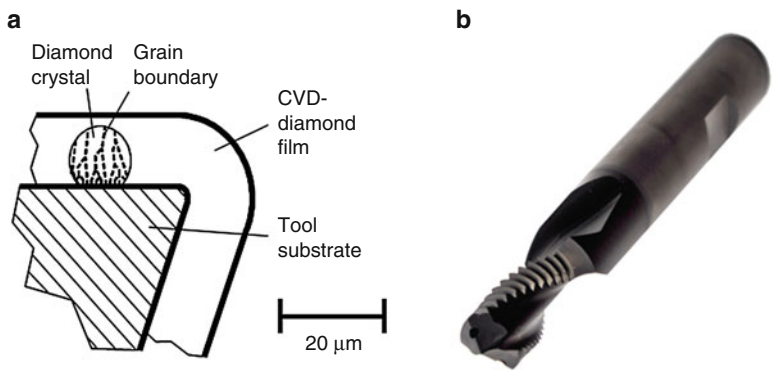
PCD tooling materials are usually employed in very large volume manufacturing processes such as automotive engine and gearbox manufacturing, wheel turning, and wood machining. In such applications, tool lives may be 10–1,000 h and it is generally found that PCD will exceed cemented carbide tool lives by at least two orders of magnitude. PCD is also employed in the machining of fiber-reinforced composites, cemented carbides, and cast irons (precision finishing) (Kalpakjian and Schmid 2001).

Diamond can also be produced via Chemical Vapor Deposition (CVD). Using this process, diamond is deposited from the gas phase directly onto a substrate. CVD diamond can be deposited as a thin-film tool coating directly onto the tool substrate (Fig. 2) or as a thick-film insert, similar to PCD. A diamond thin film is usually no thicker than 40 µm whereby thick films typically have thicknesses of 0.2–1 µm. CVD diamond can be deposited in the form of nanocrystalline, microcrystalline, or multilayer films, referring to the size of the diamond crystals in the film which can be generated through the selection of particular process parameters during the CVD coating process. The properties of these film types differ with regard to the surface roughness, hardness, oxidation resistance, and elasticity.

For example, the surface roughness of a nanocrystalline film is very low and can be implemented for high precision surface finishing operations. The surface roughness of a microcrystalline diamond film is higher due to the larger diamond crystals (Fig. 3); however, these coatings possess a higher hardness than the nanocrystalline films. The appropriate diamond film must therefore be chosen depending on the application of the tools. Further types of CVD diamond film are doped films. Both thick and thin films have doped variants for the purpose of providing electrical conductivity. CVD diamond has undergone significant development in recent years and can now challenge PCD tools in many applications



Superhard Tools, Fig. 3 Microcrystalline CVD diamond thin film on a cemented carbide substrate (Source: Institute for Machine Tools and Factory Management (IWF), Technische Universität Berlin)



Superhard Tools, Fig. 2 CVD diamond thin film (**a**) schematically and (**b**) deposited on a cemented carbide thread milling drill (Source: (**a**) VDI 2841 2008, (**b**) Institute for Machine Tools and Factory Management (IWF), Technische Universität Berlin)

(Clark and Sen 1998; VDI 3824 2002; VDI 2840 2005; VDI 2841 2008).

CVD diamond coated tools are implemented in the machining of lightweight materials in the automotive and aeronautical sectors, for example, in the machining of aluminum silicon and aluminum lithium alloys. These alloys challenge the cutting material due to the hard particles in the soft aluminum matrix and depending on the percentage of hard particles, the tools exhibit adhesive or abrasive wear as well as coating delamination.

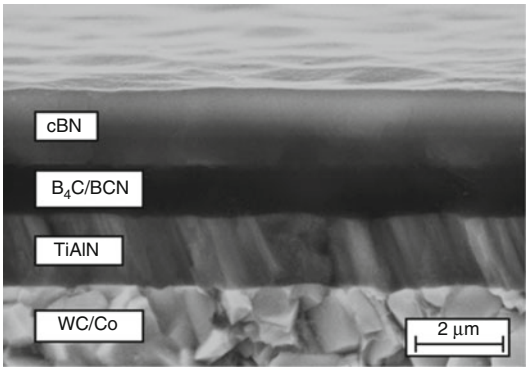
Although diamond is one of the most inert materials, it will rapidly graphitize in contact with ferrous metals at temperatures above 1,100 K, and as such, diamond is only by exception used in the machining of ferrous metals (and typically with relatively low cutting speeds). PcBN tools, however, have a greater inertness in the presence of ferrous metals at high temperature.

Polycrystalline Cubic Boron Nitride (PcBN)

Single crystal cBN is not of industrial importance for cutting applications as it offers few benefits over monocrystalline diamond and is intrinsically more difficult to synthesize. As such, all industrial cBN tools are composite materials prepared using powder metallurgical techniques, but sintered under similar conditions to those used for the synthesis of diamond and cBN from their softer allotropes (graphite in the case of diamond).

The majority of PCBN tooling materials contain 40–70% cBN with the remainder being primarily TiC, TiCN, or TiN ceramic. These tooling materials exhibit excellent abrasion resistance and chemical wear resistance and are mostly used in the machining of hardened steels. Grades with lesser amounts of cBN tend to be used for continuous turning operations, while grades for interrupted turning and milling more commonly contain 60–70% cBN. Despite being somewhat counterintuitive, thin PVD ceramic coatings applied to PcBN tools can increase tool lives or enable operation at higher cutting speeds. Approximately half of all PcBN tools in use today are coated.

High-content PcBN grades (containing >80% cBN) are used for machining of cast irons, heavy turning, and milling of hardened steels and for finish-machining of powder metal components. These materials tend to have better thermal shock resistance and strength but exhibit more rapid crater wear in comparison to low-cBN grades.



Superhard Tools, Fig. 4 CBN coating on a cemented carbide substrate with interlayers to increase coating adhesion (Source: Uhlmann et al. 2009)

Due to the fact that PcBN composites exhibit only moderate strength values, yet are used to machine many of the highest strength metallic work materials, virtually all tools are employed with negative cutting geometries, and cutting edge chamfers of 20–30° are commonly employed, as are edge hones (cutting edge radii). This is in contrast to diamond tools, which invariably are used in a neutral or positive cutting geometry.

Research is currently being undertaken on the topic of cubic boron nitride (cBN) tool coatings (Fig. 4). These coatings possess high hardness and wear resistance and can be used in the machining of unalloyed, alloyed, and hardened steels as well as nickel-based alloys. For example, cBN-coated cutting inserts with a chip breaker geometry were used in the machining of nickel-based alloys and compared with TiAlN-coated inserts. At a cutting speed of $v_c = 50$ m/min, the cBN-coated inserts exhibited a tool life of 15 min, a 100% longer tool life than that of TiAlN-coated inserts (Uhlmann et al. 2009).

Cross-References

► Cutting, Fundamentals

References

Clark IE, Sen PK (1998) Fortschritte bei der Entwicklung ultraharter Schneidstoffe [Proceedings in the development of ultra-hard cutting materials]. Ind Diam Rundsch 32(4):274–284 (in German)
Kalpakjian S, Schmid SR (2001) Manufacturing engineering and technology, 4th edn. Prentice Hall, Upper Saddle River Jersey, pp 571–585
König W, Hiding M (1994) Feindrehen und Bohren gehärteter Stahlwerkstoffe [Fine turning and drilling of hardened steel]: Fortschrittliche Werkzeugtechnologie, Grob- und Feinbearbeitung harter Eisenwerkstoffe mit PKB-Schneidstoffen: Grundlagen. Technische Information der De Beers Industrie Diamanten (in German)
Reuter U (2001) Verschleißmechanismen bei der Bearbeitung von Gusseisen mit PCBN-Schneidstoffen [Wear mechanism of PcBM cutting material for the machining of cast iron]. Dissertation, TU Darmstadt, pp 3–9 (in German)
Tönshoff HK (2011) Cutting, fundamentals. In: CIRP encyclopedia of production engineering. Springer, Heidelberg
Uhlmann E, Oyanedel Fuentes JA, Keunecke M (2009) Machining of high performance workpiece materials with cBN coated cutting tools. Thin Solid Films 518:1451–1454
VDI 2840 (2005) Carbon films: basic knowledge, film types and properties. Beuth, Berlin
VDI 2841 (2008) CVD diamond tools: categorisation, production and characterisation. Beuth, Berlin
VDI 3824 (2002) PVD and CVD hard coatings: quality assurance: characteristic profiles and fields of application of hard coatings. Beuth, Berlin

Supersonic Spray

▶ Cold Spray

Supply Chain

▶ Logistics

Supply Chain Management

Paul Schönsleben
Betriebswissenschaft, ETH Zürich, Zürich, Switzerland

Synonyms

Management of the extended enterprise

Definitions

There are several definitions of this term, as there also are several definitions of the term "supply chain." In 2014, two important professional organizations in the field, APICS (the Association for Operations Management, founded 1957; see also www.apics.org) and the Supply Chain Council (SCC, founded 1996; see also www.supply-chain.org), merged. Together, they represent a (if not the) leading professional association for supply chain and operations management. Therefore, it seems adequate to refer to APICS and SCC definitions.

According to APICS, a *supply chain* is "the global network used to deliver products and services from raw materials to end customers through an engineered flow of information, physical goods, and cash" (Blackstone 2013). Thus, a supply chain can include procurement networks, production networks, distribution networks, logistics networks, and service networks.

According to APICS, *supply chain management* (SCM) is "the design, planning, execution, control, and monitoring of supply chain activities with the objective of creating net value, building a competitive infrastructure, leveraging worldwide logistics, synchronizing supply with demand, and measuring performance globally" (Blackstone 2013).

These definitions are generally applied to the *comprehensive* design, manufacturing, and delivery process of a product or a service within and across companies. In competitive markets, the underlying concept perceives a supply chain to be in competition with other supply chains in order to sell its products to potential end customers (users).

Clearly, SCM principles can and should be applied to the other phases of the entire product life cycle, that is, also to the use phase (e.g., the after-sales service network), and also – as already the SCOR model suggests – to the return, recycling, and disposal phase (see Fig. 3 below).

Theory and Application

Figure 1 shows one of the characteristic tasks handled by SCM, namely, the ongoing synchronization of supply with demand in the

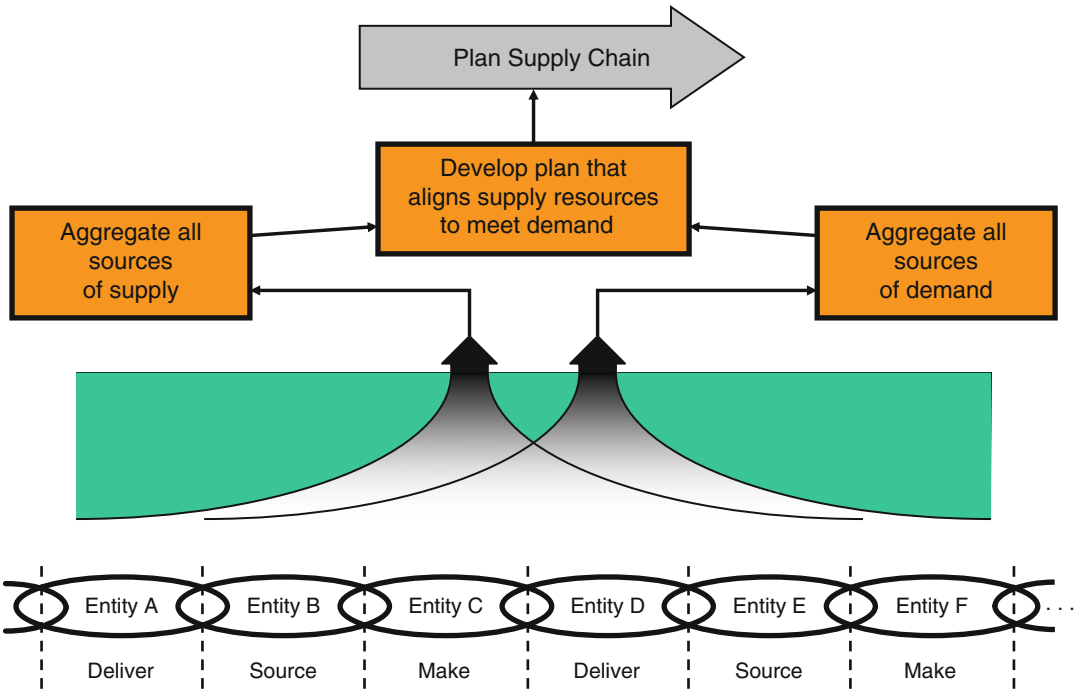
comprehensive supply chain. The organizational units (here called entities) involved can be legally independent companies, as well as profit or cost centers within a company.

Figure 2 groups the tasks in which all organizational units of a supply chain invest in different areas, namely, supply chain structure, supply chain organization, and the required information technology.

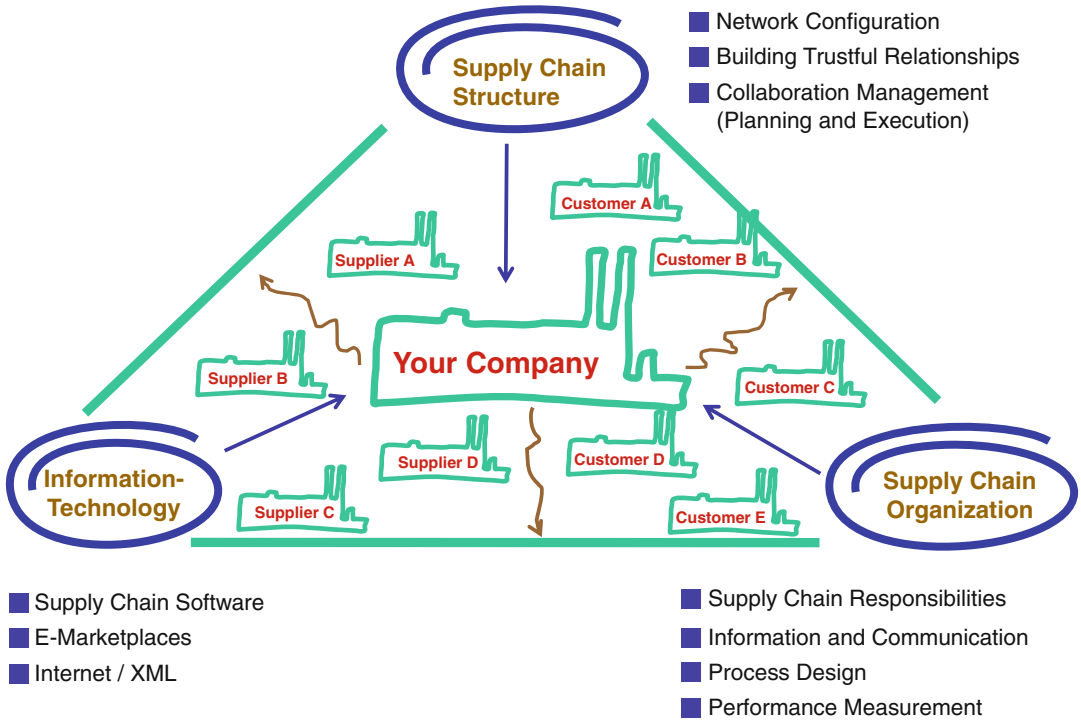
Building Trust and Establishing Principal Relationships

To achieve an effective synchronization, cooperation between the entities of the comprehensive supply chain is essential. In contrast to relationships that are buyer dominated or supplier dominated, in a *balanced* partnership relationship, the intensity of the cooperation can be significantly greater. This demands a high degree of competency in organizing the network, particularly of the leading partner in the supply chain. As a supply chain is in competition with other supply chains, the aim of each company in the supply chain should be to get *a bigger pie*, that is, a bigger share of the market and not *a bigger piece of the pie*, that is, a bigger share of the value added of the supply chain.

In order to cooperate long term and intensively in a supply chain, the following *trust-building measures* have proven significance. *Firstly, create the required conditions in your own company.* This comprises the necessary mentality for a mutual win–win situation, openness to suggestions from internal and external participants, orientation toward procedures and value-adding tasks, and delegation, teamwork, and the like. *Secondly*, w*here possible, place emphasis upon local networks (local sourcing).* This principle is extensively applied by Toyota in Nagoya, where its system suppliers' premises are in a 30 km periphery around the assembly lines of Toyota. Local proximity affects not only logistics favorably (speed, transport, and carrying cost) but also has a particularly favorable effect on relationships among the participants. *Thirdly, do not exploit strengths in your company's negotiating position.*



Supply Chain Management, Fig. 1 Ongoing synchronization of supply with demand in the comprehensive supply chain (Taken from SCOR (2004, p. 22). Copyright APICS Supply Chain Council. Used with permission)



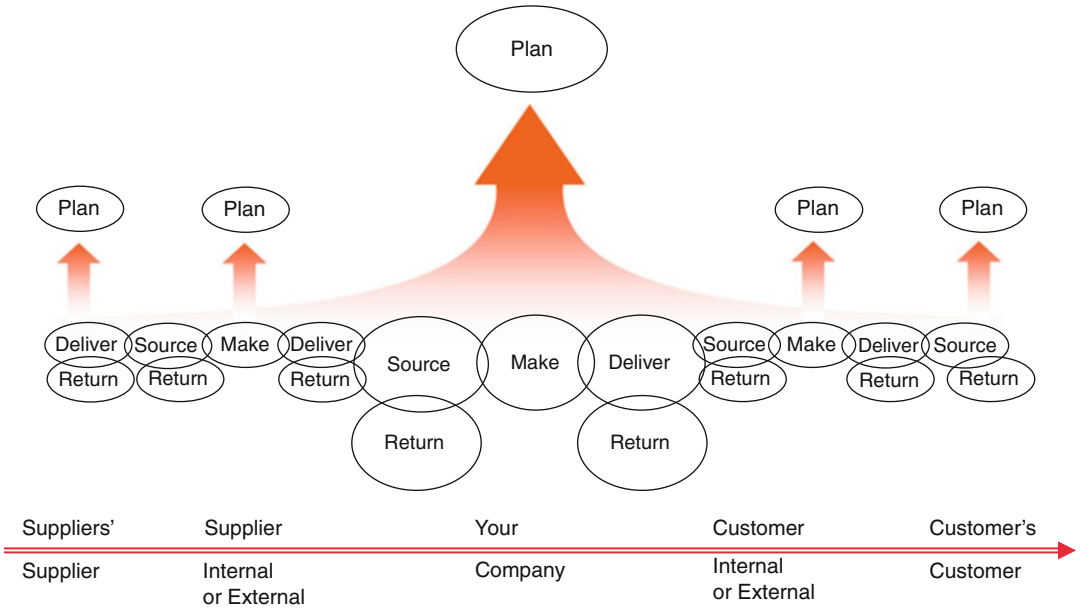
Supply Chain Management, Fig. 2 Tasks and investment areas for intensive cooperation in the partnership relationship (Taken from (Schönsleben 2016), Section 2.3.1)

Present all intentions openly (no hidden agendas). Formulate the objectives of the cooperative venture clearly for all. The primary competition is competition of the entire supply chain against other supply chains for the favor of the user. Competition between buyer and supplier within the supply chain is of secondary importance. The optimum depth of added value of a partner in a supply chain is not necessarily optimal for the *total* supply chain. It is advisable to distribute gains from a cost reduction or increase in earnings equally, because it is the partnership that is the primary factor in success and not the individual contribution of a partner. A balanced win–win situation for all companies involved is prerequisite to long-term or intensive cooperation.

Working Out Collaborative Processes in the Supply Chain

The ongoing synchronization of supply with demand is based upon the internal chain of "source," "make," and "deliver" in each of the entities involved. All demand and capabilities of fulfilling them are carried by the network as a whole and reconciled jointly. Based on this idea, the Supply Chain Council published the supply chain operations reference model or simply SCOR model (SCOR®). SCOR is an aid to standardization of process chains within and across companies. See here (SCOR 2014). The aim of SCOR is to foster a common understanding of processes in the various companies participating in a supply chain. The SCOR model is iteratively updated; the most recent version available only to members. The model is comprised of multiple levels. Level 1 can be seen in Fig. 3, depicted here from the now-superseded version 10.0.

Emphasis is put on commonly agreed performance measurement along the supply chain. The SCOR metrics are made up by five groups of SCM key performance indicators (KPI): firstly, supply chain reliability (perfect order fulfillment); secondly, supply chain responsiveness (order fulfillment cycle time); thirdly, supply chain agility



Supply Chain Management, Fig. 3 The SCOR model, version 10.0, level 1 (Copyright APICS Supply Chain Council. Used with permission)

(upside supply chain flexibility, upside supply chain adaptability, downside supply chain adaptability, overall value at risk (VaR)); fourthly, supply chain costs (SCM costs, cost of goods sold); and fifthly, supply chain asset management (cash-to-cash cycle time, return on supply chain fixed assets, return on working capital).

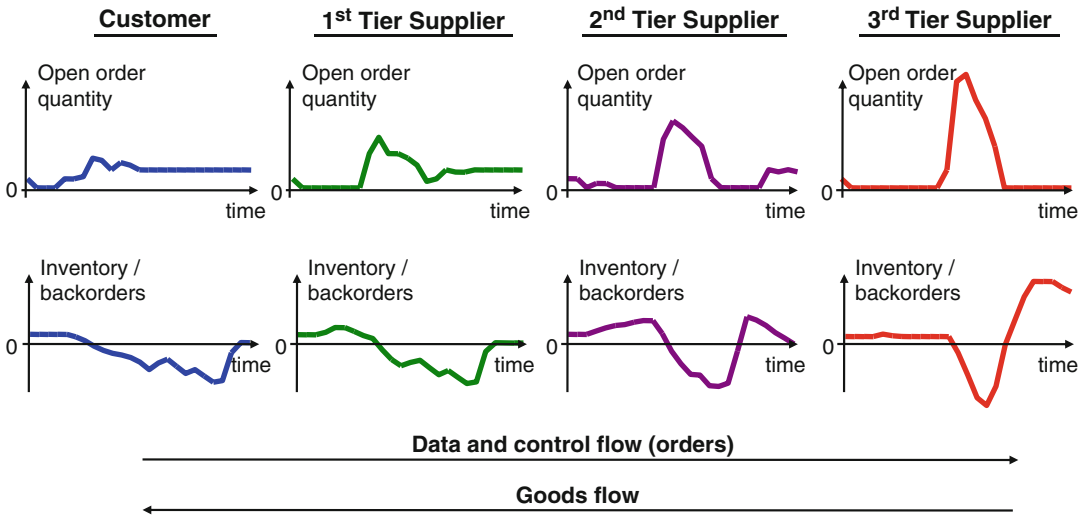
Collaborative Order Processing: Avoiding the Bullwhip Effect

To fulfill the objectives of the collaboration, not only must planning and control systems be linked but close contact among the participants is also necessary. According to APICS, collaborative planning, forecasting, and replenishment (CPFR) is "a process whereby supply chain trading partners can jointly plan key supply chain activities from production and delivery of final products to end customers. Collaboration encompasses business planning, sales forecasting, and all operations required to replenish raw materials and finished goods" (Blackstone 2013).

A key tool for CPFR and thus for effective SCM is blanket orders. A blanket order is a long-term agreement for a great number of deliveries, either material or capacities. Sometimes, a long-term blanket order may be precision-tuned for shorter-term time intervals. According to Blackstone (2013), a blanket release is "the authorization to ship and/or produce against a blanket agreement or contract."

In SCM, it is very important to implement countermeasures to prevent the bullwhip effect, also called the Forrester effect (Forrester 1958). The bullwhip effect is an extreme increase of the variation of (excess) inventory and back orders as well as an increase of open order quantities upstream in the supply chain from customer to the various tiers of suppliers, generated by a small change or no change in customer demand. See here also Lee et al. (1997) and Simchi-Levi et al. (2007). In addition, the longer the lead times of goods, data, and control flow are, the stronger the bullwhip effect is. Fig. 4 shows this effect.

A famous example, analyzed and published by Procter & Gamble, is the demand for Pampers disposal diapers. Besides the fluctuation of the demand, the bullwhip effect is also caused by information processing obstacles in the supply chain; the obstacles are information time lag and distortion (by the actual orders). Appropriate countermeasures are reducing the manufacturing lead times in



Supply Chain Management, Fig. 4 Open order quantities and inventories/back orders in a supply chain: the bullwhip effect (or Forrester effect) (Taken from Schönsleben (2016), Section 2.3.1)

general and quickly adapting manufacturing lead times on demand, based on rapid information exchange by point-of-sale scanning.

Cross-References

- ▶ ERP Enterprise Resource Planning
- ▶ Operations Management

References

Blackstone JH (ed) (2013) APICS dictionary, 14th edn. APICS (The Association for Operations Management), Chicago

Forrester J (1958) Industrial dynamics: a major breakthrough for decision makers. Harv Bus Rev 36(4):37–66

Lee H, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: the bullwhip effect. Manag Sci 43(4):546–558

Schönsleben P (2016) Integral logistics management – operations and supply chain management within and across companies, 5th edn. CRC Press/Taylor & Francis Group, New York

SCOR (2004) SCOR, version 6.1 overview, APICS Supply Chain Council, Chicago

SCOR (2014) SCOR, APICS Supply Chain Council, Chicago. www.supply-chain.org/scor. Accessed 18 Nov 2014

Simchi-Levi D, Kaminsky P, Simchi-Levi E (2007) Designing and managing the supply chain. Concepts, strategies, and case studies, 3rd edn. McGraw-Hill Irwin, New York

Supply Chain Network

▶ Production Networks

Supply Network

▶ Production Networks

Surface and Subsurface Characteristics

▶ Surface Integrity

Surface and Subsurface Properties

▶ Surface Integrity

Surface Degradation

▶ Corrosion

Surface Finishing

► Finishing

Surface Integrity

Bernd Breidenstein
Institut für Fertigungstechnik und Werkzeugmaschinen, An der Universität 2, Garbsen, Germany

Synonyms

Surface and subsurface characteristics; Surface and subsurface properties

Definition

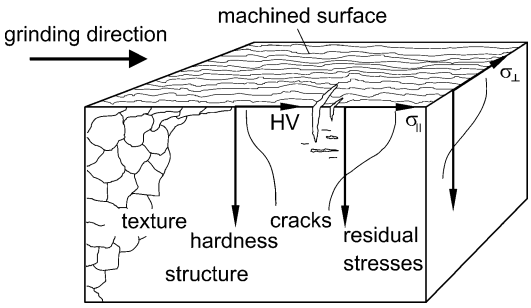
Surface integrity means the inherent or enhanced condition of a surface produced in a machining or other surface generating operations (according to Field and Kahles 1964).

Theory and Application

History

The term "surface integrity" in the context of machined surfaces was launched by Field and Kahles in the middle of the 1960s (Field and Kahles 1964). Subsequently more and more publications made use of this expression understanding plastic deformation, microcracking, phase transformations, hardness variations, tears and laps, residual stresses, etc. as typical measures to investigate.

In 1991 Brinksmeier indicated the following quantities to be measured for the description of surface integrity, including the subsurface for characterization (see Fig. 1): surface – profile, orientation of machining grooves, topography, roughness, etc. – and subsurface, material, microstructure, texture, residual stress, hardness, etc. (Brinksmeier 1991).



Surface Integrity, Fig. 1 Surface and subsurface properties (Brinksmeier 1991)

Terminology

"Surface" in common sense describes the external limitation of a component. In natural science, "surface" means the boundary between two media. For production technology, another specification is necessary. The DIN standard 4760 gives three definitions to be distinguished for "surface" (DIN 4760 1982):

1. Real surface: surface that separates the component from the surrounding medium
2. Actual surface: metrologically detected, approximated image of the real surface of a shaped element, may depend on the measurement method
3. Nominal surface: ideal surface, shape defined by engineering drawing or other technical document

As many effects are not strictly limited to the surface, it makes sense to include the subsurface into surface integrity considerations. "Subsurface" denotes that volume range of the material below the surface, where its properties have been modified by a machining process.

Characteristics

The surface of a component is the part being in contact with the surrounding medium, another component, or the eye of a viewer. The subsurface of a component must be able to compensate the demands acting upon it, in combination with a maximized lifetime. Herefrom the requirements on machined surfaces and subsurfaces result.

In the following, important surface and subsurface properties are discussed, including possibilities for their quantification.

Topography

Most of the properties of a machined surface depend on its topography. For the specification of surface topography, a multitude of different roughness parameters has been defined, the most common being the mean roughness R_a, the average peak-to-valley height or roughness depth R_z, and the maximum roughness depth R_{max}.

The mean roughness R_a is able to detect gradual changes in surface topography, e.g., as a result of tool wear. Peaks and grooves, however, cannot be distinguished, just as little as different profile shapes. For this, additional roughness values have to be determined.

R_a is regarded as a reliable parameter, as its determination is a result of intense averaging (Volk 2005). R_a denotes the mean deviation of the profile from the centerline (see Fig. 2). Its calculative determination is done by (DIN EN ISO 4287 2010).

$$R_a = \frac{1}{l}\int_0^l |Z(x)|dx$$

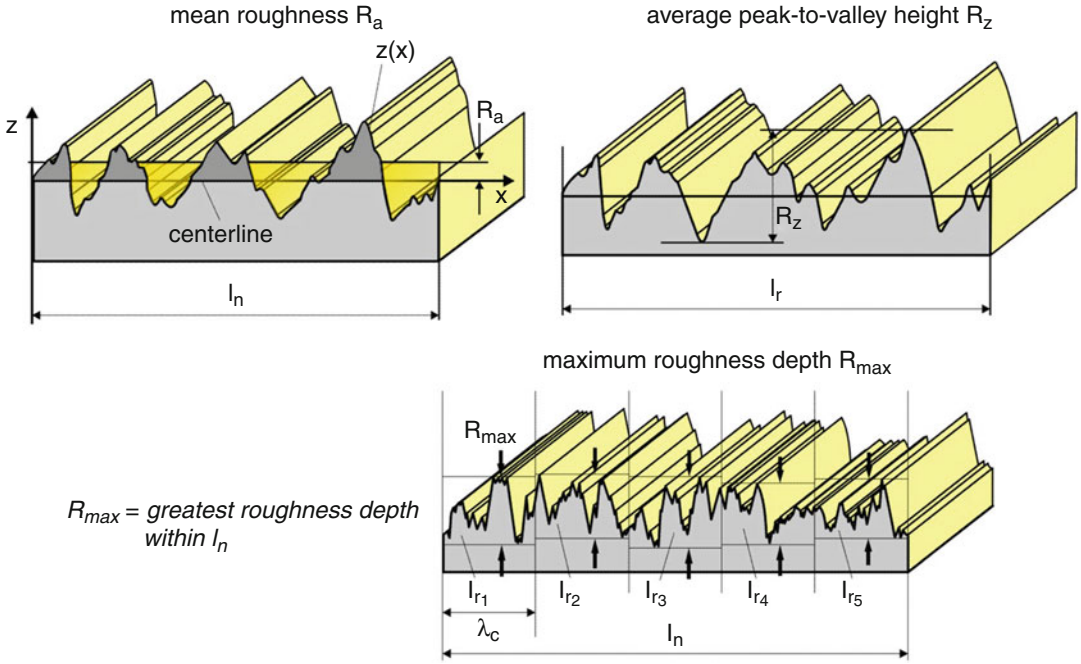
where l means the evaluated fraction of the measured length and z is the distance of the profile from the centerline within the measured length.

The application of average peak-to-valley height R_z and maximum roughness depth R_{max} as a combination allows the detection of individual outliers. If R_{max} is significantly higher than R_z, an extra high peak exists.

For the determination of R_z and R_{max}, the measured surface profile is filtered by the threshold wavelength λ_c, which separates roughness from waviness. The measuring distance l_n is divided into five equal sections l_r with the length of λ_c. From each section, the maximum peak value z_i is taken, and the mean average is calculated according to (see Fig. 2).

$$R_z = \frac{1}{5}\sum_{i=1}^{5} z_i$$

For each section, the distance from the highest peak to the deepest groove is determined. The



Surface Integrity, Fig. 2 Topographic parameters R_a, R_z, and R_{max} (Volk 2005)

highest distance within sections l_1–l_5 is defined as R_{max} (see Fig. 2) (Volk 2005).

The measurement procedures for roughness parameters work either in contact or contactless. Basically three measuring principles can be distinguished:

- Tactile methods
- Optical methods
- Scanning probe microscopy methods

Classical measurement procedures perform a section of the surface, which means that they work two-dimensionally. All standards refer to these methods. Scanning or the application of special sensors moreover allows the determination of planar roughness values. For the three-dimensional case, it is necessary to define the roughness parameters anew. This is currently being done within the standard DIN EN ISO 25178 (2013), which at present has the status of a technical draft.

For the measurement of machined surfaces, mostly two-dimensionally working tactile stylus instruments are applied. At this a stylus tip (as a rule made of diamond) is moved along the surface, and its displacement against a reference plane is measured. The obtained profile finally is treated mathematically. On mild surfaces, the stylus tip may leave marks, which is a disadvantage of tactile measuring methods.

This problem does not occur, if contactless procedures are applied. Optical systems have great potential for the application in surface measurement (Inasaki and Karpuschewski 2001). Optical point sensors with different measurement principles are applied.

Scanning probe microscopy methods use a probe to obtain information on the surface micro-geometry. The probe as a rule is a solid body, except for the scanning electron microscope, where the electron beam is regarded as a probe. The probe scans the surface with or without contact, depending on the kind of interaction. An image of the surface is created from the measurement data. These methods capture only a small surface section but deliver a high resolution. Important methods are, e.g., scanning tunneling microscopy, atomic force microscopy, and scanning electron microscopy.

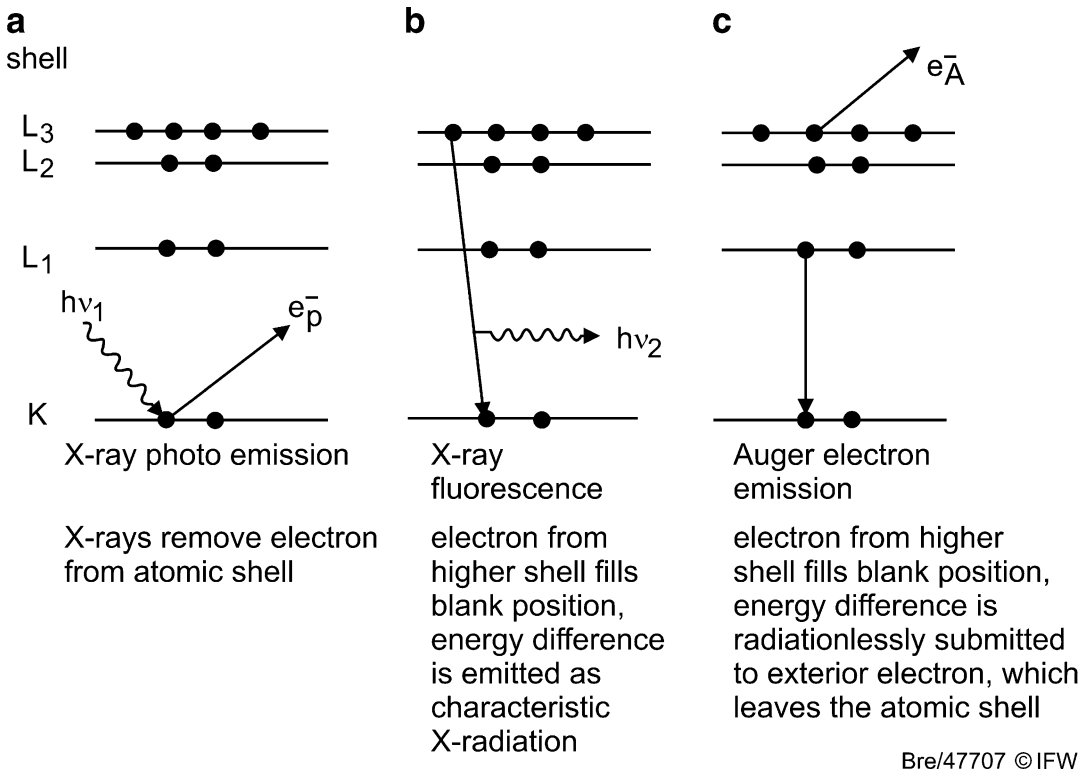
Chemical Composition

The chemical composition of a component surface may differ from that in the core. This can be a result of chemical reactions between tool and workpiece material or lubricant and workpiece material during machining as well as diffusion processes. Hence, it is important to have information on the chemical composition of the surface. The methods for analyses of the chemical composition, which are applied in production technology, are based upon interactions of energy-rich radiation (electromagnetic waves or electrons) with the atoms of the specimen surface (Fig. 3). Some important analysis procedures are briefly described in the following.

Possibilities for the Quantification of Surface and Subsurface Properties

Energy Dispersive X-Ray Analysis (EDX) If an electron near the nucleus gets an impulse by energy-rich radiation, it will be removed from its position and leave the atomic shell (see Fig. 3a). An electron from a higher shell will fill the blank position. The energy which is set free by this is emitted as X-radiation (see Fig. 3b). The energy of this X-radiation is specific to the atom (i.e., the chemical element) that is emitting it. This radiation, also called characteristic radiation, thus can be used for the identification of chemical elements. As a rule, different electron changes occur at the same time so that different characteristic energy levels can be detected for each chemical element. The detection of characteristic X-radiation is performed by the use of energy dispersive analyzers (detectors). Elements with an atomic number >5 can be evaluated qualitatively and quantitatively. Often scanning electron microscopes are equipped with energy dispersive X-ray detectors.

Auger Electron Spectroscopy (AES) During the generation of characteristic X-radiation, there exists a competitive effect, where the succeeding electron submits its energy radiationlessly to an exterior electron, which leaves the atomic shell (see Fig. 3c). This is the so-called Auger effect; the released electron is called Auger electron. The Auger effect is more distinct for light atoms, in opposition to characteristic X-radiation. The



Surface Integrity, Fig. 3 Physical basics for the determination of the chemical composition (Briggs and Grant 2003)

specific attribute of Auger electrons is their energy. Measuring this energy by an electron spectrometer allows identification of the emitting chemical element.

Microprobe Analysis This method belongs to the most exact physical methods for the determination of local chemical compositions (Ammann 1994). Such as the energy dispersive X-ray analysis, also the microprobe analysis is based upon measurement of the intensity of characteristic X-radiation, excited by an electron beam. In contrast to EDX, the X-ray quanta here are detected no longer as energy dispersive but wavelength dispersive. Curved focusing X-ray monochromators are used for high angular resolution. In opposition to scanning electron microscopes, the electron beam is fixed in microprobes. This wavelength dispersive method allows a detection limit which is two to three magnitudes lower than for energy dispersive methods. All elements with an atomic number >3 can be analyzed.

X-Ray Photoelectron Spectroscopy (XPS) This method is based upon the photoelectric effect (see Fig. 3a), which is activated by low-energy X-radiation (usually Al Kα or Mg Kα radiation). The specimen chamber is evacuated. The released electrons are emitted into the vacuum (photo emission), where their energy is determined by the use of an electron spectrometer, comparable to AES. Quantitative determination of the chemical composition requires substantial corrections of the measured spectrum (Briggs and Grant 2003). By variations of the incident angle, this method can deliver very surface-near information (monolayer).

Glow Discharge Optical (Emission) Spectroscopy (GDO(E)S) This method is mentioned here, though it is not focused on surface analysis but on quantitative analysis of the chemical composition of materials in general. By development of GDOS methods in the 1960s, the significantly more time- and cost-consuming wet chemical

analyses could be replaced, e.g., in the steel and aluminum industries (Wilken 2004).

Material from a specimen is removed and excited by cathodic evaporation via glow discharge. The light, emitted by the atoms, is detected wavelength dispersively by the use of an optical spectrometer. The determined intensities are proportional to the mass of the particular chemical element and its concentration in the specimen. All elements from the periodic table can be analyzed. As this method removes material from the surface, GDOS can be used for the detection of gradients of the chemical composition in the subsurface.

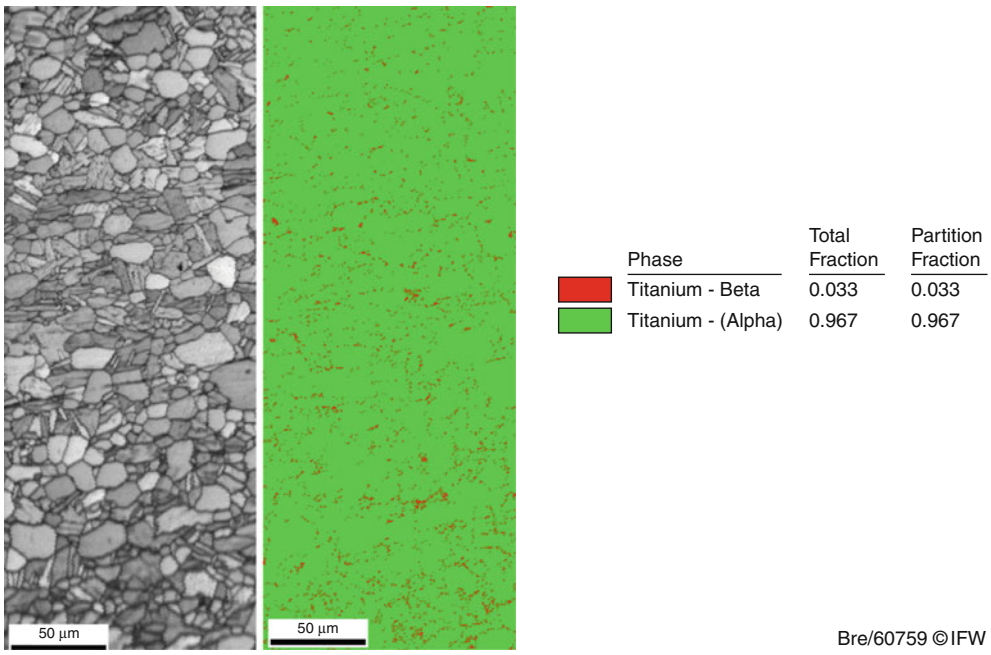
Phase Composition

All analysis methods described before cannot distinguish different phases of the same material, e.g., α-Ti and β-Ti. For phase determination, scattering methods have to be applied. The most common is X-ray diffractometry (XRD), which can perform qualitative and quantitative phase analysis of surface-near regions, as the penetration (information) depth of the X-rays used by this method is limited to a few microns (depending on X-ray energy and specimen material). Monographs on XRD methods describe the application in detail (Jenkins and Snyder 1996; Cullity and Stock 2001; Spieß et al. 2009).

Also electron diffraction methods like electron backscatter diffraction (EBSD) can deliver phase information (see Fig. 4). EBSD, applied in scanning electron microscopes, assigns to each surface grain its phase affiliation and its orientation as a result of detected Kikuchi diagrams during specimen scanning (Schwartz et al. 2009). Figure 4 shows an EBSD analysis result for a two-phase titanium specimen.

Hardness

Hardness as resistance of a material against penetration of another body marks a decisive material characteristic. The hardness of important construction materials can be influenced or selectively set by special hardening processes. Hardening is based upon different principles, e.g., the formation of martensite in steels as a result of thermal treatment. For nonferrous metals, precipitation hardening plays an important role. Alloying elements are deposited by a multilevel



Surface Integrity, Fig. 4 Quantitative phase determination by EBSD for a two-phase titanium material

thermal process. Their phase boundaries and size influence the increase of hardness and stability decisively. This attribute improvement is based upon the hindrance of the motion of dislocation (Bargel and Schulze 1988).

For hardness determination, different methods are possible: scratching the surface, penetration of an indenter with static or dynamic loads, or rebound as a result of elastic material behavior. The methods with a penetrating indenter are the most important ones. The applied methods are distinguished, e.g., by the shape of the indenter. Brinell hardness is determined by a ball-shaped indenter, while Vickers hardness applies a pyramid-shaped one. After the indenting test with a certain load, the surface area of the indentation is measured which delivers a value for material hardness. Determination of Rockwell hardness uses the depth of the indentation instead of the surface area (Bargel and Schulze 1988). Independent of the method, the so-called surface hardness can be determined normal to the corresponding surface, while hardness-depth gradients have to be determined on cross-section polishes.

Microstructure

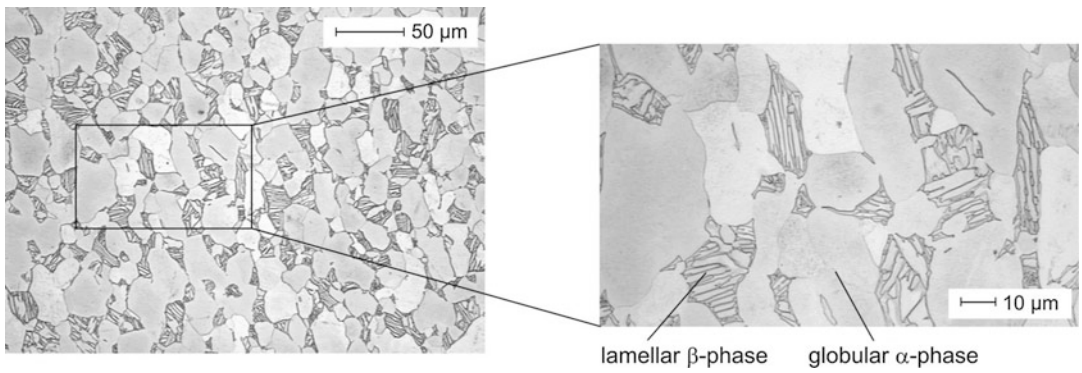
Most of the important construction materials are polycrystalline. The assembly of grains of the material forms its microstructure, which is characterized by grain size and shape. Microstructures can be homogeneous or heterogeneous (Bargel and Schulze 1988). Often the microstructure of the subsurface is modified compared to that of the basic material as a result of machining. Classical methods for uncovering microstructures are based upon metallographic methods. A specimen of the material in question is ground and polished and in most cases etched, before it is investigated microscopically. Figure 5 shows a polished micrograph section of a two-phase titanium material (Gey 2002).

Residual Stress

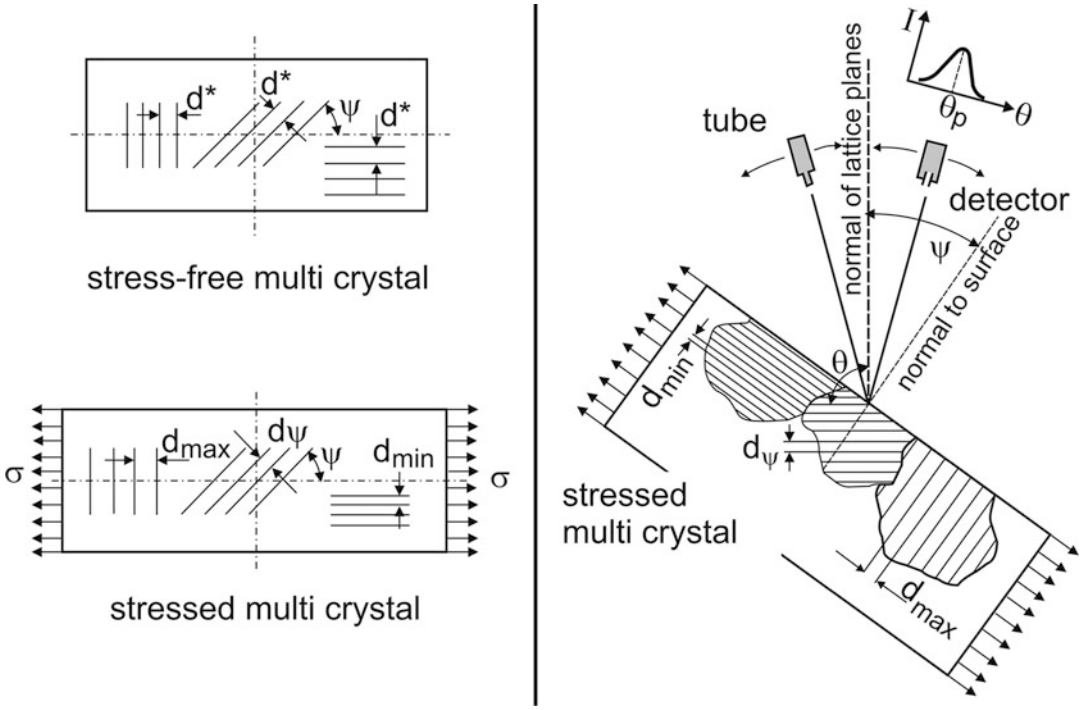
Residual stress can be determined with indirect or with direct methods. In both cases, strain is measured, and from this residual stress is calculated. The indirect method is based upon measuring the distortion of a stressed part after separation of a small part or layer. Measurement is done by strain gauges or via optical interference. Residual stress computation needs the application of algorithms from elasto-mechanics. Because of the "deviation" by springback of the remaining body, this procedure is called indirect method (Toenshoff and Denkena 2004).

Between the direct methods, the X-ray diffractometric $\sin^2\psi$ method is the most important one (Macherauch and Müller 1961). With this method, according to Bragg's diffraction condition, net plane distances d_{hkl} are measured, i.e., existing strains are measured directly. As strain of a specimen is direction dependent, it is necessary to incline the specimen around the angle ψ (Fig. 6).

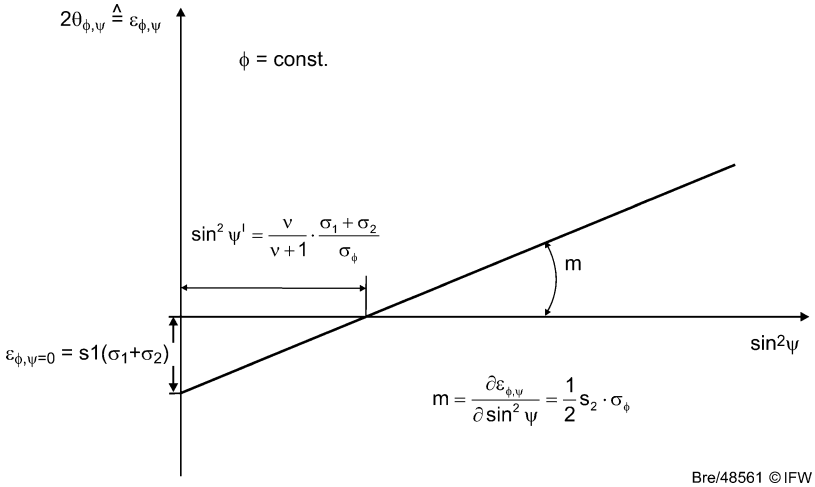
The net plane distances d_{hkl} in ideal case show a linear distribution versus $\sin^2\psi$, from which the residual stress state in the measured



Surface Integrity, Fig. 5 Polished micrograph section of a two-phase titanium material (Gey 2002)



Surface Integrity, Fig. 6 Direction-dependent lattice strain by residual stress (Brinksmeier 1982)



Surface Integrity, Fig. 7 Correlations of strain and stress at the $\sin^2\psi$ method (Macherauch and Müller 1961)

direction can be determined (Fig. 7). The residual stresses determined by this method are as a rule two-dimensional.

For the determination of residual stress-depth gradients, material has to be removed gradually, e.g., by electrolytic polishing. This arises from the low penetration (information) depth of the X-rays into the material, which lies in the range of few microns only. Based upon the $\sin^2\psi$ method, several new methods have been developed, like the scattering vector method (Genzel 1999), which allow nondestructive determination of residual stress-depth gradients in depths lower than penetration depth. These methods are applied, e.g., for

the determination of gradients in PVD and CVD coatings. Due to the development of sensitive detectors, also energy dispersive X-ray scattering experiments for the nondestructive determination of residual stress-depth distributions in the X-ray lab gain importance (Genzel et al. 2010).

Cutting and grinding processes generate residual stresses in the machined component. In the surface itself, only two axial stresses can occur. There are mainly two sources that create residual stresses during machining: mechanical loads and thermal effects. Mechanical load leads to compressive residual stress as a result of plastic strain in the surface-near regions. In front of the cutting edge, elastic and plastic compressions occur. Behind the tool contact, plastic and elastic strain arises, which partly springs back. After load relieving, compressive residual stresses remain (Toenshoff and Denkena 2004).

During machining, the material subsurface is heated temporarily to high temperatures. By this the surface-near layers expand as a result of thermal compressive stress. The surface-near layers deform plastically in a state of reduced yield strength. After cooling down to room temperature, tensile residual stresses result. As in machining both mechanical and thermal influences act concurrently, they interfere with each other strongly nonlinear. A simple superposition is not valid, but a formal overlap can give indications for thermal and mechanical relations in the chip formation zone (Toenshoff and Denkena 2004).

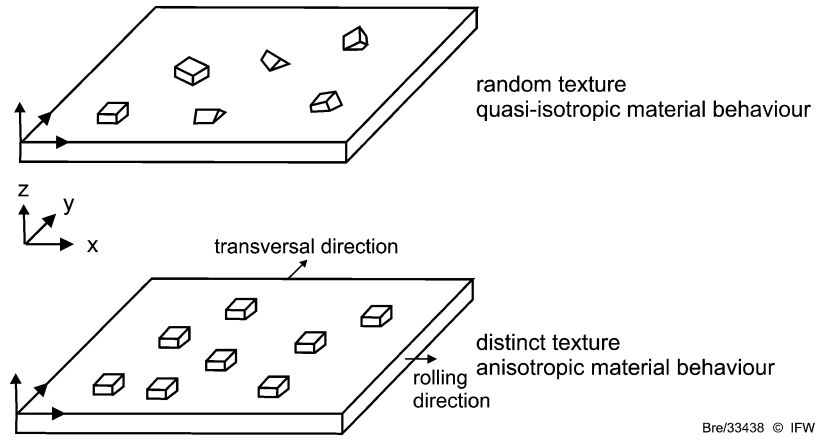
Texture

Crystallographic texture describes the orientation of crystallites to each other or to a specimen coordinate system. The two extremes of texture are random orientation (Fig. 8, top), where there is no preferred orientation and the material behaves quasi-isotropic, and distinct texture (Fig. 8, bottom), where the material behaves anisotropic.

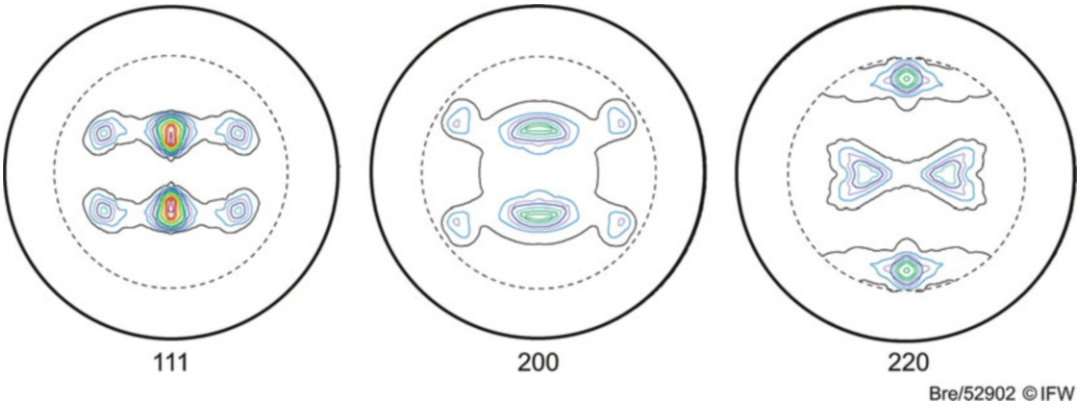
Textures mostly are a result of forming processes like rolling and extruding, but they may also occur after cutting processes, which lead to modified physical material properties like hardness and corrosion resistance. Textures can be identified by scattering experiments, applying electrons, neutrons, or X-rays. Most common are X-ray diffractometric methods, though the application of electron backscatter diffraction (EBSD) gains importance. In X-ray diffractometry, the specimen has to be rotated around its normal and tilted gradually in a range of 0–70°. The diffracted intensity of low-indexed peaks is measured in dependency of rotation (azimuth) and inclination (polar distance) angle and displayed in the so-called pole figures (Fig. 9). From the hereby obtained information, orientation distribution functions (ODF) for quantitative texture analysis and theoretical pole figures can be computed (Bunge 1969).

Systematic Setting of Surface and Subsurface Characteristics by Machining

The above knowledge can be used for a targeted design of surface and subsurface during

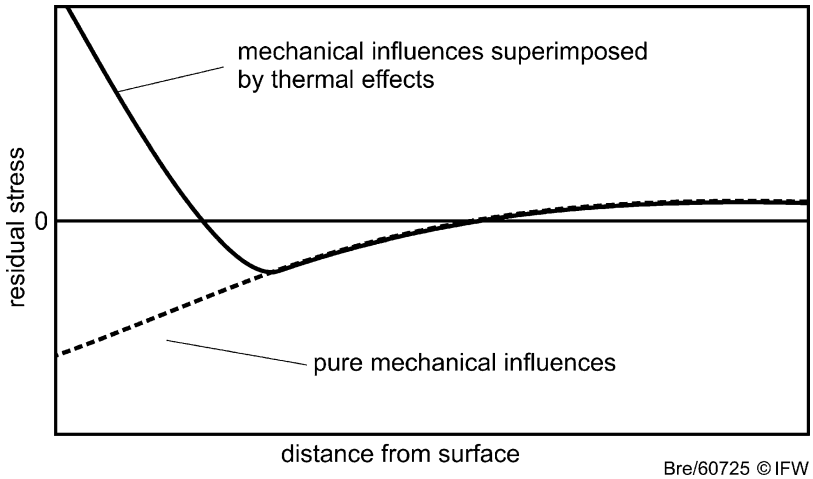


Surface Integrity, Fig. 8 Random texture (top) and cube texture (bottom) (Bunge 1969)



Surface Integrity, Fig. 9 Pole figures of aluminum foil, measured by X-rays

Surface Integrity, Fig. 10 Model of residual stress development (according to Jacobus et al. 2000)

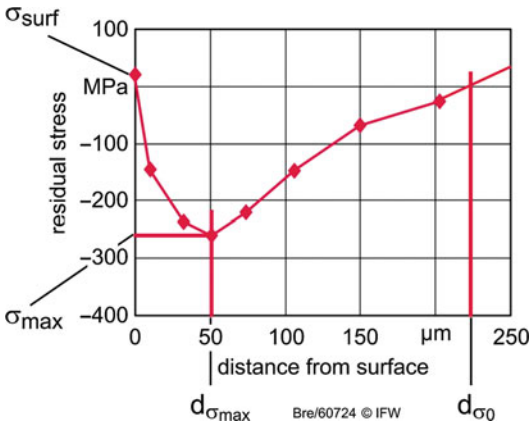


machining. As most surface properties are a result of micro-geometry, roughness plays the most important role for surface quality. In machining processes, roughness can be controlled by a proper choice of machining parameters. In turning processes, e.g., feed is the critical parameter. In a wide range, roughness is increasing nearly linear with increasing feed. At very low feeds, secondary effects, e.g., vibrations, penetration of the tool into the material, material properties, or minimum chip thickness, gain importance and lead to higher roughness values than expected. Other machining parameters like cutting speed or depth of cut have lower influence on surface roughness in turning.

In grinding processes, increasing single-grain chip thickness leads to increasing roughness as well as a decreasing number of active cutting edges. Thus, by decreasing grain size, decreasing pressure of tool engagement, or increasing cutting speed, surface quality in grinding can be controlled in a certain range.

Residual stress in the surface reflects the residual stress state in the subsurface only very insufficiently. For the understanding of the development and the actual state, it is thus required to determine residual stress-depth gradients. According to a model assumption, pure mechanic influences would cause compressive residual stress with its maximum in the surface, while the superimposing thermal effects lead to the typical residual stress-depth gradients shown in Fig. 10 (Jacobus et al. 2000).

Another model, which has been derived from end milling of aluminum, explains the typical



Surface Integrity, Fig. 11 Curve parametrization and characteristic values of residual stress-depth gradients (according to de León 2009)

gradient as a result of pure mechanical influence; as in the investigated process, the temperature development could be neglected (de León 2009). This model allows a parametrization of the curve and a definition of typical values, e.g., surface residual stress σ_{surf}, maximum residual stress σ_{max}, depth of maximum residual stress d_{max}, or depth of stress absence $d_{\sigma 0}$ (Fig. 11) (de León 2009).

With the knowledge that mechanical loads during machining lead to a shift of the stress curve into direction of compressive stress and thermal influences lead to a shift into direction of tensile stress, it is possible to generate a tailored residual stress-depth distribution by variation of machining parameters.

In practice, however, this seems to be very difficult as a round-robin collaborative work of 12 international researchers as a round robin test shows. Target was to machine a steel sample of AISI 52100 (100Cr6) in order to get −200 MPa compressive stress at the surface. As processes milling, grinding, turning, fine grinding, and EDM were applied. Only turning and fine grinding reached about the given value; the others scattered in a range of −800 MPa to +600 MPa. The results will be forthcoming in the CIRP Annals Manufacturing Technology (Jawahir et al. 2011).

Correlations Between Component Characteristics and Component Performance

Component performance is influenced decisively by surface and subsurface characteristics. They may either extend or shorten the lifetime of a component considerably. Many of the characteristics are defined by the final machining process so that it is of particular importance for component quality.

For a functional surface, e.g., a sealing face, the roughness needs to go below a certain limit, in order to guarantee a correct function of the component. Also corrosion resistance may be a function of roughness. An ideal smooth surface possesses a small area, which is enlarged with increasing roughness, giving a greater corroding surface.

In case of tensile, bending, torsion, or alternating loads, roughness may be of particular importance for the component's lifetime, namely, if roughness values exceed a certain limit, so that stress concentration as a result of the notching effect occurs. Near a notch stress increases disproportionately intense and may reach unexpectedly high values (Thum et al. 1960). Size and type of the notch influence the dimension of the appearing stress, which may shorten the component's lifetime considerably.

The microstructure of a material is responsible for specific material characteristics. It can be set selectively, e.g., by heat treatment. Machining processes, which work near a technologically justifiable load limit, bear the danger of accidental microstructure modification. One prominent example for this is grinding burn. The term grinding burn is used, if in grinding processes a too high thermal influence of the microstructure has taken place, with the result of a microstructure modification, which shortens component lifetime, in some cases drastically.

For instance, during grinding of a case hardened steel, the heat input into the workpiece may lead to an undesired modification of the microstructure from martensitic to pearlitic. Additionally the existing and essential compressive residual stress is reduced or in disadvantageous cases even shifted to tensile residual stress, thus making the component not suitable for its

intended use. Grinding burn is detected by complex etching procedures (nital etching) or by Barkhausen noise analysis (BNA). Currently methods are developed that take regulating action of the process in order to detect the danger of grinding burn and avoid its development (Liu et al. 2006; Saxler 1997).

Another prominent example for accidental microstructure modification is white layers, more or less thick surface layers in hardened materials. They show a clearly increased hardness and a resistance against etching so that they appear white in metallographic polished sections. The phenomenon appears in certain machining processes like grinding or hard machining but also on railroad tracks, as a result of mechanic and thermal influences. Below the white layer, there is also more or less thick, tempered zone with reduced hardness. Microstructural analyses of white layers lead to different results (Guo and Sahni 2004).

Concerning residual stresses, it is basically valid that, in case of compressive residual stress, a gain of fatigue strength, depending on the compressive residual stress amount, can be expected. This effect, however, is depending on material hardness. Hardened materials as well as heat-treated steels show the effect clearly, while recrystallized material shows no change of fatigue strength at different residual stresses (Sollich 1994). Tensile residual stress forwards crack generation and propagation, which affects fatigue strength. An exception can be found at hard-turned components, which in spite of tensile residual stress in the surface show a comparable fatigue strength to ground specimens with compressive residual stress. The reason for this is most likely the fact that the tensile residual stress is limited to very surface-near regions only, while in deeper regions, compressive residual stress hinders crack propagation (Borbe 2001; Renner et al. 2000).

Cross-References

- ▶ Cutting, Fundamentals
- ▶ Machinability of Aluminum and Magnesium Alloys

References

Ammann N (1994) Quantitative Tiefenprofilanalyse mit der Elektronenstrahlmikrosonde – Entwicklung der Technik und Untersuchungen zur Diffusion von Gallium in ZnSe/GaAs (Quantitative depth profile analysis by applying electron beam micro sensors – development of the technique and investigations on the diffusion of gallium in ZnSe/GaAs). Thesis RWTH Aachen, VDI-Verlag, Düsseldorf (in German)

Bargel H-J, Schulze G (eds) (1988) Werkstoffkunde (Materials science), 5th edn. Düsseldorf, VDI-Verlag. (in German)

Borbe C (2001) Bauteilverhalten hartgedrehter Funktionsflächen (Component behaviour of hard turned functional faces). Thesis Universität Hannover, Fortschritt-Berichte VDI, Reihe 2, Nr. 583, Berichte aus dem Institut für Fertigungstechnik und Spanende Werkzeugmaschinen, Universität Hannover, VDI-Verlag, Düsseldorf (in German)

Briggs D, Grant JT (eds) (2003) Surface analysis by Auger and X-ray photoelectron spectroscopy. IM Publications, Charlton

Brinksmeier E (1982) Randzonenanalyse geschliffener Werkstücke (Subsurface analysis of ground workpieces). Thesis, Universität Hannover, Hannover (in German)

Brinksmeier E (1991) Prozess- und Werkstückqualität in der Feinbearbeitung (Process- and workpiece quality in fine machining). Postdoctoral lecture qualification, Universität Hannover, Hannover (in German)

Bunge HJ (1969) Mathematische Methoden der Texturanalyse (Mathematical methods in texture analysis). Akademie Verlag, Berlin. (in German)

Cullity BD, Stock SR (2001) Elements of X-ray diffraction, 3rd edn. Prentice Hall, Upper Saddle River

de León LR (2009) Residual stress and part distortion in milled aerospace aluminium. Thesis, Universität Hannover, Bd. 1 von Berichte aus dem IFW, Berichte aus dem IFW, PZH Produktions-technisches Zentrum, Hannover

DIN 4760: 1982-06 (1982) Gestaltabweichungen; Begriffe, Ordnungssystem (Form deviations; concepts; classification system). Beuth, Berlin

DIN EN ISO 25178 (2013) Geometrische Produktspezifikation (GPS) – Oberflächenbeschaffenheit: Flächenhaft (Geometrical product specifications (GPS) – surface texture: Areal). Beuth, Berlin. (in German)

DIN EN ISO 4287 (2010) Geometrische Produktspezifikation (GPS) – Oberflächenbeschaffenheit: Tastschnittverfahren – Benennungen, Definitionen und Kenngrößen der Oberflächenbeschaffenheit (Geometrical product specifications (GPS) – surface texture: profile method – terms, definitions and surface texture parameters). Beuth, Berlin (in German)

Field M, Kahles JF (1964) The surface integrity of machined and ground high strength steels. DMIC Rep 210:54–77

Genzel C (1999) A self-consistent method for X-ray diffraction analysis of multiaxial residual-stress fields in

the near-surface region of polycrystalline materials. I. Theoretical concept. J Appl Crystallogr 32:770–778
Genzel C, Krahmer S, Klaus M, Denks IA (2010) Energy-dispersive diffraction stress analysis under laboratory and synchrotron conditions: a comparative study. J Appl Crystallogr 44:1–12
Gey C (2002) Prozessauslegung für das Flankenfräsen von Titan (Process design for flank milling of titanium). Dissertation, Universität Hannover, Verein Deutscher Ingenieure: (Fortschritt Berichte VDI/2) Fortschrittberichte VDI: Reihe 2, Fertigungstechnik; Nr. 625, Berichte aus dem Institut für Fertigungstechnik und Werkzeugmaschinen, Universität Hannover, VDI-Verlag, Düsseldorf
Guo YB, Sahni J (2004) A comparative study of hard turned and cylindrically ground white layers. Int J Mach Tools Manuf 44(2–3):135–145
Inasaki I, Karpuschewski B (2001) Abrasive processes. In: Toenshoff HK, Inasaki I (eds) Sensors applications. vol 1: sensors in manufacturing. Wiley, Weinheim, pp 236–271
Jacobus K, DeVor RE, Kapoor SG (2000) Machining-induced residual stress: experimentation and modeling. Trans ASME 122:20–31
Jawahir IS, Brinksmeier E, M'Saoubi R, Aspinwall DK, Outeiro JC, Meyer D, Umbrello D, Jayal AD (2011) Surface integrity in material removal processes: recent advances. CIRP Ann Manuf Technol 60(2): 603–626
Jenkins R, Snyder RL (1996) Introduction to X-ray powder diffractometry. Wiley, New York
Liu Q, Chen X, Gindy N (2006) Investigation of acoustic emission signals under a simulative environment of grinding burn. Int J Mach Tools Manuf 46(3–4):284–292. https://doi.org/10.1016/j.ijmachtools.2005.05.017
Macherauch E, Müller P (1961) Das $\sin^2\psi$-Verfahren der röntgenographischen Spannungsmessung (The $\sin^2\psi$-method in X-ray residual stress measurement). Z Angew Phys 13(7):305–312. (in German)
Renner F, Zenner H, Borbe C, Toenshoff HK (2000) Forschungsbericht P 337: Vergleich der Schwingfestigkeit hartgedrehter und geschliffener Bauteile (Comparison of the dynamic strength of hard turned and ground workpieces). Studiengesellschaft Stahlanwendung e. V., Stiftung Industrieforschung, Verlag und Vertriebsgesellschaft mbH, Düsseldorf (in German)
Saxler W (1997) Erkennung von Schleifbrand durch Schallemissionsanalyse (Detection of grinding burn by acoustic emission analysis). Thesis RWTH Aachen, VDI-Verlag, Düsseldorf (in German)
Schwartz AJ, Kumar M, Adams BL, Field DP (eds) (2009) Electron backscatter diffraction in materials science, 2nd edn. New York, Springer
Sollich A (1994) Verbesserung des Dauerschwingverhaltens hochfester Stähle durch gezielte Eigenspannungserzeugung (Improvement of the fatigue behaviour of high strength steels by targeted generation of residual stress). Thesis Universität Kassel, VDI-Verlag, Düsseldorf (in German)
Spieß L, Teichert G, Schwarzer R, Behnken H, Genzel C (2009) Moderne Röntgenbeugung: Röntgendiffaktometrie für Materialwissenschaftler, Physiker und Chemiker (Modern X-ray diffraction: X-ray diffractometry for material scientists, physicists and chemists), 2nd edn. Vieweg-Teubner, Wiesbaden (in German)
Thum A, Petersen C, Svenson O (1960) Verformung, Spannung und Kerbwirkung (Strain, stress and notch effect). VDI-Verlag, Düsseldorf
Toenshoff HK, Denkena B (2004) Spanen (Cutting), 2nd edn. Springer Verlag, Berlin. (in German)
Volk R (2005) Rauheitsmessung – Theorie und Praxis (Roughness measurement – theory and practice). Beuth Verlag, Berlin. (in German)
Wilken L (2004) Theoretische und experimentelle Untersuchungen zur Optimierung einer Glimmentladungsquelle für spektroskopische Messungen von elektrisch leitenden und nichtleitenden Materialien (Theoretical and experimental investigations for the optimization of a glow discharge source for spectroscopic measurements of electrically conductive and not-conductive materials). Thesis TU Dresden, Shaker Verlag, Aachen (in German)

Surface Mounted Device (SMD) Placement

► SMD Component Placement

Surface Parameter

Han Haitjema
Mitutoyo RCE, Best, The Netherlands
KU Leuven, Department of Mechanical Engineering, Leuven, Belgium

Synonyms

Roughness parameter

Definition

Parameter, defined as a number and a unit that characterizes an averaged dimensional property of a surface.

Theory and Application

Surfaces and Parameters

A surface can be characterized by many, many parameters. In fact, it is easier to define a new parameter than to come with a thorough analysis of the usefulness of the already existing parameters. Parameters are defined in many ISO standards (see references). They can be separated in 2-D and 3-D parameters, and further separation is possible in amplitude parameters, spacing parameters, hybrid parameters, parameters derived from integrated probability density curves, and topological parameters.

This section is confined to surface texture characterizing parameters. For the defined *R* (roughness) parameters, there are equivalent *P* (unfiltered profile) parameters and *W* (waviness) parameters that are defined for longer wavelength limits.

Roughness parameters naturally depend on the length scale where they are defined. In dimensional metrology, this scale is confined between the short-wavelength cutoff lengths λs and the long-wavelength cutoff λc. In the ISO 25178 series these terms are replaced by S-nesting index and L-nesting index respectively. It is a common requirement that within the short-wavelength cutoff length, at least some five measurement points should be taken, to keep the filtering well defined for short wavelengths. Parameters and specifications can only be usefully compared when the same cutoff lengths are used and meant (Leach and Haitjema 2010).

Indication of Surface Parameters on Technical Drawings

In ISO 1302 (2002), it is described how surface parameter requirements should be indicated on technical drawings.

Uncertainty in Surface Parameters

When measuring surface parameters and comparing these to specifications, it is necessary that an uncertainty in the parameter is calculated. Some of the factors that must be taken into account are:

Inhomogeneity of the workpiece

The parameter may depend on the location on the workpiece and the direction in which the measurement is taken (at least for 2-D measurements). Usually, this is the dominant uncertainty factor.

Uncertainty and traceability of the surface measuring instrument

Uncertainty and traceability of the surface measuring instrument

The traceability and uncertainty of the coordinates as measured (see section "Definitions") must be taken into account, as well of uncertainties in the used probe size, filtering, sampling, reference line or surface, etc. For 3-D measurements, it is a problem that the default measurement system is a mechanical probing system. As 3-D mechanical roughness measurements are very time-consuming, most measurements are taken using optical techniques that may have artifacts that are not easily recognized.

A problem with these calculations is that they can become rather tedious and surface dependent (Haitjema 2015; Haitjema and Morel 2000).

2-D Roughness Parameters

In this section, roughness parameters are defined, derived from the measured profile, with a base length *lr* and the cutoff wavelength λc, filtered with the standard Gaussian filter. This base length is also called sampling length.

Unless otherwise stated, the parameter is calculated for every base length *lr* and averaged over the evaluation length *ln* that usually consists of five consecutive base lengths.

The defined parameters *R* are valid analogous as *W* and *P* parameters, respectively, for the waviness and unfiltered profile.

Amplitude Parameters (Top-Bottom)

Rz The maximum profile peak height *Rp* is the largest profile peak height *Zp* within the base length (ISO 4287 1997). Analogously, the maximal profile valley depth *Rv* is the largest profile valley depth *Zv* within the sampling length. The

sum of *Rp* and *Rv* is the maximal profile height *Rz*. As commonly the measured length *ln* consists of several (typically 5) sampling lengths *lr*, *Rz* is the average of the *Rz* within every sampling length, at least according to the present ISO standard. Confusingly this definition was different in the past and still may be different in some national standards.

Rt The total profile height *Rt* is the sum of the largest profile peak height *Zp* and the largest profile valley depth *Zv* within the evaluation length *l n*. This evaluation length usually consists of several (typically 5) of the sampling length as it is illustrated in Fig. 1, where the measurement of one sampling length is given.

Amplitude Parameters (Average of Ordinates)

Ra *Ra*

is the arithmetic average deviation of the absolute ordinate values $Z(x)$ within the base length, defined in Eq. 1:

$$Ra = \frac{1}{l}\int_0^l |Z(x)|\mathrm{d}x \tag{1}$$

Due to the relatively easy calculation, the insensitivity for outliers, and for historical reasons, *Ra* is by far the most used parameter.

Rq *Rq*

is the root mean square and standard deviation of the ordinate values $Z(x)$ within the base length, defined in Eq. 2:

$$Rq = \frac{1}{l}\int_0^l Z^2(x)\mathrm{d}x \tag{2}$$

Spacing Parameter

RSm *RSm*

is the average width of profile elements *Xs* within the base length. A profile element is the area of a peak plus the adjacent valley (Fig. 2):

RSm is defined in Eq. 3, with the terms as defined in Fig. 2:

$$RSm = \frac{1}{m}\sum_{i=1}^{m} Xs_i \tag{3}$$

Hybrid Parameters

RΔq* or *Rdq *RΔq*

is the square average of the local slope *dZ*/*dx* within the base length as defined in Eq. 4:

$$R\Delta q = \sqrt{\frac{1}{l}\int_0^l \left(\frac{\mathrm{d}Z(x)}{\mathrm{d}x}\right)^2 \mathrm{d}x} \tag{4}$$

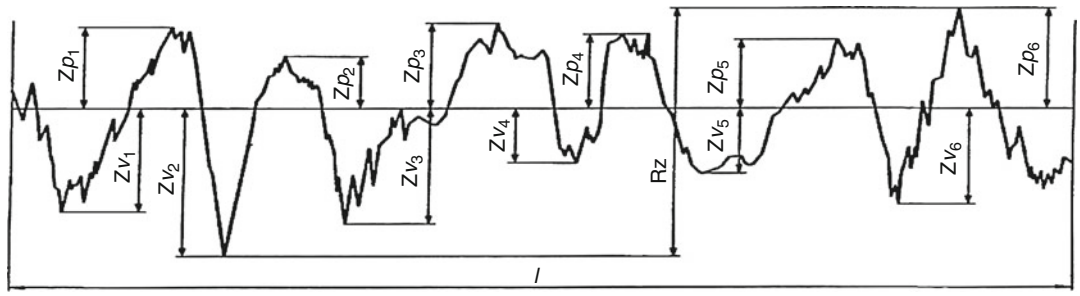
Rλq

The square average mean wavelength *Rλq* is a measure for the distance between peaks and valleys and can be calculated from Eq. 5:

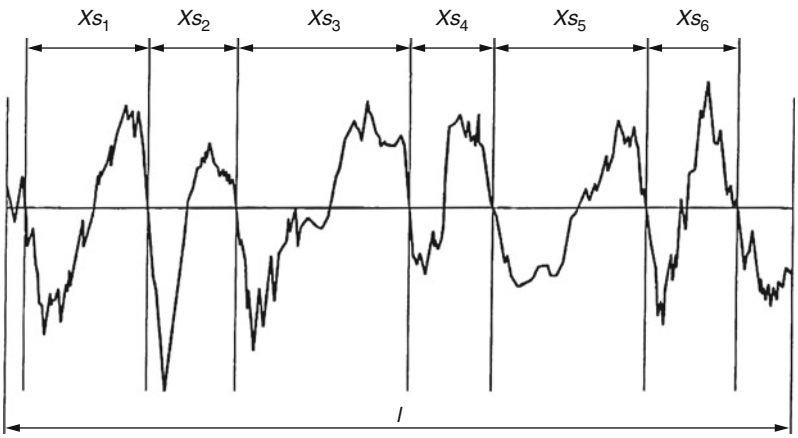
$$R\lambda q = \frac{2\pi Rq}{R\Delta q} \tag{5}$$

Topological/Field Parameters

A completely different approach of surface parameters is based on topography of peaks and valleys. Definitions are given in ISO 12085 (1996).



Surface Parameter, Fig. 1 Definition of *Rz*. *l* is the sampling length



Surface Parameter, Fig. 2 Definition of *RSm*. *l* is the sampling length. *RSm* is the average length of the profile elements *Xsn*

Parameters Derived from Probability Density Curves

The integrated probability density curve of the ordinates is also known as the Abbott-Firestone curve. From this curve several parameters can be derived as it is described in ISO 13565-3 (1998).

Abbott-Firestone curves are used in important sliding applications, particularly cylinder liners.

3-D Roughness Parameters

3-D roughness parameters are generally defined similarly as the 2-D parameters, with some additional parameter that depend on the direction such as the *Std* parameter The definitions are given in ISO 25178-2 (2012).

Value and Usefulness of Surface Parameters

One use of roughness parameters is for quality control and commerce so that a buyer and seller can agree on the quality of a product. In quality control it is necessary to be able to discriminate good and bad surfaces.

Another use is in design, so the designer can specify the texture as roughness influences function.

Yet another use of the roughness parameters is in research. In order to optimize the design of products and processes, design and manufacturing engineers need to know about the relation between roughness and performance on one hand and with processing on the other. Functional correlations of both kinds are valuable and depend on selecting the correct characterization parameter.

Surface parameters are also used in research in forensics, physical anthropology, archeology, and cultural heritage conservation.

Cross-References

- ▶ Form Error
- ▶ Functional Correlation
- ▶ Roughness
- ▶ Surface Integrity
- ▶ Surface Texture

References

Haitjema H (2015) Uncertainty in measurement of surface topography. Surf Topogr Metrol Prop 3(3):035004

Haitjema H, Morel M (2000) Traceable roughness measurements of products. In: De Chiffre L, Carneiro K (eds) Proceedings of the 1st Euspen topical conference on fabrication and metrology in nanotechnology, vol 2. Department of Manufacturing Engineering, TU Denmark, Copenhagen, pp 354–357

ISO 12085 (1996) Geometrical product specifications (GPS) – surface texture: profile method – motif parameters, ISO. Beuth, Berlin

ISO 1302 (2002) Geometrical product specifications (GPS) – indication of surface texture in technical product documentation, ISO. Beuth, Berlin

ISO 13565-3 (1998) Geometrical product specifications (GPS) – surface texture: profile method; surfaces having stratified functional properties – part 3: height characterization using the material probability curve, ISO. Beuth, Berlin

ISO 25178-2 (2012) Geometrical product specifications (GSP) – surface texture: areal part 2: terms, definitions and surface texture parameters, ISO. Beuth, Berlin

ISO 4287 (1997) Geometrical product specifications (GPS) – surface texture: profile method – terms, definitions and surface texture parameters, ISO. Beuth, Berlin
Leach R, Haitjema H (2010) Bandwidth characteristics and comparisons of surface texture measuring instrument. Meas Sci Technol 21(3):032001

Surface Roughness

► Roughness

Surface Texture

Richard Leach
Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, Nottingham, UK

Definition

Surface texture is the geometrical irregularities present at a surface. Surface texture does not include those geometrical irregularities contributing to the form or shape of the surface.

Theory and Application

Most manufactured parts rely on some form of control of their surface characteristics. The surface is usually defined as the feature on a component or device that interacts with the environment, in which the component is housed or in which the device operates, or with another surface (Leach et al. 2014). The surface topography, and of course the material characteristics, of a part can affect how two bearing parts slide together, how fluids interact with the part, or how the part looks and feels (Bruzzone et al. 2008; Thomas 2014).

Surface topography is defined (in Leach 2014) as the overall surface structure of a part (i.e., all the surface features treated as a continuum of spatial wavelengths), surface form as the underlying shape of a part (e.g., a cylinder liner has cylindrical form), and surface texture as the features that remain once the form has been removed (e.g., machining marks on the cylinder liner).

The origins of the surface texture on an object come from a number of sources. Any surface that has been manufactured will have some production process marks, which are inherent and which are caused by the tool removing material or any additive processes. Surface texture also comprises other marks on the surface which are not inherent and which are produced by errors of one sort or another, such as the machine tool lack of stiffness resulting in chatter of the tool against the workpiece, thermal effects such as deflections, or agglomeration of particulates in an additive process, and material effects producing surface irregularities, such as grain boundaries.

Surface Texture Measurement

Surface geometry can be measured using a variety of physical principles, but the most common methods use contacting or optical techniques. The measurement principles fall in to two categories as follows:

1. Those that measure surface topography directly. Such methods produce a topographical image of the surface that may be represented mathematically as a height function, $z(x)$ in the profile case and $z(x,y)$ in the areal case (see below). From the topography image, various filtering processes are used to extract the surface texture.
2. Those that measure a representative area of a surface and produce numerical results that depend on area-integrated properties of the surface. Such methods interrogate physical models of the interaction of the instrument with the surface to directly produce parameters of the surface texture.

The reader is referred to several references for more in-depth treatments of surface measurement (Leach 2014, 2011; Mainsah et al. 2010; Whitehouse 2010).

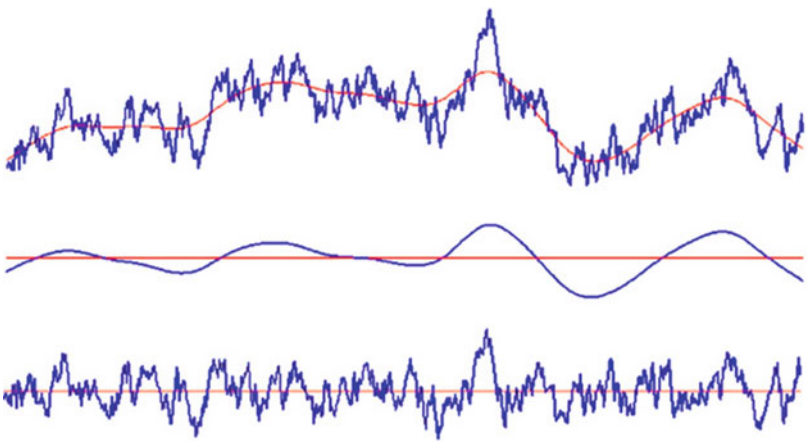
Surface Texture Characterization

Measuring and obtaining the digital representation of surface topography as either a profile $z(x)$ or an areal map $z(x,y)$ is the first step in determining surface texture. Subsequently, the form is removed and the resulting surface texture is filtered to remove (or attenuate) unwanted spatial frequencies. In the profile case, filtering is applied to split the surface texture into primary profile, waviness, and roughness (see Fig. 1) (ISO 4287 1997; Leach 2014; Whitehouse 2010). ISO specification standards define default filter cutoff frequencies for determining primary, waviness, and roughness profiles (ISO 4287 1997; ISO 4288 1996) and which type of filter to use (ISO 16610 part 21 2011). It is important to stress here that the default values are just defaults – they are not requirements. For example, if five sampling lengths would produce an evaluation length that is larger than the length of the surface available for measurement, then fewer sampling length or a smaller sampling length can be used. Finally, surface texture profile parameters can be calculated, the most well-known parameter being *Ra*, but there are many others (see ISO 4287 (1997), Leach (2014) for the ISO parameter descriptions and Whitehouse (2010) for many more).

A surface texture parameter is used to give the surface texture of a part a quantitative value. Such a value may be used to simplify the description of the surface texture, to allow comparisons with other parts (or areas of a part), and to form a suitable measure for a quality system. Surface texture parameters are also used on engineering drawings to formally specify a required surface texture for a manufactured part. Some parameters give purely statistical information about the surface texture and some can describe how the surface may perform in use, that is to say, its functionality (Thomas 2014).

Whereas the profile method may be useful for showing manufacturing process change, much more functional information about the surface can be gained from an analysis of the areal surface texture. There are a number of significant differences between profile and areal analysis. Firstly, whereas it may be possible to use the profile method to control quality once a machining process has been shown to be sufficiently stable, for problem diagnostics and function prediction, an areal measurement is often required. Also, with profile measurement and characterization, it is often difficult to determine the exact nature of a topographic feature. Lastly, an areal measurement will have more statistical significance than an equivalent profile measurement, simply because there are more data points and an areal map is a closer representation of the "real surface."

Distinct from the profile system, areal surface characterization does not require three different groups (profile, waviness, and roughness) of surface texture parameters. For example, in areal parameters, only *Sa* is defined for the average parameter rather than the primary surface *Pa*, waviness *Wa*, and roughness *Ra* as in the profile case (ISO 25178 part 2 2012; Leach 2013).



Surface Texture, Fig. 1 *Top*: primary profile, *center*: waviness, *bottom*: roughness (Courtesy Leach 2014)

The meaning of the *Sa* parameter depends on the type of scale-limited surface used. Filtering for areal analysis again requires the use cutoff frequencies (called nesting indexes) to define the spatial bandwidth for analysis. Defaults are given in ISO specification standards (ISO 25178 part 3 2012; Leach 2013), and there are a range of filter types to choose from (Leach 2013) depending on the application. The ISO areal surface texture parameters are divided into two categories: field parameters, defined from all the points on a scale-limited surface, and feature parameters, defined from a subset of predefined topological features from the scale-limited surface. The areal parameters are described in detail elsewhere (Leach 2013), including a range of case studies illustrating their use.

Cross-References

- ▶ Roughness
- ▶ Surface Parameter

References

Bruzzone AAG, Costa HL, Lonardo PM, Lucca DA (2008) Advances in engineering surfaces for functional performance. Ann CIRP 57:750–769

ISO 4287 (1997) Geometrical product specification (GPS) – surface texture: profile method – terms, definitions and surface texture parameters. International Organization of Standardization, Geneva

ISO 4288 (1996) Geometrical product specification (GPS) – surface texture: profile method – rules and procedures for the assessment of surface texture. International Organization of Standardization, Geneva

ISO 16610 part 21 (2011) Geometrical product specifications (GPS) – filtration – part 21: linear profile filters: gaussian filters. International Organization of Standardization, Geneva

ISO 25178 part 2 (2012) Geometrical product specification (GPS) – surface texture: areal – part 2: terms, definitions and surface texture parameters. International Organization for Standardization, Geneva

ISO 25178 part 3 (2012) Geometrical product specification (GPS) – surface texture: areal – part 3: specification operators. International Organization for Standardization, Geneva

Leach RK (2011) Optical measurement of surface topography. Springer, Berlin

Leach RK (2013) Characterisation of areal surface texture. Springer, Berlin

Leach RK (2014) Fundamental principles of engineering nanometrology, vol 2. Elsevier, Amsterdam

Leach RK, Weckenmann A, Coupland JM, Hartmann W (2014) Interpreting the probe-surface interaction of surface measuring instruments, or what is a surface? Surf Topogr Metrol Prop 2:035001

Mainsah E, Greenwood JA, Chetwynd DG (2010) Metrology and properties of engineering surfaces. Kluwer, Boston

Thomas TR (2014) Roughness and function. Surf Topogr Metrol Prop 2:014001

Whitehouse DJ (2010) Handbook of surface and nanometrology, vol 2. CRC Press, Boca Raton

Surface Texture Filtering

Xiangqian Jiang and Paul J. Scott
Centre for Precision Technologies, University of Huddersfield, Huddersfield, UK

Synonyms

Roughness filtering; Surface topography filtering

Definition

A filter separates the small-scale texture from the larger-scale texture in a surface. The value of the scale at the defined separation is called the nesting index although other names are used for specific filters (e.g., cut-off for linear filters).

Scale can be defined in terms of: wavelength for linear filters, size of the structuring element (e.g., radius of a disk) for morphological filters. In Segmentation filters, the scale can be: the height difference between the highest (or lowest) points in the interior and on the boundary of a segment, the area of a segment, length of the boundary of a segment, etc.

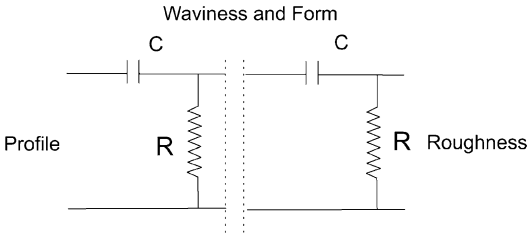
Theory and Application

Introduction

Filtration has always been important in surface metrology: It is the means by which the surface

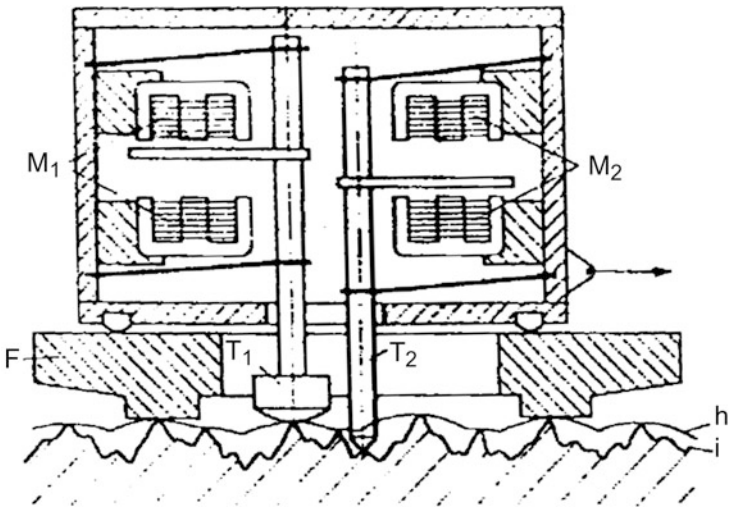
features of interest are extracted from the measured data for further analysis.

The first filters started with the fully analogue 2CR filter implemented as a two-stage Capacitor-Resistance network as shown in Fig. 1. Originally it was designed to remove the DC and slope of the measured surface profile but was quickly realized to have other useful properties, i.e., surface filtration. The 2CR filter formed the basis of the M-system (mean line system) of filtration. Unfortunately, it was found that the 2CR filter could badly distort profile features due to phase shifting of the different harmonic components in the profile. A symmetrical version, the phase corrected 2CR filter, resolved most of this distortion, but it still had problems, one being that it badly distorted the profile at the ends. This led to its eventual replacement by the modern digital Gaussian filter as the general default surface texture filter.



Surface Texture Filtering, Fig. 1 Two-stage 2RC high pass filter (Whitehouse 2002). (Reproduced with permission from DJ Whitehouse)

The Envelope system (E-system) was first developed by Von Weingraber (1956). The E-system bases the reference lines upon the loci of centers of circles of different radii rolled along the profile. As Fig. 2 demonstrates, the locus of the center of the larger tip gives the curve of form, while that of the smaller tip gives the contacting profile. The difficulty appeared in building practical instruments as two elements are needed: a spherical skid (T1) to approximate the "enveloping circle" and a needle-shaped stylus (T2) moving in a diametric hole of the skid to measure the roughness as deviation with respect to the "generated envelope." The advantages of the E-system were claimed to be that it is more physically significant in that many engineering properties of a surface are determined by its peaks. Standard radii were 25 mm for roughness and 250 mm for waviness, though other radii have been proposed by Radhakrishnan (1972). The standing objection was that the choice of the rolling circle radius is as arbitrary as the choice of cut-off in the M-system and no practical instrument using mechanical filters could be made at the time. The advent of fast practical computers, which can be used in association with measurement instruments, has virtually eliminated the need for any hardware implementation for the E-system (Tholath and Radhakrishnan 1999); today digital morphological filters are used instead.



Surface Texture Filtering, Fig. 2 Probe for E-system: T1 Skid, T2 stylus. (Reprinted from (Peters et al. 2001) with permission from Elsevier)

The French automobile industry adopted an alternative approach to filtration called R&W (roughness and waviness). The method began as a purely graphical approach, where an experienced operator would draw on a profile graph an upper envelope that subjectively joined the highest peaks of the profile. This was an attempt at a simulation of the E-system. The technique was based on the concept of the "motif": a motif is the portion of a profile between local peaks. The basic approach (for both the graphical and the digital versions) was to determine all motifs between adjacent local peaks; then a series of rules would combine "insignificant" motifs with neighboring motifs, to create larger combined motifs, until only "significant" motifs were left, from which surface texture parameters could be calculated, see Fig. 3. The R&W approach proved to be very unstable with very small changes in the profile producing significant changes in the analysis. Today stable digital segmentation filters are being proposed as a replacement.

In 1996, ISOTC/213: *Dimensional and geometrical product specifications and verification* set up a working group to investigate filtration. The Gaussian filter, although a good general filter, is not applicable for all functional aspects of a surface, for example, in contact phenomena, where the upper envelope of the surface is more relevant. This work has resulted in the establishment of a standardized framework for filters, giving a mathematical foundation for filtration, together with a toolbox of different filters. Information concerning these filters has been published as a series of ISO standards (ISO 16610 series). The core parts of the ISO 16610 series of filters are now described in the next sections.

Linear Filters

General

Linear filters: The mean line filters (M-system) belong to this class and include the Gaussian filter (ISO 11562 1996, replaced by ISO 16610-21 2011), Spline filter (Krystek 1996, 1997), and the Wavelet filter (Jiang et al. 2000).

A linear filter is a filter operation $w(x) = F(z(x))$, that is also a linear functional, which satisfies:

$$F(y(x) + z(x)) = F(y(x)) + F(z(x)) \;\&\; F(a \times z(x)) = a \times F(z(x)) \quad (1)$$

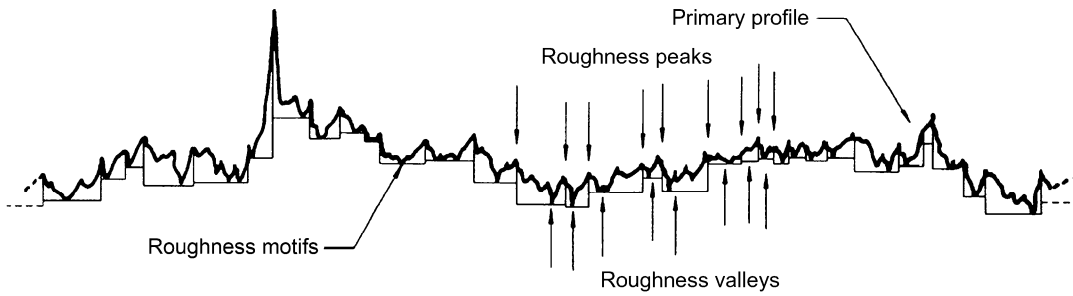
where $z(x)$ and $y(x)$ are the input data/measurement data, $F(\cdot)$ is the filter operation to get the filtered data $w(x)$, and a is an arbitrary real numbers.

Linear filters are usually defined by their associated weighting function $s(x)$ which is used to calculate the filtered mean line $w(x)$. The weighting function indicates for each point x the weight attached by the profile in the vicinity of that point, that is, to say the filtered data $w(x)$ is a convolution of the input data/measurement data $z(x)$ and the weighting function $s(x)$.

$$w(x) = z(x) * s(x) \quad (2)$$

Gaussian

The Gaussian filter is the most popular standardized filtering technique for surface texture. Gaussian filtration follows the convolution



Surface Texture Filtering, Fig. 3 R&W filtration into motifs (ISO 12085 1996)

procedure which convolves the input data/measurement data $z(x)$ with a Gaussian weighting function $s(x)$ to get the filtered data $w(x)$. The Gaussian weighting function for a profile has the form of a bell-shaped curve as shown in Fig. 4 is defined in (ISO 11562 1996; ISO 16610-21 2011):

$$s(x) = \begin{cases} \left(\frac{1}{\alpha\lambda_c}\right)\exp\left(-\pi\left(\frac{x}{\alpha\lambda_c}\right)^2\right), & -L_c\lambda_c < x < L_c\lambda_c \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where, L_c is a truncation constant with default value 1, α is given by $\alpha = \sqrt{ln2/\pi} \approx 0{,}4697$, to provide a 50% transmission characteristic at the cut-off wavelength λ_c. Figure 4 shows a weighting function of the linear profile Gaussian filter.

Compared with the historically 2RC filter (Whitehouse 2002), the most important advantage of the linear Gaussian filter is its zero/linear phase property, and 50% transmission at the cut-off wavelength as shown in Fig. 5. The linear Gaussian filter has the advantage that it has very efficient and fast algorithms and thus can process large data sets very quickly.

The linear areal Gaussian filter for areal surface analysis can be seen as a procedure of the measured areal surface convolving with a Gaussian weighting function, while the Gaussian weighting has the same shape over each data point. The Gaussian weighting function (Fig. 6) is defined as:

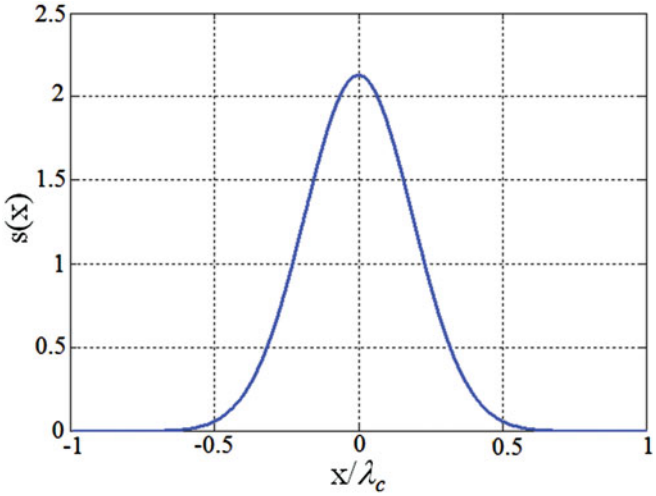
$$s(x,y) = \left(\frac{1}{\alpha^2\lambda_{cx}\alpha\lambda_{cy}}\right)\exp\left[-\pi\left(\frac{x}{\alpha\lambda_{cx}}\right)^2 - \pi\left(\frac{y}{\alpha\lambda_{cy}}\right)^2\right] \quad (4)$$

Linear areal Gaussian filter can be implemented by using the linear profile Gaussian filtration in the x-direction followed by the linear profile Gaussian filtration in the y direction, or vice versa. Areal Gaussian filters are standardized in ISO 16610-61 (2015). An example can be seen in Fig. 7.

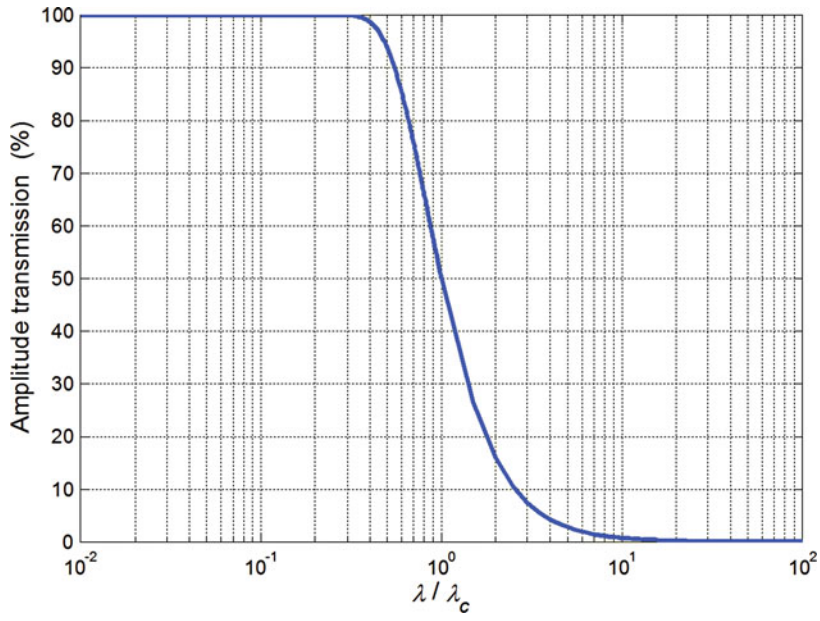
For linear Gaussian filtering of surfaces, the following preconditions apply: (*a*) the measured profile/areal surface will be truncated at the boundaries of both sides to remove the effect of boundary distortion, so a filtered surface is much smaller than the measured surface (see the right side of Fig. 7); (*b*) the form must be removed before filtering; (*c*) the surface texture is assumed to be nearly symmetric otherwise there will be distortion due to the nonsymmetry, e.g., Plateau honed surfaces (ISO 13565-1).

Spline

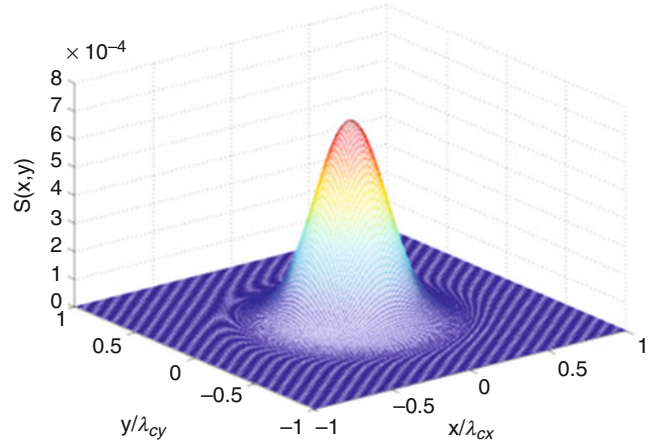
Another type of linear filter is the Spline filter, originally proposed by M. Krystek (1996, 1997). The weighting function of a spline filter cannot be given by a simple closed formula. A filter



Surface Texture Filtering, Fig. 4 weighting function of the linear profile Gaussian filter (Eq. 3)



Surface Texture Filtering, Fig. 5 Amplitude transmission characteristic (transfer function) of the Linear Gaussian filter



Surface Texture Filtering, Fig. 6 weighting function of the linear areal Gaussian filter (Eq. 4)

equation is therefore used instead of the weighting function to describe the profile spline filter (ISO 16610-22 2015):

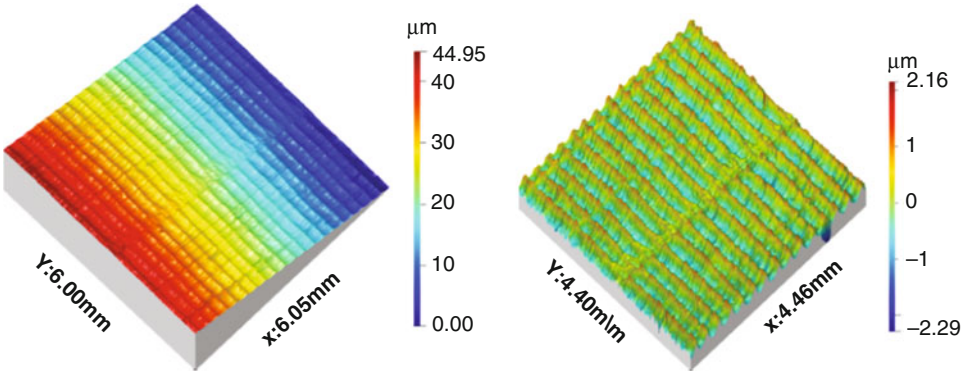
$$\left[\mathbf{I} + \beta\alpha^2\mathbf{P} + (1-\beta)\alpha^4\mathbf{Q}\right]\mathbf{W} = \mathbf{Z} \qquad (5)$$

where, $\mathbf{I}$ is an $n \times n$ identity matrix. $\mathbf{P}$ is an $n \times n$ tridiagonal symmetric matrix, Z is the vector of the measurement data, W is the vector of the filtered data, Q is an $n \times n$ five-diagonal symmetric matrix, $\alpha = 1/2 \sin(\pi\Delta x/\lambda c)$ and β is tension parameter which controls how tightly the spline curve fits through the data points.

Compared with the linear Gaussian filter, the linear spline filter has two advantages: there is a reduced boundary effect, which means the measured surface can be fully used; and it has very good form-following property, which means that it can be used as a form filter (Fig. 8).

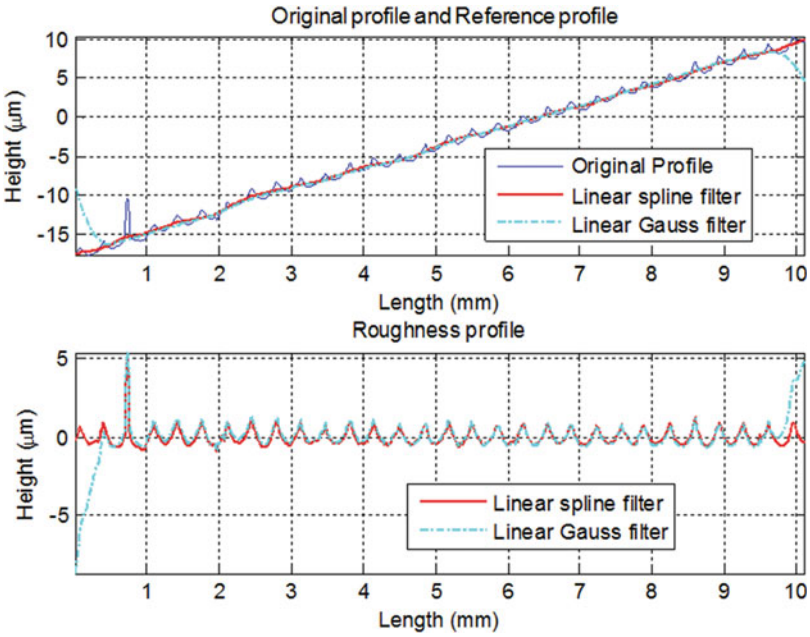
Wavelet

It is recognized that the manufacture of ultra-precision surfaces leaves multiscale topography signatures within the topography of the surface. In practice, this means that a surface will contain fine-scale texture, from multiple sources, superimposed on other scales of texture.



Surface Texture Filtering, Fig. 7 Milled surface filtered by linear areal Gaussian filter (Left: measured surface without tilt removal; Right: small scale (roughness) surface $\lambda_{cx} = \lambda_{cy} = 0.8\ mm$)

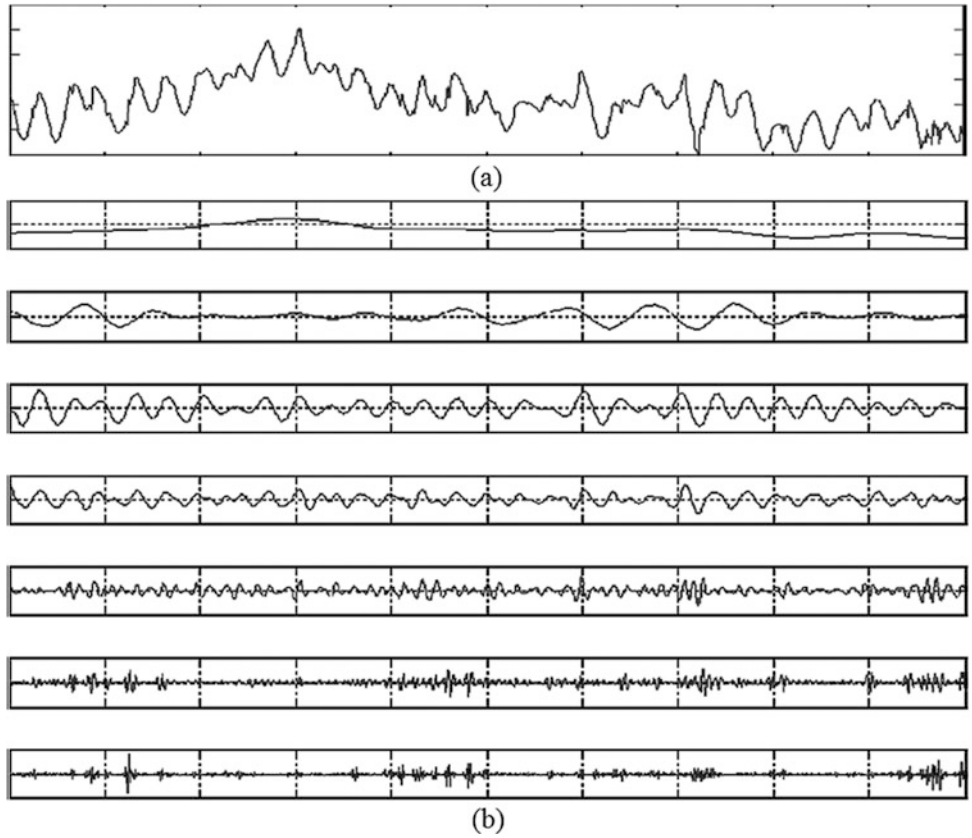
Surface Texture Filtering, Fig. 8 Comparison of the linear spline filter and linear Gaussian filter (milling profile $\lambda c = 0.8\ mm$)



Wavelet theory provides a new analysis tool capable of resolving this fundamental problem; it employs space-frequency windows and offers the relevant space-frequency analysis. Wavelet transforms can divide functions into different scale – frequency components – and then each component can be studied with a resolution matched to its scale (Chui 1992). Progress in wavelet analysis created an opportunity that enabled researchers to use orthogonal wavelets to decompose turned, milled, and ground surfaces, as well as evaluation of tool marks, machining vibrations, and machine-tool errors. An example of a milled surface is shown in Fig. 9.

Multiscale analysis can be seen as multiple band-pass filters, which can output different frequency band signals, while keeping their space/time information. Jiang et al. (1997), however, soon identified the phase distortion problems in orthogonal wavelets. Consequently, a biorthogonal wavelet model, using a fast algorithm called "the lifting scheme" (firstly developed at Bell Laboratory by Swelden (1995)), Jiang et al. (2000) proposed a new wavelet

S



Surface Texture Filtering, Fig. 9 Multiscale analysis of a milling surface: (**a**) Milled profile; (**b**) multiscale wavelet representation

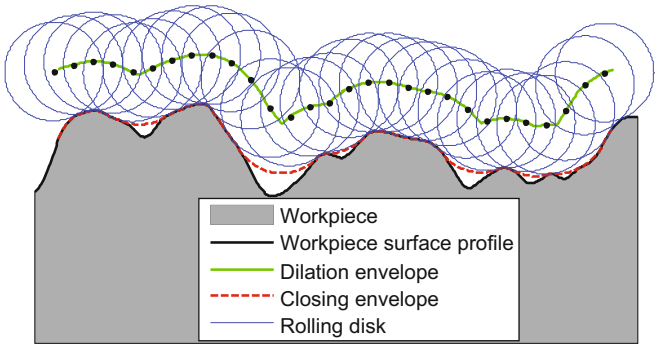
representation for use with multi-scale surface topography. This biorthogonal wavelet overcomes the phase distortion problems that occur with orthogonal wavelets (Jiang et al. 2000; ISO 16610-29 2015).

The biorthogonal wavelet transform is very good for isotropic surfaces but has difficulty with surface topography with various scale scratches, such as those in plateau-honed surfaces and worn biomedical surfaces because in a discrete biorthogonal wavelet transform a small shifts in position of the surface can result in a completely different distribution of "energy" among the wavelet components. In order to extract and reconstruct a surface topography with various scale scratches, such as those in plateau-honed surfaces and worn biomedical surfaces, a complex wavelet model (Ma et al. 2005) was developed, according to Kingsbury's dual-tree complex wavelet theory (Kingsbury 1999) to avoid shift-variant problems from affine shifts of the surface. The dual-tree complex wavelet is designed to minimize the shift-variant problem and reconstruction of scratches has considerably less distortion. The metrological characteristics of the complex wavelet were examined (Zeng et al. 2005).

Morphological Filters

Morphological filters: The envelope filters (E-system) belong to this class and include closing and opening filters using either a disk or a horizontal line, ISO 16610-41 (2015). The E-system, which had problems as a purely mechanical approach, can be simulated digitally (Scott 1992; Srinivasan 1998).

Morphological filters emerged as the superset of the early envelope filter but offer more tools



Surface Texture Filtering, Fig. 10 Dilation and closing envelope of an open profile by a disk

and capabilities by incorporating mathematical morphology theory. The basic variation of morphological filters includes the closing filter and the opening filter, see ISO 16610-40 (2015). The closing filter is obtained by placing an infinite number of identical spherical balls in contact with the object's surface from the air side and taking the lower boundary of the balls (see Fig. 10). Likewise, the opening filter is archived by rolling the balls from the material side of the object. The closing filter suppresses surface valleys whose curvatures are larger than that of the ball; meanwhile the peaks remain unchanged. Conversely, the opening filter suppresses the large peaks and retains the valleys. They can be even combined to achieve superimposed effects, referred to as alternating symmetrical filters, by which surface peaks and valleys are both suppressed, ISO 16610-49 (2015). Scale-space techniques further developed morphological filters, which provide a multi-resolution analysis to surface textures whereby various scales of geometrical features can be extracted from a surface and assessed separately (Scott 2000).

Morphological filters are useful for surface filtration. However, the broader application of morphological filters was restricted by the traditional algorithms based on image processing techniques. The implement of traditional morphological filters are: very time-consuming for large data set and large structuring elements, only calculate the extreme values in height restricted to "planar" surfaces, i.e., height maps. Recently developed computational methods based on the alpha shape have removed these limitations, which enable the computation of closing and opening operations on arbitrary shapes of surfaces (Lou et al. 2013). The Alpha shape method is based on the Delaunay triangulation from which the boundary facets of the alpha shape are extracted. Figure 11 is an example of computing the closing envelope of a profile data.

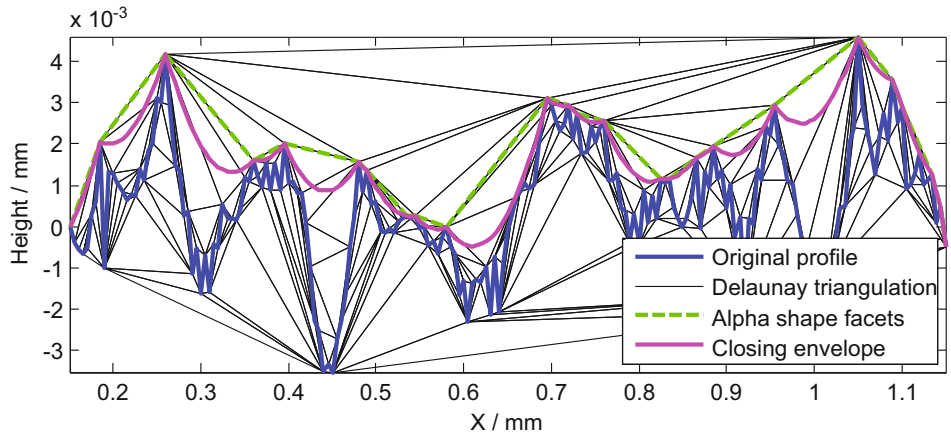
Morphological filters are considered as useful tools for the functional evaluation of component surfaces, e.g., wear. Figure 12 illustrates an example of using morphological closing filter to extract the wear marks of an artificial knee joint component. The left of Fig. 12 shows a portion of the worn joint surface, presenting significant deep scratches produced during its functional life. The right side of Fig. 12 is the residual surface obtained by subtracting the envelope surface from the original surface using a morphological closing filter with a 5 mm ball.

Robust Filters

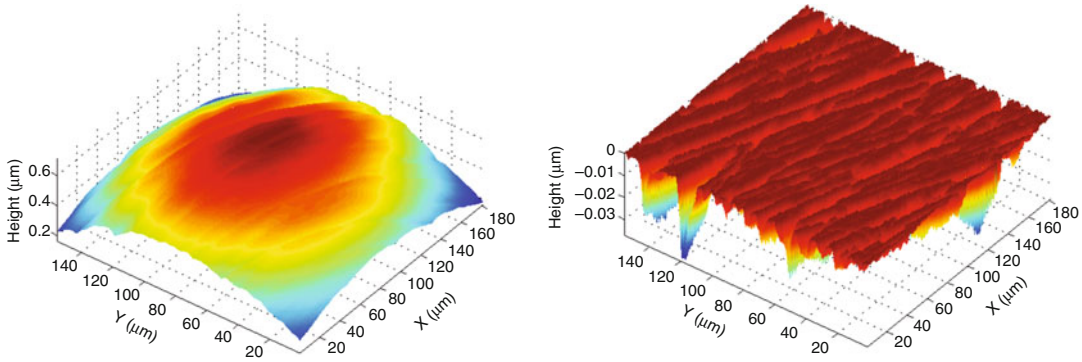
General

Robust filters are defined as filters that are robust with respect to specific profile/surface phenomena such as spikes, scratches, steps etc. These filters include the Robust Gaussian filter (Seewig 1999, 2005). There are several approaches to design a robust filer of which the following are defined in ISO 16610-30 (2015):

Metric based approaches: Here, the metric used to fit the filtered profile to the profile is altered to a more "robust" metric. For example, the metric based on the L1-norm is more robust against spike discontinuities than the metric based on the least squares norm (L2-norm),



Surface Texture Filtering, Fig. 11 Alpha shape facets extracted from the Delaunay triangulation of a profile data (Lou et al. 2013)



Surface Texture Filtering, Fig. 12 The morphological closing filter on a worn artificial knee replacement component. Left: original worn surface; Right: residual surface

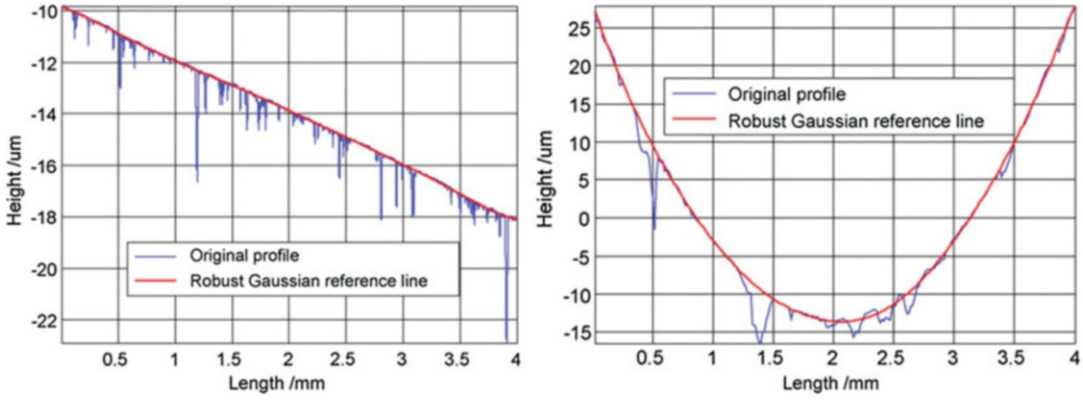
which in turn is more robust than the metric based on the Chebychev norm (L∞-norm);

Statistic based approaches: Each point on the profile is weighted according to an influence function using the filter's low-pass response as a reference. As a result, points further away are given less relative weight than would be the case with points nearer to the low pass response. There are several common influence functions used to allocate the weights to points (Huber, Beaton, bi-weight functions, etc.);

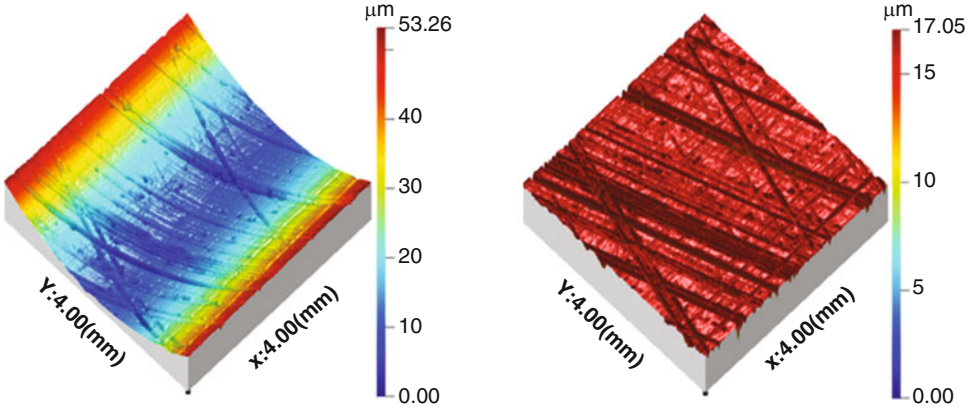
Preprocessing approaches: A technique where an unwanted phenomenon in the profile/surface is removed or greatly reduced, by other means, before filtration, thus removing or greatly reducing any effect the unwanted phenomenon can have on the filter's response This approach has the advantage that once a method has been found to remove unwanted phenomena, then it will work with any filter. Preprocessing based on Wavelets or Scale-space filtration is contained in ISO 16610-30 (2015).

Robust Gaussian Regression Filter

This filter uses a statistical-based approach and is a combination of a Gaussian filter, a bi-weight influence function and an M-estimator based on the local polynomial modelling of the profile/surface of up to a second order polynomial, see ISO 16610-31 (2016) and ISO 16610-71 (2014) for details.



Surface Texture Filtering, Fig. 13 Robust Gaussian filtrating for the stratified surface profiles. Left: honed surface profile (first order Robust Gaussian Regression filter); Right: honed surface profile with form components (second order Robust Gaussian Regression filter). (Reprinted from (Jiang 2010) with permission from Elsevier)



Surface Texture Filtering, Fig. 14 Cylinder surface filtered by second order Robust Gaussian Regression filter (Left: measured surface; Right: small scale (roughness) surface)

Figure 13 demonstrates the ability of the first order robust Gaussian regression filter for processing a honed profile (left side). The right side of Figure shows the merits and flexibility of the second order robust solution: It does not suppress outlier distortion yet but can remove form components from the measurement data without distorted boundaries.

Figure 14 illuminates that the second order Robust Gaussian Regression filtering for areal surface analysis and shows a result from the robust second order Gaussian regression filter against outliers on the measured cylinder surface (left), which maintains all small-scale components (roughness) (right) phenomena.

The advantages of this filter is that not only is it robust against spike discontinuities, such as deep valleys and high peaks and has a much reduced end effect as compared to the Gaussian filter, but also if a second order polynomial is incorporated, the robust Gaussian regression filter follows form components up to the second degree.

Segmentation

General

Segmentation filters are filters that partition a profile/surface into portions according to specific rules. The motif approach belongs to this class and has now been put on a firm mathematical

basis, using pattern analysis, which has solved the stability problem in R&W techniques and other segmentation filters (Scott 1992, 2004).

Segmentation is a filtration operation that spatially decomposes a surface into mutually exclusive portions of that surface. Segmentation filtration requires: the definition of the objects being filtered, i.e., the segments; the rule for segment combination and the rule which states which segments are significant/insignificant. Associated with each segmentation method is a nesting index such that large values of the nesting index correspond to large surface portions and smaller values of the nesting index correspond to smaller surface portions.

Watershed Segmentation

The watershed method consists of gradually filling insignificant dales/motifs with water. The water will eventually flow out of each dale/motif, into an adjacent dale/motif. If the dale/motif is significant combine the two dales/motifs. Otherwise continue to fill the new lake until the water flows into a significant dale/motif. All the filled insignificant dales are then combined with the significant dale. Segmentation that uses the watershed method as the combination rule is called watershed segmentation. A nesting index threshold is used to determine which dales/motifs are significant or insignificant. Here the scale used, to define the nesting index, is the height difference between the highest (or lowest) points in the interior and on the boundary of a dale/motif, see ISO/WD 16610-45 2018 (profile) and ISO 16610-85 for further details.

Watershed segmentation is the proposed replacement for R&W segmentation, which is retained in ISO/WD 16610-45 2018 for legacy reasons until watershed segmentation is generally accepted by the surface texture community.

Crossing-the-Line Segmentation (profile only)

Crossing-the-line segmentation is based on crossings of the reference line (mean line, *x*-axis etc.) by a scale-limited profile. The scale-limited profile is initially segmented based on crossings of the reference line, and a procedure is then applied to prune out the least significant through a combination algorithm (based on discrimination thresholds for the peaks and for the valleys) to leave a set of significant segments which is the resulting crossing-the-line segmentation, see ISO/WD 16610-45 2018 (profile) for further details.

The nesting index for crossing-the-line segmentation is defined as the vector (Hu, Hl): where Hu is the height discrimination threshold for the peaks and Hl is the depth discrimination threshold for the valleys, both values shall be predefined.

Cross-References

- ▶ Metrology
- ▶ Micro Geometry
- ▶ Roughness
- ▶ Surface Parameter

References

Chui CK (1992) An introduction on wavelets. SIAM, Philadelphia

ISO 11562 (1996) Geometrical product specifications (GPS) – surface texture: profile method: metrological characteristics of phase correct filters. International Organization for Standardization, Geneva

ISO 12085 (1996) Geometrical product specifications (GPS) – surface texture: profile method: motif parameters. International Organization for Standardization, Geneva

ISO 16610 part 21 (2011) Geometrical product specification (GPS) – filtration: linear profile filters: Gaussian filters. International Organization for Standardization, Geneva

ISO 16610 part 22 (2015) Geometrical product specification (GPS) – filtration: linear profile filters: spline filters. International Organization for Standardization, Geneva

ISO 16610 part 29 (2015) Geometrical product specifications (GPS) – linear profile filters: spline wavelets. International Organization for Standardization, Geneva

ISO 16610 part 30 (2015) Geometrical product specifications (GPS) – robust profile filters: basic concepts. International Organization for Standardization, Geneva

ISO 16610 part 31 (2016) Geometrical product specifications (GPS) – robust profile filters: Gaussian regression filters. International Organization for Standardization, Geneva

ISO 16610 part 40 (2015) Geometrical product specifications (GPS) – morphological profile filters: basic concepts. International Organization for Standardization, Geneva

ISO 16610 part 41 (2015) Geometrical product specifications (GPS) – morphological profile filters: disk and horizontal line-segment filters. International Organization for Standardization, Geneva

ISO 16610 part 49 (2015) Geometrical product specifications (GPS) – morphological profile filters: scale space techniques. International Organization for Standardization, Geneva
ISO 16610 part 61 (2015) Geometrical product specification (GPS) – filtration: linear areal filters: Gaussian filters. International Organization for Standardization, Geneva
ISO 16610 part 71 (2014) Geometrical product specifications (GPS) – robust profile filters: robust areal filters: Gaussian regression filters. International Organization for Standardization, Geneva
ISO 16610 part 85 (2013) Geometrical product specifications (GPS) – filtration: morphological areal filters: segmentation. International Organization for Standardization, Geneva
ISO/WD 16610 part 45 (2018) Geometrical product specifications (GPS) – morphological profile filters: segmentation. International Organization for Standardization, Geneva
Jiang X (2010) Robust solution for the evaluation of stratified functional surfaces. CIRP Ann Manuf Technol 59(1):573–576
Jiang X, Blunt L, Stout KJ (1997) Recent development in the characterization technique for bioengineering surfaces, In: Transactions of 7th international conference of metrology and properties of engineering surface, 2–4 April 1997, ISBN 91-7197-470-9, pp 215–221
Jiang X, Blunt L, Stout KJ (2000) Development of a lifting wavelet representation for surface characterization. Proc R Soc Lond A 456:2283–2313
Kingsbury N (1999) Image processing with complex wavelets. Phil Trans R Soc A 357:2543–2560
Krystek M (1996) Form filtering by splines. Measurement 18:9–15
Krystek M (1997) Transfer function of discrete spline filters. In: Ciarlini P et al (eds) Advanced mathematical tools in metrology III. World Scientific Publishing, Singapore, pp 203–210
Lou S, Jiang X, Scott PJ (2013) Geometric computation theory for morphological filtering on freeform surfaces. Proc R Soc A 469(2159):20130150
Ma J, Jiang X, Scott PJ (2005) Complex ridgelets for shift invariant characterization of surface topography with line singularities. Phys Lett A 344:423–431
Peters J, Bryan JB, Elster WT, Evans C, Kunzmann H, Lucca DA et al (2001) Contribution of CIRP to the development of metrology and surface quality evaluation during the last fifty years. CIRP Ann Manuf Technol 50(2):471–488
Radhakrishnan V (1972) Selection of an enveloping circle radius for E-system roughness measurements. Int J Mach Tools Manuf 12:151–159
Scott PJ (1992) The mathematics of motif combination and their use for functional simulation. Int J Mach Tool Manu 32(1/2):69–73
Scott PJ (2000) Scale-space techniques. In: Proceedings of the X. Internationales Oberflachenkolloquium, Chemnitz
Scott PJ (2004) Pattern analysis and metrology: the extraction of stable features from observable measurements. Proc R Soc Lond A 460:2845–2864
Seewig J (1999) Praxisgerechte Signalverarbeitung zur Trennung der Gestaltabweichungen technischer Oberflächen [Practical signal processing for the separation of the shape deviations of technical surfaces], Shaker Verlag GmbH. Herzogenrath, Aachen, Germany
Seewig J (2005) Linear and robust Gaussian regression filters. J Phys Conf Ser 13:254–257
Srinivasan V (1998) Discrete morphological filters for metrology. In: Proceedings of the 6th ISMQC IMEKO symposium on metrology for quality control in production, Viden, Czech Republic, pp 623–628
Swelden W (1995) The lifting scheme: a custom-design construction of biorthogonal wavelets. Appl Comput Harmon Anal 3(2):186–200
Tholath J, Radhakrishnan V (1999) Three-dimensional filtering of engineering surfaces using envelope system. Precis Eng 23:221–228
Von Weingraber H (1956) Zur definition der Oberflachenrauigkeit. Werkstattstechnik, Masch, Bass 46
Whitehouse DJ (2002) Surfaces and their measurement. Hermes Penton Science, London
Zeng W, Jiang X, Scott P (2005) Metrological characteristics of dual tree complex wavelet transform for surface analysis. Meas Sci Technol 16:1410–1417

Surface Texture Measuring Instrument Specifications

► Surface Texture Metrological Characteristics

Surface Texture Metrological Characteristics

Han Haitjema[1,2] and Richard Leach[3]
[1]Mitutoyo RCE, Best, The Netherlands
[2]KU Leuven, Department of Mechanical Engineering, Leuven, Belgium
[3]Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, Nottingham, UK

Synonym

Surface texture measuring instrument specifications

Definition

The surface texture metrological characteristics establish a common calibration framework for surface topography measuring instruments. A metrological characteristic is defined in ISO specification standards as a "Characteristic of measuring equipment, which may influence the results of measurements."

Theory and Application

Introduction

In general, surface topography measuring instruments measure the height z of a surface, relative to a nominally horizontal x-y plane. For instruments that scan in the horizontal plane, the x-axis is the tracing axis, and the y-axis is the stepping axis. For optical measurements where the x-y plane is projected on a camera, the pixels correspond to the (x,y) coordinates, and the z coordinates are obtained by some operation, e.g., from a stack of images while the surface is moved in the z-direction.

The metrological characteristics of surface topography measuring instruments are given in the literature (Leach 2011; Leach et al. 2015) and in several ISO specification standards. Several standards (ISO 25178-601 2010; ISO 25178-602 2010; ISO 25178-603 2013; ISO 25178-604 2013; ISO 25178-605 2014; ISO 25178-606 2015) have been developed to define terms and metrological characteristics for individual measurement methods. At the time of writing, the common aspects of the methods are being summarized into one standard: ISO DIS 25178-600 (2018). The metrological characteristics for surface topography measuring instruments are summarized in ISO DIS 25178-600 and presented in Table 1.

In the following sections, the metrological characteristics and their determination (calibration) will be briefly explained.

Amplification Coefficient α_x, α_y, α_z

The amplification coefficient is the slope of the linear regression curve obtained from the response function. Amplification coefficients apply to the x, y, and z coordinates. The ideal response is a straight line with a slope equal to 1, which means that the values of the measurand are equal to the values of the input quantities. The amplification coefficient quantity may also be termed "scaling factor."

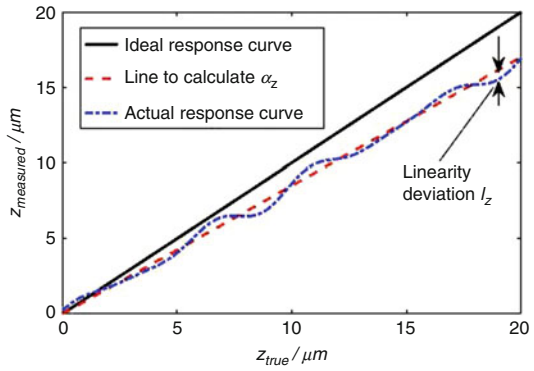
Surface Texture Metrological Characteristics, Table 1 Metrological characteristics for surface topography measuring instruments

Metrological characteristic	Symbol	Main potential error along
Amplification coefficient	α_x, α_y, α_z	x, y, z
Linearity deviation	l_x, l_y, l_z	x, y, z
Flatness deviation	z_{FLT}	z
Measurement noise	N_M	z
Topographic spatial resolution	W_R	z, x, y
x-y mapping deviation	$\Delta_x(x,y)$, $\Delta_y(x,y)$	x, y
Topography fidelity	T_{FI}	x, y, z

Linearity Deviation l_x, l_y, l_z

The linearity deviation is the maximum local difference between the curve from which the amplification coefficient is derived and the response function.

For the linearity and amplification in the z-direction, measurement of step heights is the default method for determination of the amplification coefficient α_z and linearity deviation l_z. The measured depth is compared with the calibrated depth of the material measure (Giusca et al. 2012b). The major drawback of this type of artifact is that they only reproduce a single value, usually much smaller than the instrument's measurement range; consequently, it is often not possible to establish the amplification coefficient to within a useful uncertainty. A solution to this drawback is to use multiple step height artifacts of different values (Giusca et al. 2012a) or a material measure with multiple heights of different values (Brand et al. 2006). Alternative methods may use interferometric displacement sensors, either directly (de Groot and Beverage 2015) or indirectly by a separate calibration device, a so-called vibration table (Haitjema and Morel 2005a; Koops et al. 2015). An efficient way of checking the z-linearity deviation l_z in a limited range is by measuring a tilted optical flat (see ISO 12179 2000).



Surface Texture Metrological Characteristics, Fig. 1 Example of a response curve and derivation of amplification factor and linearity deviation

For the *x*-*y* calibration, the measurement of a calibrated line scale, calibrated sinusoidal profile, or a calibrated cross-grating is most appropriate (Giusca et al. 2012a). A virtual lateral calibration standard has been developed where a specific feature is given a known lateral displacement while it is measured or scanned by the topography measuring instrument (Koops et al. 2016).

Figure 1 shows an example of how quantities α_z and l_z can be derived from a response curve that is obtained by calibration.

Flatness Deviation z_{FLT}

The flatness deviation is the deviation of the measured topography of an ideally flat object from a plane. Flatness deviation can be caused by residual flatness of an imperfect areal reference or by imperfection in the optical setup of an instrument.

An effective way of determining the flatness deviation is the measurement of an optical flat. Commonly, the flatness deviation of optical flats is in the nanometer region for millimeter square regions and can be negligible compared to the flatness deviation of the instrument being calibrated. The influence of the used optical flat can be reduced further by taking a measurement at various locations on the optical flat and averaging the result for every point (pixel) of the measurement (Evans and Tyler Estler 1996; Giusca et al. 2014; Creath and Wyant 1990). Further reduction of the influence of the used flat surface can be achieved by considering the randomness of the Fourier components of the surface (Haitjema and Morel 2005b).

Measurement Noise N_M

Measurement noise is the noise added to the output signal occurring during the normal use of the instrument. Measurement noise includes the instrument noise as well as components arising from the environment (thermal, vibration, air turbulence) and other sources. The instrument noise may be due to electronic noise, such as that arising in amplifiers, or optical noise, such as that arising from stray light or from fluctuations in the intensity of the used light source. The noise is commonly specified as a standard deviation in the *z* coordinates, possibly with a specified filtering applied.

Commonly, only the measurement noise in the *z*-direction is relevant and is determined. An effective method to determine the measurement noise is by repeating the measurement and determining the standard deviation in every coordinate (Haitjema and Morel 2005b; Giusca et al. 2014).

Topographic Spatial Resolution W_R

The topographic spatial resolution is the metrological characteristic describing the ability of a surface topography measuring instrument to distinguish closely spaced surface features. Several parameters and functions may be used to quantify the topographic spatial resolution, depending on the application and the method of measurement. These include:

- Lateral period limit DLIM, that is, the spatial period of a sinusoidal profile at which the height response falls to 50%
- Stylus tip radius rTIP in case of a mechanical measurement
- Lateral resolution Rl, that is, the smallest distance between two features which can be measured
- Width limit for full height transmission Wl, that is, the width of the narrowest rectangular groove whose step height is measured within a given tolerance
- Small-scale fidelity limit TFIL, that is, smallest lateral surface feature for which the reported topography parameters deviate from accepted values by less than a specified amount, e.g., 10%

- Rayleigh criterion, that is, the quantity characterizing the optical lateral resolution given by the separation of two point sources at which the first diffraction minimum of the intensity image of one point source coincides with the maximum of the other
- Sparrow criterion, that is, the quantity characterizing the optical lateral resolution given by the separation of two point sources at which the second derivative of the intensity distribution vanishes between the two imaged points

The latter two parameters are useful for characterizing the instrument response to features with heights much less than the used effective light wavelength for optical surface topography measuring instruments.

For optical instruments, it is common to specify the topographical spatial resolution based on the theoretically achievable lateral resolution using the numerical aperture of the used objective. However, based on the measurement principle and evaluation, the lateral resolution can be much worse in practice (Giusca and Leach 2013; de Groot et al. 2012). The measurement/calibration of the topographic spatial resolution can be carried out using material measures, such as star-shaped (Giusca and Leach 2013; Xu et al. 2012) or chirped artifacts characterized by uniform amplitude (Krüger-Sehm et al. 2007) or with varying amplitudes (Fujii et al. 2011).

For mechanical measurements, the stylus tip radius is an appropriate measure, provided the stylus follows the surface faithfully.

x-y Mapping Deviation Δ_x (*x,y*), Δ_y (*x,y*)

The *x-y* mapping deviation is generally displayed as a gridded image of *x*- and *y*-deviations of actual coordinate positions on a surface from their nominal positions. The mapping deviations may be used to calculate the *x*- and *y*-amplification and linearity deviations and the *x-y* axis perpendicularity.

The *x-y* mapping is most conveniently carried by the use of a calibrated grid (Giusca et al. 2012a).

Topography Fidelity T_{FI}

Topographic fidelity is the closeness of agreement between a measured surface profile or measured topography and reference values whose uncertainties are insignificant by comparison.

A material measure may be used for the determination of the topography fidelity. The shape of the material measure should be close to the shape of the measurand. To quantify the topography fidelity, a mathematical model or calibrated values of the topography represent the shape of the calibrated material measure. An example structure is a square-profile grating having a depth and periodicity consistent with the intended measured surface structure.

Additional Characteristics

Other metrological characteristics are in use that are not currently in the ISO 25178-600. A few are mentioned here for completeness and reference.

Maximum Measurable Local Slope

The maximum measurable local slope is the greatest local slope of a surface feature that can be assessed by the measurement system. The maximum measurable slope is an important limitation to be specified for a surface topography measurement instrument. However, calibration of this parameter is rarely useful. For optical systems, the maximum measurable local slope is commonly considered as dependent of the numerical aperture of the used objective, and for mechanical probing systems, it depends on both the probe tip radius and the cone angle. However, this quantity depends on the definition of "local" (e.g., a small local roughness on a tilted area of a specimen); therefore, it is difficult to give a generally valid specification.

RMS Repeatability

The RMS repeatability is the ability of an instrument to reproduce the "RMS" of a surface, where with "RMS" commonly the *Sq* parameter is intended. The problem with this quantity is that a smaller value does not necessarily imply a "better" instrument; e.g., a high but constant measurement noise may result in a very low (good) RMS repeatability. A critical review of this and some more commonly used specification values (e.g., "vertical resolution") was given by de Groot (2017).

Conclusions and Further Requirements

At present, instrument manufacturers use a wide range of specifications. The general adoption of the ISO standards as mentioned, and specification of the surface texture metrological characteristics as listed above, would mean a major step forward to clarify and compare the various instruments' performances. From a user's perspective, calibration of the metrological characteristics may be the starting point of an uncertainty assessment of measurements taken by an instrument; however, for a complete understanding and uncertainty estimation for specific measurements, additional information may be needed (Haitjema 2015). Also, a sound and generally applicable topography fidelity assessment method is still to be developed.

Cross-References

- ▶ Calibration
- ▶ Flatness
- ▶ Roughness
- ▶ Surface Parameter
- ▶ Surface Texture
- ▶ Topography
- ▶ Traceability

References

Brand U, Schädelnach H, Schödel R, Feist C, Hinzmann G (2006) New depth-setting standards with grooves up to 5 mm depth. In: Proceedings of 6th international EUSPEN conference, Baden, 2006, p 438

Creath K, Wyant JC (1990) Absolute measurement of surface roughness. Appl Opt 29(26):3823–3826

de Groot PJ (2017) The meaning and measure of vertical resolution in optical surface topography measurement. Appl Sci 7(1)., art. no. 54

de Groot P, Beverage J (2015) Calibration of the amplification coefficient in interference microscopy by means of a wavelength standard. Proc SPIE 9526:10–11

de Groot P, Colonna de Lega X, Sykorda DM, Deck L (2012) The meaning and measure of lateral resolution for surface profiling interferometers. Opt Photonics News 23(4):10–13

Evans CJ, Tyler Estler W (1996) Self-calibration: reversal, redundancy, error separation, and 'absolute testing'. Ann CIRP 45(2):617–634

Fujii A, Suzuki H, Yanagi K (2011) Development of measurement standards for verifying functional performance of surface texture measuring instruments. J Phys Conf Ser 311:012009

Giusca CL, Leach RK (2013) Calibration of the scales of areal surface topography measuring instruments: part 3 – resolution. Meas Sci Technol 24:105010

Giusca CL, Leach RK, Helery F (2012a) Calibration of the scales of areal surface topography measuring instruments: part 2 – amplification coefficient, linearity and squareness. Meas Sci Technol 23:065005

Giusca CL, Leach RK, Helery F, Gutauskas T, Nimishakavi L (2012b) Calibration of the scales of areal surface topography measuring instruments: part 1 – measurement noise and residual flatness. Meas Sci Technol 23:035008

Giusca CL, Claverley JD, Sun W, Leach RK, Helmli F, Chavigner MPJ (2014) Practical estimation of measurement noise and flatness deviation on focus variation microscopes. Ann CIRP 63(1):545–548

Haitjema H (2015) Uncertainty in measurement of surface topography. Surf Topography Metrol Prop 3(3)., art. no. 035004

Haitjema H, Morel MAA (2005a) Accurate roughness measurements by laser interferometer calibration, VFM-uncertainty calculations and noise reduction. Proc SPIE 58790I:1–7

Haitjema H, Morel MAA (2005b) Noise bias removal in profile measurements. Measurement 38(1):21–29

ISO 12179 (2000) Geometrical product specifications (GPS) – surface texture: profile method – calibration of contact (stylus) instruments. ISO, Geneva

ISO 25178-601 (2010) Geometrical product specifications (GPS) – surface texture: areal – part 601: nominal characteristics of contact (stylus) instruments. ISO, Geneva

ISO 25178-602 (2010) Geometrical product specifications (GPS) – surface texture: areal – part 602: nominal characteristics of non-contact (confocal chromatic probe) instruments. ISO, Geneva

ISO 25178-603 (2013) Geometrical product specifications (GPS) – surface texture: areal – part 603: nominal characteristics of non-contact (phase-shifting interferometric microscopy) instruments. ISO, Geneva

ISO 25178-604 (2013) Geometrical product specifications (GPS) – surface texture: areal – part 604: nominal characteristics of non-contact (coherence scanning interferometry) instruments. ISO, Geneva

ISO 25178-605 (2014) Geometrical product specifications (GPS) – surface texture: areal – part 605: nominal characteristics of non-contact (point autofocus probe) instruments. ISO, Geneva

ISO 25178-606 (2015) Geometrical product specifications (GPS) – surface texture: areal – part 606: nominal characteristics of non-contact (focus variation) instruments. ISO, Geneva

ISO-DIS 25178-600 (2018) Geometrical product specifications (GPS) – surface texture: areal – part 600: metrological characteristics for areal-topography measuring methods. ISO, Geneva

Koops R, Van Veghel M, Van De Nes A (2015) A dedicated calibration standard for nanoscale areal surface texture measurements. Microelectron Eng 141:250–255

Koops R, Van Veghel M, Van De Nes A (2016) A virtual lateral standard for AFM calibration. Microelectron Eng 153:29–36
Krüger-Sehm R, Bacucz P, Jung L, Wilhelms H (2007) Chirp calibration standards for surface measuring instruments. Tech Mess 74:572–576
Leach RK (ed) (2011) Optical measurement of surface topography. Springer, Berlin/Heidelberg
Leach RK, Giusca CL, Haitjema H, Evans C, Jiang X (2015) Calibration and verification of areal surface texture measuring instruments Ann. CIRP 64(2):797–813
Xu ZW, Fang FZ, Gao HF, Zhu YB, Wu W, Weckenmann A (2012) Nano fabrication of star structure for precision metrology developed by focused ion beam direct writing Ann. CIRP 61(1):511–514

Surface Topography Filtering

► Surface Texture Filtering

Sustainability

Andreas Moltesen[1], Michael Z. Hauschild[2], David Dornfeld[3] and Sami Kara[4]
[1]Danish Energy Agency, Copenhagen, Denmark
[2]Department of Management Engineering, Division of Quantitative Sustainability Assessment, Technical University of Denmark, Lyngby, Denmark
[3]University of California, Berkley, CA, USA
[4]School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW, Australia

Synonyms

Sustainable development

Introduction

Basically, "sustainable" means – as the word implies – that something can be sustained, i.e., continued at that level, in principle indefinitely. Sustainability has many synonyms, such as durability and viability. In the environmental debate, sustainability is often used interchangeably with the term "sustainable development" even though they are two different concepts. For example, Waas et al. (2011) mention that some scholars assert that the term "sustainable development" is primarily about development/economic growth, whereas sustainability gives priority to the environment. Reid (1995) and Martin (2008) argue that the difference is rather that sustainable development should be seen as the process or journey to achieving sustainability which then according to Graedel and Howard-Grenville is an absolute concept "...as are pregnant and unique, to use two common examples. A sustainable world is not one that is slightly more environmentally responsible than it was yesterday" (Graedel and Howard-Grenville 2005).

Definition

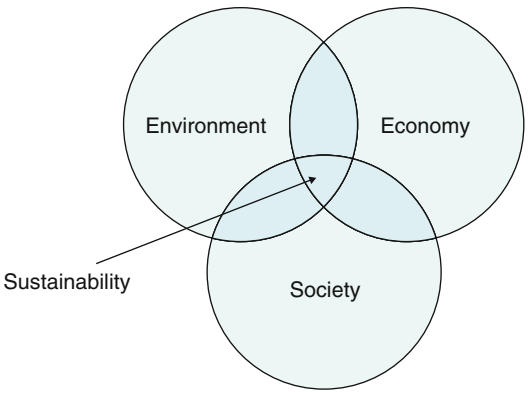
Most discussions of sustainability refer to the definition of a sustainable development that is given in the so-called Brundtland report, "Our common future" (WCED 1987), as:

> Development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Many other definitions are given in literature. For a review of definitions in literature, see, for example, Pezzey (1992). Despite the number and variation of definitions of sustainability, they all include considerations for the welfare of future generations. Welfare may be described in many different ways, for example, in terms of meeting of needs, well-being, or utility. Different emphasis in the definitions may also be put on the extent to which concerns for the present are considered. For example, in economic literature, sustainability has been defined as "non-decreasing utility over time" (Pezzey 1992), which does not indicate a concern for equity within the present generation. Less commonly, definitions also include concerns for other living species than humans (Waas et al. 2011). In general, however, sustainability is an anthropocentric concept.

Theory and Application

The sustainability term has been adopted and is influencing a very broad variety of scientific fields (Fig. 1). Examples are economy (e.g., Costanza et al. 1997), ecology (Gunderson and Holling 2002), and urban planning (Haughton 1999). Sustainability has also been adopted in management literature, where Elkington (1997) with the concept of "triple bottom line" has attempted to make sustainability manageable in a business context. Elkington argues that sustainability calls for not only considering the economic aspects of a production but also its economic and social aspect – in total a triple bottom line.



Sustainability, Fig. 1 Three dimensions of sustainability

In 2015, the member states of the United Nations adopted 17 Sustainable Development Goals (SDGs) to "end poverty, protect the planet, and ensure prosperity for all as part of a new sustainable development agenda" (UN 2015). Each goal is accompanied by specific targets that member states must strive to achieve by 2030. The SDGs specify in more detail what is encompassed by a sustainable development according to the United Nations and as such offer more guidance than the three sustainability dimensions in Fig. 1 but also emphasize the importance of a holistic development approach since there are strong synergies between many of them. Figure 2 shows the 17 SDGs.

The existing ambiguity in the interpretation of sustainability into practical use may have been an enabling factor in its broad adoption in the many varied contexts and applications in which it can be found. Although the debate continues, most authors recognize that sustainability involves integrating economic, environmental, and social factors in such a way that the human experience is able to continue undiminished over time (Harding 1996), referred to by the International Institute for Sustainable Development as "environmental, economic and social well-being for today and tomorrow" (IISD 2010).

The different interpretations of sustainability may be classified into "hard" sustainability and

Sustainability, Fig. 2 United Nations' 17 Sustainable Development Goals (UN 2015)

"soft" sustainability in accordance with the degree to which they allow trade-offs between the involved dimensions. Soft sustainability thus allows an increase in the economic or social sustainability to offset a decrease in environmental sustainability, while a hard sustainability interpretation does not (Neumayer 2003).

The assessment of the sustainability of products, production processes, and technologies suffers from the rather loose and ambiguous state of the definition of the concept. There is, however, agreement that an assessment must encompass all dimensions of sustainability and needs to take a life cycle perspective on the activity or process to ensure that shifting of problems in the value chain is taken into account when optimizing the sustainability.

Cross-References

- ▶ Sustainability of Machining
- ▶ Sustainable Manufacturing

References

Costanza R, d'Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'Neill RV, Paruelo J, Raskin RG, Sutton P, van den Belt M (1997) The value of the world's ecosystem services and natural capital. Nature 387:253–260

Elkington J (1997) Cannibals with forks: the triple bottom line of 21st century business. Capstone, Oxford

Graedel T, Howard-Grenville J (2005) Greening the industrial facility: perspectives, approaches and tools. Springer, Heidelberg/Berling, p 126

Gunderson LH, Holling CS (2002) Panarchy – understanding transformations in human and natural systems. Island Press, Washington, DC

Harding R (1996) Sustainability: principles to practice: outcomes. In: 1994 Fenner conference on the environment, proceedings, 13–16 November 1994. Department of the Environment, Sport and Territories, Canberra, pp 13–18

Haughton G (1999) Environmental justice and the sustainable city. J Plan Educ Res 18(3):233–243

IISD (2010) What is sustainable development? [online]. International Institute for Sustainable Development (IISD). Available: http://www.iisd.org/sd/. Accessed 8 May 2013

Martin S (2008) Sustainability, systems thinking and professional practice. J Educ Sustain Dev 2(1):31–40

Neumayer E (2003) Weak versus strong sustainability: exploring the limits of two opposing paradigms, 2nd edn. Edward Elgar, Cheltenham

Pezzey J (1992) Sustainable development concepts; an economic analysis. World Bank environment paper number 2. Report no 11425. World Bank, Washington, DC

Reid D (1995) Sustainable development. An introductory guide. Earthscan, London

Waas T, Hugé J, Verbruggen A, Wright T (2011) Sustainable development: a bird's eye view. Sustainability 3(19):1637–1661

UN (2015) UN sustainable development goals. http://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 11 July 2017

WCED (World Commission on Environment and Development) (1987) Our common future. Oxford University Press, Oxford

Sustainability of Machining

David Dornfeld
University of California, Berkley, CA, USA

Synonyms

Sustainable cutting; Sustainable material removal

Definition

Sustainable machining should be defined in the context of sustainability and sustainable manufacturing. The entry on "sustainability" states that sustainability implies that something can be sustained, i.e., continued at that level, in principle indefinitely. "Sustainable manufacturing" is defined as the creation of manufactured products that use processes that are nonpolluting, conserve energy and natural resources, and are economically sound and safe for employees, communities, and consumers, including the manufacturing of sustainable products and the sustainable manufacturing of all products. Thus, applied to machining, sustainable machining can be defined as the manufacture of products (components, etc.) by a subtractive process based on cutting (material removal by the cutting action of a tool usually with a machine tool to create surfaces and features) in a way that is nonpolluting, minimizes and conserves energy and natural resources, and is economically sound and safe for employees, communities, and consumers. And, finally, the process can be continued at that level

indefinitely. Further, there must be consideration of the social and economic elements as well as environmental for sustainable machining.

Theory and Application

The challenge in the application of any concept of sustainability to a process is great. Processes are subset of large systems of production that range from the enterprise level of supply chains and distributions to facilities with lines of machinery, to individual machines, to the tooling, and finally to the actual removal process (in the case of machining). The discussion here is focused closer to the production floor with specific interest in one of the primary pieces of manufacturing hardware – the machine tool.

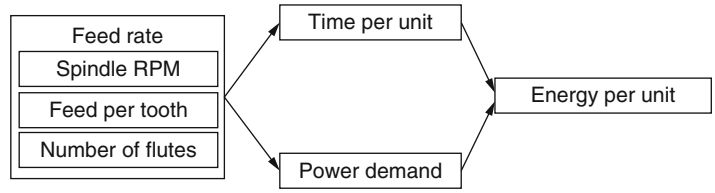
Recent research includes power consumption analyses of machine tool use. Dahmus and Gutowski (2004) conducted an environmental analysis of machining that quantified the energy consumption of four types of milling machines varying in automation as well as accounted for material production and cutting fluid preparation. Taniguchi et al. (2006) studied the effects of downsizing a CNC milling machine tool on its energy and resource requirements. Ultimately, a life-cycle energy assessment is required to determine the appropriate strategy to reduce the impact of a machining process. This type of analysis yields two general possibilities: (1) high constant energy demand due to the dominance of noncutting operations and peripheral equipment or (2) low constant energy demand due to the dominance of cutting operations. Sample strategies to address the first case include using machine design to minimize the energy requirements of peripheral equipment (e.g., kinetic energy recovery systems used in conjunction with the spindle) and focusing on machine operation to increase the production rate of the machine tool. Strategies to address the second case generally require optimization of the cutting process itself. This may be difficult to accomplish from a design perspective due to the influence of desired process parameters, but energy savings may be achieved by considering typical machine tool use in design (e.g., ensure that axes with high motion carry less weight).

While the current literature provides an extensive knowledge of the life-cycle energy consumption of machining, it is limited by the assumption that machine tool operation dominates the overall impact such that other aspects of the machine tool's life cycle, such as its manufacture, are neglected. Furthermore, much of the literature neglects transportation, material inputs (e.g., cutting fluid), or facility inputs (e.g., HVAC and lighting), which may all have a significant impact on the overall energy consumption.

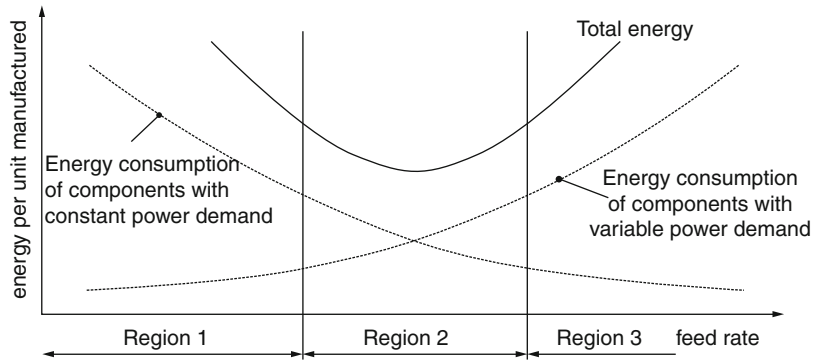
It was the goal of Diaz et al. (2010) to study the effect of these aspects as well as that of the manufacturing environment and degree of automation on the life-cycle energy requirements of milling machine tools. This analysis of two milling machines placed in three environments quantified the CO_2-equivalent emissions associated with producing a standardized part over the lifetime of a machine tool. Several findings show significant differences from previously published literature, such as the manufacturing phase impact, for example, HVAC, and lighting requirements are a significant portion of the use phase emissions. In addition to the significant impacts of parameters that have been disregarded in the literature, the results suggest areas for reducing energy consumption. For example, a more energy-efficient thermal control system could reduce the overall energy usage since HVAC is energy intensive. This analysis may also provide greater clarity to other LCA studies by highlighting potential impacts on a product's manufacturing phase since machine tools are key to manufacturing all other products. Finally, these results may be extended to other manufacturing processes by showing new areas to consider for energy and environmental impact reductions.

Given the significant environmental impact of the use phase of a machine tool, an analysis was conducted by Diaz et al. (2009) on end milling process parameter selection for energy consumption reduction. The energy per unit manufactured is determined by both the power demand of the machine tool during machining and the processing time, Fig. 1.

The power demand of a machine tool may be divided into a constant and a variable component (Diaz et al. 2009). The constant power can be attributed to the computer, fans, lighting, etc., of the machine tool. This component of the total power demand is independent of process parameter



Sustainability of Machining,
Fig. 1 Influence of process parameters on the energy per unit manufactured



Sustainability of Machining,
Fig. 2 Regions of machining process

selection. The variable power demand, though, is dependent on process parameter selection and can be attributed to the spindle or the drives of the table axes. The processing time per unit manufactured is determined by the feed rate. Cutting conditions at a specific feed rate depend on the selection of the number of revolutions per minute of the spindle, the feed per tooth, and the number of flutes.

Figure 2 identifies two opposing effects on the energy per unit manufactured. First, as the feed rate increases, the processing time is reduced. Therefore, the contribution of the constant power demand of the machine tool to the energy per unit manufactured decreases. Second, an increase in the feed rate demands more power from the machine tool with or without adjustment of the cutting speed. Depending on which effect prevails, three different machining regions may be found.

The sum of both energy contributions results in a parabolic-total energy plot also shown in Fig. 2. In Region 1 the decrease due to a shorter processing time dominates the increased variable power demand. In this region, the feed rate will be chosen as fast as technically possible. In Region 2 the energy per unit manufactured is fairly constant, whereas the increase of the variable power demand dominates in Region 3. If the process is located in Region 3, slower feed rates would lead to lower energy per unit manufactured.

The tool limits the possibilities to reduce the energy per unit manufactured by process parameter selection because elevated tool wear was observed at process parameter combinations, which reduce the energy demand of the machine tool during the manufacturing process. Coated end mills enable high-speed cutting, whereas the uncoated end mills can be used for conventional cutting only. High-speed machining is carried out at drastically elevated cutting speeds and reduced feeds per tooth compared to conventional cutting. In general, higher cutting speeds increase the power demand of the machine tool, whereas a lower feed per tooth reduces the power demand. The experimental studies show that high-speed cutting results in smaller energies per unit manufactured compared to a machining operation at conventional cutting speeds. The decrease in the processing time has a greater impact on the energy demand per unit manufactured than the increase in power demand. Therefore, it can be concluded that high-speed cutting is more energy efficient than cutting at conventional speeds. Future work should include the variation of coating as well as the consideration of the additional production energy of coating.

From an economical and environmental point of view, it is desirable to minimize the coolant use in machining. In the metal cutting industry, coolants play a crucial role in the production of high-quality products and are often used in large quantities. For many machining processes, the use of coolant lubricants is not imperatively necessary. However, simply turning off the coolant supply can deteriorate the cutting result, since the functions of the coolant are not fulfilled anymore. In detail, the primary functions of coolant lubricants comprise lubrication, cooling, and cleaning. Without these functions, more friction and adhesion between the tool and the workpiece occur. Moreover, the heat produced in the cutting process has to be discharged only by the chips, the tool, and the workpiece and not by the lubricant anymore. The result is a higher thermal load on the tool, the workpiece, and the machine tool, causing a shorter tool life and a reduced workpiece and machine tool accuracy. Therefore, the absence or minimization of coolant lubricant as with dry cutting and minimum quantity lubrication (MQL) necessitates an analysis of the boundary conditions and the work dependencies between the process, the cutting tool, the workpiece, and the machine tool. Finally, storage and disposal of these fluids add cost and can require special processing for recycling or disposal.

Cross-References

- ▶ Sustainability
- ▶ Sustainable Manufacturing

References

Dahmus JB, Gutowski TG (2004) An environmental analysis of machining. In: Proceedings of the 2004 ASME international mechanical engineering congress and R&D exposition (IMECE2004), Anaheim, 13–19 Nov 2004, pp 643–652 [Paper no. IMECE2004–62600]

Diaz N, Helu M, Jarvis A, Toenissen S, Dornfeld D, Schlosser R (2009) Strategies for minimum energy operation for precision machining. In: Yamazaki K (ed) Machine tool technologies research foundation (MTTRF) Shanghai 2011. Proceedings of machine tool technologies research foundation and ASCENTi-CNC 2009, annual meeting, Shanghai, 8–9 July, pp 47–52

Diaz N, Helu M, Jayanathan S, Chen Y, Horvath A, Dornfeld D (2010) Environmental analysis of milling machine tool use in various manufacturing environments. In: 2010 I.-E. international symposium on sustainable systems and technology, Arlington, 17–19 May 2010, pp 1–6

Taniguchi M, Kakinuma Y, Aoyama T, Inasaki I (2006) Influences of downsized design for machine tools on the environmental impact. In: Proceedings of the MTTRF 2006 annual meeting, San Francisco, pp 1–4

Sustainable Cutting

- ▶ Sustainability of Machining

Sustainable Development

- ▶ Sustainability

Sustainable Manufacturing

Michael Z. Hauschild[1], David Dornfeld[2], Margot Hutchins[3], Sami Kara[4] and Francesco Jovane[5]
[1]Department of Management Engineering, Division of Quantitative Sustainability Assessment, Technical University of Denmark, Lyngby, Denmark
[2]University of California, Berkley, CA, USA
[3]Sandia National Laboratories, Livermore, CA, USA
[4]School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW, Australia
[5]Dipartimento di Meccanica, Politecnico di Milano, Milan, Italy

Synonyms

Clean production; Corporate social responsibility; Green manufacturing

Definition

Sustainable manufacturing has been defined by the International Trade Administration under

the US Department of Commerce as *"...the creation of manufactured products that use processes that minimize negative environmental impacts, conserve energy and natural resources, are safe for employees, communities, and consumers and are economically sound"* (USDOC 2012). The US National Council for Advanced Manufacturing (NACFAM) contrasts this definition with the UN definition of a sustainable development given in the Brundtland report as a *"Development that meets the needs of the present without compromising the ability of future generations to meet their own needs"* and extends the definition to address both the manufacturing of "sustainable" products and the sustainable manufacturing of all products, taking into account the full sustainability life cycle issues related to the products manufactured (NACFAM 2012).

The double focus on both manufacturing activities and the life cycle of the products that are being manufactured is also found in the *sustainable production* element in the concept of sustainable production and consumption that is being promoted by international institutions like the United Nations (UNEP 2013) and the European Union (Council of the European Union 2008) as a central element in strategies for the development toward a sustainable society.

Theory and Application

The absolute form used in the term *sustainable manufacturing* is potentially misleading unless it is clearly defined. What is truly sustainable, i.e., an activity that we will be able to sustain indefinitely, depends very much on the context in which the activity is performed, in terms of, e.g.:

- The total population of the earth whose needs must be met today as well as in the future
- The equity among the people of the earth and the material level at which they wish to fulfill their needs
- The available resource base for the resources that are critical in terms of availability (today as well as in the future with the changes in demand and availability that can be foreseen)
- The state of the environment in terms of distances to critical pollution levels affected by the manufacturing activity (today as well as in the future)

Classification of a manufacturing activity as sustainable in absolute terms hence requires a number of strong assumptions in order to be substantiated and meaningful.

For practical purposes, it is often more relevant to talk about the sustainability of manufacturing in relative terms, i.e., as one form of manufacturing being more sustainable (or green) than another, both in terms of the manufacturing practice and in terms of the products that it manufactures. The relative environmental sustainability of a product is also represented by its eco-efficiency, the product with the highest eco-efficiency being the most environmentally sustainable among the compared products. It may, however, not actually be sustainable in the absolute sense.

Conflicts Between Relative and Absolute Sustainability of Manufacturing

While relative sustainability of products or their manufacturing is more operational in many situations, there is a potential conflict between increasing the relative environmental sustainability and moving toward absolute environmental sustainability (Hauschild 2015). The conflict may arise when the consumption aspect is ignored in the sustainability assessment, and there is a positive feedback coupling between an increase in the relative environmental sustainability or eco-efficiency and the consumption. In these cases, what seems to be sustainable from a manufacturing perspective turns out not to be so when assessed at the societal level.

An example is offered by the case of automotive person transport in Europe where the increase in fuel efficiency over the last decade is an accomplishment of the automotive manufacturing that has clearly made passenger cars more environmentally sustainable. It has, however, also made personal mobility via private cars more affordable and hence led to an increase of the total transport work which has more than neutralized the improvements in fuel efficiency. The net result is

an increase in the environmental impacts from person transport and hence a development away from the environmental sustainability of this sector, sometimes referred to as the "rebound effect" (Clark 2007).

Another example is offered by the example of indoor lighting for which the energy efficiency (which here serves as a proxy for environmental sustainability) has increased by more than three orders of magnitude over the last centuries going from candles via incandescent bulbs to diode-based lighting. Also here, the improvements in energy efficiency have been accompanied by an increase in economic efficiency, leading to increases in the use of lighting that have neutralized the energy efficiency gains, again showing that an increase in relative environmental sustainability does not alone lead us in the direction of a sustainable society (Gutowski 2011). Along the same vein, LCD screens offer a more eco-efficient alternative to the old CRT technology, providing a similar functionality (screen size, image quality) at lower environmental impact. However, the enabling LCD technology also supports the construction of much larger screens, so the improvement in eco-efficiency is countered by an increase in the size of TV and PC screens (Kim et al. 2014).

Therefore, the manufacturing industry must consider not only the environmental impact of products during the product design now but also future technology change and the anticipated market growth in order to qualify claims of sustainability (Kim and Kara 2012; Kim et al. 2014).

The relationship between eco-efficiency, consumption, and environmental impact is reflected by the I = PAT equation that expresses the environmental impact I as a product of the size of the human population P, the material standard of living or affluence A (the value that is created or consumed per capita), and the technology factor T, often expressed as the reciprocal of the eco-efficiency (impact per created value). Considering the growth in population (P) and the economic development and increase in material standard of living (A) and resource use in many parts of the world and considering the fact that the current level of impact (I) is not sustainable for many types of environmental impact, there is a need to increase the eco-efficiency by which we provide technological services by a factor of 10–20 over the next 40 years in order to ensure a sustainable production (von Weizsäcker et al. 1998; Schmidt-Bleek 2008; Hauschild et al. 2005). This macro level need trickles down to the meso level of manufacturing as a daunting challenge. But as the examples given above illustrate, even such high increases in eco-efficiency may not lead to sustainable production when A and T are not mutually independent, so a reduction in T (an increase in eco-efficiency) triggers a growth in A that more than outbalances it resulting in an overall increase in environmental impact I (Pogutz and Micale 2011).

All these examples also stress the need to consider the whole life cycle of the technology or product to address the use versus manufacturing stage trade-off. Things that do not move or need power to operate like bridges, furniture, etc. are dominantly manufacturing-stage consumers of resources and, by extension, impact. Things that do move and need power to operate like automobiles, airplanes, etc. are generally use-stage heavy. If the use stage dominates, efforts to reduce consumption (meaning giving the consumer products that deliver the required functionality or service but at a lower environmental impact or energy/resource consumption) are appropriate. If the manufacturing stage dominates, the product's life cycle impact, say for a structure with low energy consumption in use, then manufacturing stage impacts, and resource consumption must be addressed in order to increase eco-efficiency.

Addressing the Three Dimensions of Sustainability

Sustainability is generally considered as covering three dimensions addressing the impacts on the environment, the impacts on society and central stakeholders, and the impacts on the economy of the manufacturer, as represented in the concept of the triple bottom line.

In its "10 principles of sustainable production," the Lowell Center for Sustainable Production addresses these three dimensions in 10 concrete focus points (Lowell Center 2012):

1. Products and packaging are designed to be safe and ecologically sound throughout their life cycle.
2. Services are organized to satisfy real human needs and promote equity and fairness.
3. Wastes and ecologically incompatible by-products are reduced, eliminated, or recycled.
4. Chemical substances or physical agents and conditions that present hazards to human health or the environment are eliminated.
5. Energy and materials are conserved, and the forms of energy and materials used are most appropriate for the desired ends.
6. Work places and technologies are designed to minimize or eliminate chemical, ergonomic, and physical hazards.
7. Work is organized to conserve and enhance the efficiency and creativity of employees.
8. The security and well-being of all employees is a priority, as is the continuous development of their talents and capacities.
9. The communities around workplaces are respected and enhanced economically, socially, culturally, and physically.
10. The long-term economic viability of the enterprise or institution is enhanced.

These principles need to be operationalized to be effective in design and manufacturing. Metrics with specific definitions are needed to ensure these principles are upheld. A thorough understanding of material, energy, and water flows as well as any associated hazard or risk to humans and ecosystems is required to address the environmental dimension of sustainability. Human and labor rights, as advocated by international bodies such as the UN and International Labour Organization (ILO), provide some insight into the more basic issues of the social dimension of sustainability. However, as the Lowell Center's principles imply, metrics are needed to assess the well-being of a range of stakeholders, including employees, communities, consumers, and others impacted by product life cycles. Finally, measures to assess long-term economic viability and the broader economic contribution of an organization are needed to address the economic dimension of sustainability.

Application

Applications of sustainable manufacturing can be found across various scales, from the global/supply chain level down to the process level of manufacturing (see Duflou et al. 2012 for a review of how energy and resource efficiency is addressed at the different scales). At the global level, techniques can be utilized to optimize the location of suppliers in order to reduce the energy and, consequently, the greenhouse gas emissions attributed to transportation. Additionally, a factory manager can opt to build the factory close to the consumer market as a means of lowering the environmental impact of the distribution of products to consumers. Facility design is another viable avenue for achieving a state of sustainable manufacturing as the installation of energy-efficient HVAC equipment and lighting would effectively lower the energy consumption of a facility, as would the implementation of machine tool prioritization techniques during the factory planning and operational phases (Diaz and Dornfeld 2012; Herrmann et al. 2011). A product's sustainability can be optimized in a life cycle perspective using Design for Environment tools (Hauschild et al. 2004) and also taking into account the social impacts along the life cycle (Jørgensen et al. 2008; Hauschild et al. 2008). At the machine tool level, the "tare" energy consumption can characterize the electrical energy of a piece of production equipment, so machine tools with lower tare energy would consequently consume less energy as well (Vijayaraghavan and Dornfeld 2010). Lastly, at the process level, we can consider tool path planning (Kong et al. 2011), tooling (Diaz et al. 2010, 2011), and the use of alternative cooling techniques (Klocke and Eisenblätter 1997; Li et al. 2012) to reduce the environmental impact of processing raw material. This highlights a few examples of methods that can be utilized to achieve the state of sustainable manufacturing, and more detailed examples can be found in Dornfeld (2012).

Competitive Sustainable Manufacturing

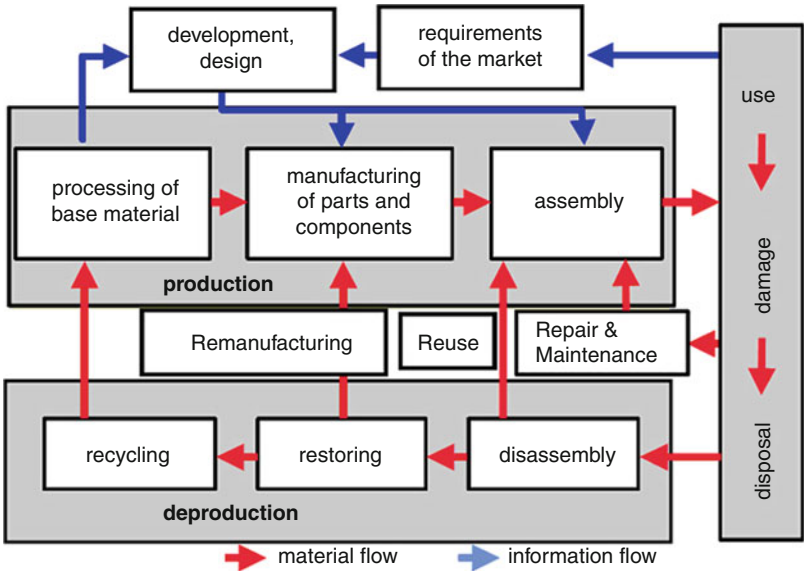
As proposed by CIRP, a holistic overarching multilevel dynamic paradigm addressing a sustainable development is Competitive Sustainable

Manufacturing (CSM) (Jovane et al. 2008), where manufacturing has to cover the entire product/service life cycle (Yoshikawa 2008), including enabling processes and business models (Fig. 1).

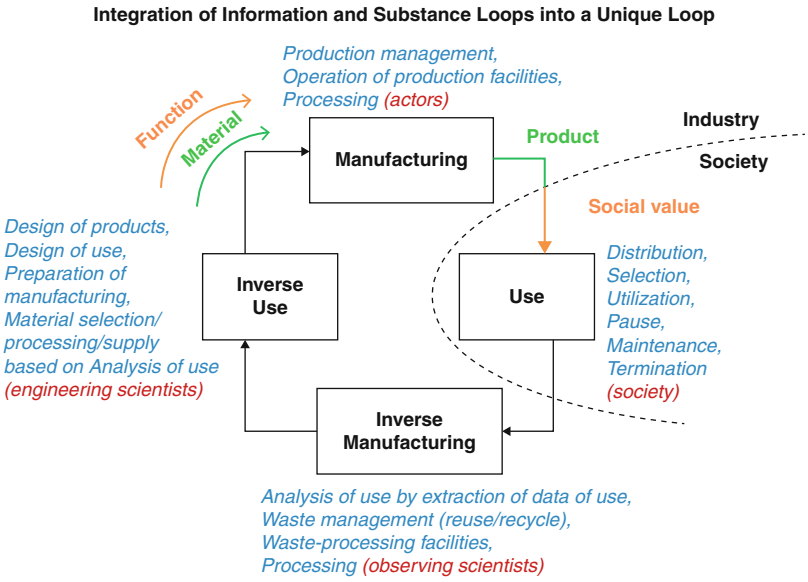
The demand paradigm, challenge-based paradigm, and the response paradigm, supported by the research-innovation-market value chain, are depicted in Fig. 2.

CSM Fundamentals

- Competitiveness (defined as "market" success) and sustainability represent mandatory conditions that must be met by products and services, by processes and business models, and by enterprises, for CSM to respond to the challenges (Jovane et al. 2008).
- At the *macro* (country) level, competitiveness was introduced by Porter. It is based on the



Sustainable Manufacturing, Fig. 1 Cycles of material and information flow. (Westkaemper et al. 2000)



Sustainable Manufacturing, Fig. 2 Manufacturing actions for sustainable evolution. (Yoshikawa 2008)

productivity with which a given country produces goods and services.
- At the *meso* level, competitiveness may be seen as the ability of a supply paradigm to respond, better than another one, to a demand paradigm.
- At the *field* level, it may be defined as a comparative concept, i.e., the ability of an actor (firms, universities, institutes, research centers, etc.) to respond to "customer demand" better than anyone else.

At the *macro* level, manufacturing sustainability is achieved when facing:

- Economic challenges by producing wealth and new services ensuring development and competitiveness through time.
- Environmental challenges by promoting minimal use of natural resources (in particular, nonrenewable) and managing them in the best possible way while reducing the environmental impact to sustainable levels.
- Social challenges by promoting social development and improved quality of life.

At the *meso* level, sustainability of manufacturing requires the development of appropriate response paradigms and enabling technologies concerning high added-value products and services, processes, and business models that meet the aforementioned economic, social, and environmental conditions.

At the *field* level, the implementation of manufacturing sustainability relies on sustainable companies that – in cooperation with other stakeholders – are able to generate and produce, by adopting competitive and sustainable processes and business models, new, high added-value products and services, responding to grand challenges (Jovane et al. 2008).

Several advanced and emerging countries are acting proactively by conceiving and launching research and innovation programs. For instance, the European Union is promoting and supporting initiatives toward CSM as proposed by the Manufacture Technology Platform. International institutions, such as the OECD and UNEP, are strategically addressing CSM.

Recently, researchers from Harvard Business School and London School of Economics demonstrated that in a matched comparison of 180 companies, the 90 companies that voluntarily adopted environmental and social policies long ago significantly outperform the 90 similar companies that do not consider sustainability an important part of their business and, hence, have adopted none of these policies. The outperformance is both in terms of performance in annual accounts and on the stock market. On the latter, return on investment was up to 50% higher over the full period 1992–2010 (Eccles et al. 2012).

Cross-References

- ▶ Cleaner Production
- ▶ Eco-Efficiency
- ▶ Energy-Efficient Manufacturing
- ▶ Life Cycle Assessment
- ▶ Life Cycle Engineering
- ▶ Recycling
- ▶ Remanufacturing
- ▶ Reuse
- ▶ Sustainability

References

Clark G (2007) Evolution of the global sustainable consumption and production policy and the United Nations environment Programme's (UNEP) supporting activities. J Clean Prod 15(6):492–498

Council of the European Union (2008) "Sustainable consumption and production and sustainable industrial policy action plan: council conclusions," 16914/08, 5 Dec 2008. http://register.consilium.europa.eu/pdf/en/08/st16/st16914.en08.pdf. Accessed 24 June 2013

Diaz N, Dornfeld D (2012) Cost and energy consumption optimization of product manufacture in a flexible manufacturing system. In: Proceedings of the 19th CIRP international conference on life cycle engineering, Berkeley, pp 411–416

Diaz N, Choi S, Helu M, Chen Y, Jayanathan S, Yasui Y, Kong D, Pavanaskar S, Dornfeld D (2010) Machine tool design and operation strategies for green

manufacturing. In: Proceedings of the 4th CIRP international conference on high performance cutting, vol 1, Gifu, pp 271–276
Diaz N, Redelsheimer E, Dornfeld D (2011) Energy consumption characterization and reduction strategies for milling machine tool use. In: Proceedings of the 18th CIRP international conference on life cycle engineering, Braunschweig, pp 263–267
Dornfeld DA (ed) (2012) Green manufacturing: fundamentals and applications (Green energy and technology). Springer, New York
Duflou JR, Sutherland JW, Dornfeld D, Herrmann C, Jeswiet J, Kara S, Hauschild M, Kellens K (2012) Towards energy and resource efficient manufacturing: a processes and systems approach. CIRP Ann Manuf Technol 61(2):587–609
Eccles RG, Ioannou I, Serafeim G (2012) The impact of a corporate culture of sustainability on corporate behavior and performance. NBER working paper series. National Bureau of Economic Research, Cambridge, MA, Issue 17950
Gutowski T (2011) Manufacturing and the science of sustainability. In: Hesselbach J, Herrmann C (eds) Glocalized solutions for sustainability in manufacturing – proceedings of the 18th CIRP international conference on life cycle engineering, Technische Universität Braunschweig, Braunschweig, Germany, May 2nd–4th, 2011. Springer, Berlin/Heidelberg, pp 32–39
Hauschild MZ (2015) Better – but is it good enough? On the need to consider both eco-efficiency and eco-effectiveness to gauge industrial sustainability. Procedia CIRP 29:1–7. https://doi.org/10.1016/j.procir.2015.02.126. (ISSN: 2212-8271)
Hauschild M, Jeswiet J, Alting L (2004) Design for environment – do we get the focus right? CIRP Ann Manuf Technol 53(1):1–4
Hauschild M, Jeswiet J, Alting L (2005) From life cycle assessment to sustainable production: status and perspectives. CIRP Ann Manuf Technol 54(2):1–21
Hauschild MZ, Dreyer LC, Jørgensen A (2008) Assessing social impacts in a life cycle perspective – lessons learned. CIRP Ann Manuf Technol 57(1):21–24
Herrmann C, Thiede S, Kara S, Hesselbach J (2011) Energy oriented simulation of manufacturing systems: concept and application. CIRP Ann Manuf Technol 60(1):45–48
Jørgensen A, Le Bocq A, Nazarkina I, Hauschild M (2008) Methodologies for social life cycle assessment – a review. Int J LCA 13(2):96–103
Jovane F, Yoshikawa H, Alting L, Boër CR, Westkaemper E, Williams D, Tseng M, Seliger G, Paci AM (2008) The incoming global technological and industrial revolution towards competitive sustainable manufacturing. CIRP Ann Manuf Technol 57(2):641–659. https://doi.org/10.1016/j.cirp.2008.09.010
Kim SJ, Kara S (2012) Impact of technology on product life cycle design: an environmental and functional perspective. In: Proceedings of the 19th CIRP international conference on life cycle engineering, Berkeley , 23–25 May, pp 191–192
Kim SJ, Kara S, Kayis B (2014) Analysis of the impact of technology changes on the economic and environmental influence of product life cycle design. Accepted for publication in the Int J Comput Integr Manuf 27(5):422–433
Klocke F, Eisenblätter G (1997) Dry cutting. CIRP Ann Manuf Technol 46(2):519–526
Kong D, Choi S, Yasui Y, Pavanaskar S, Dornfeld D, Wright P (2011) Software-based tool path evaluation for environmental sustainability. J Manuf System 30(4):241–247
LCSP (2012) What is sustainable production? Lowell Center for Sustainable Production (LCSP). University of Massachusetts Lowell, Lowell. http://sustainableproduction.org/abou.what.php. Accessed Dec 2012
Li W, Winter M, Kara S, Herrmann C (2012) Eco-efficiency of manufacturing processes: a grinding case. CIRP Ann Manuf Technol 61(1):59–62
NACFAM (2012) Sustainable manufacturing. National Council for Advanced Manufacturing (NACFAM). http://www.nacfam.org/PolicyInitiatives/SustainableManufacturing/tabid/64/Default.aspx. Accessed 20 Dec 2012
Pogutz S, Micale V (2011) Sustainable consumption and production. An effort to reconcile the determinants of environmental impact. Soc Econ 33:29–50
Schmidt-Bleek F (2008) Factor 10: the future of stuff. Sustainability 4(1):1–4
United Nations Environment Programme (UNEP) (2013) "Sustainable production and consumption" website: http://www.rona.unep.org/about_unep_rona/scp/index.html. Accessed 24 June 2013
USDOC (2012) How does commerce define sustainable manufacturing? US Department of Commerce (USDOC), International Trade Administration. http://www.trade.gov/competitiveness/sustainablemanufacturing/how_doc_defines_SM.asp. Accessed 20 Dec 2012
Vijayaraghavan A, Dornfeld D (2010) Automated energy monitoring of machine tools. CIRP Ann Manuf Technol 59(1):21–24
von Weizsäcker E, Lovins AB, Lovins LH (1998) Factor four: doubling wealth, halving resource use – a report to the club of Rome. Earthscan, London
Westkaemper E, Alting L, Arndt G (2000) Life cycle management and assessment: approaches and visions towards sustainable manufacturing. CIRP Ann Manuf Technol 49(2):501–526
Yoshikawa H (2008) Sustainable manufacturing. In: Proceedings of the 41st CIRP conference on manufacturing systems, Tokyo, 26–28 May 2008

Sustainable Material Removal

▶ Sustainability of Machining

Synthesis

Juan Manuel Jauregui-Becker
Laboratory of Design, Production and Management, University of Twente, Enschede, The Netherlands

Synonyms

Combination; Creation; Integration

Definition

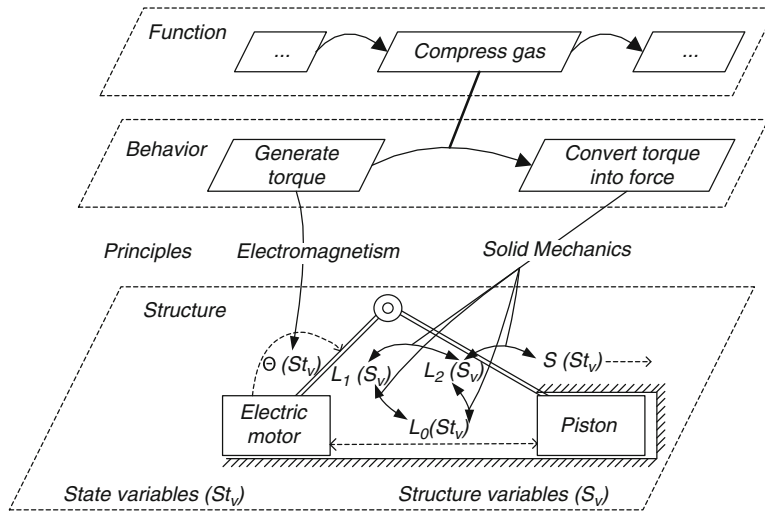
Generally speaking, synthesis can be defined as the composition or combination of parts – building blocks, elements – to create a new whole product, artifact, technology, machine, or graphic, whose ruling behavior emerges from the interaction of its constituent parts. In the context of design, synthesis follows different definitions (Chakrabarti 2002). The two most relevant to this work are "synthesis as designing" and "synthesis as solution generation." In the first definition, synthesis is defined as an iterative process of solution generation and solution evaluation. The second definition narrows its scope to that of generation solutions.

Theory and Application

A synthesis process comprehends a complex combination of cognitive and mathematical mechanism (e.g., random generation, backward reasoning, abduction, case-based reasoning, and constraint-solving). Although no unified theory exists for explaining the nature of the synthesis process, the generally accepted Function Behavior State/Structure (FBS) family of frameworks allows for making some concrete statements on the nature of this process.

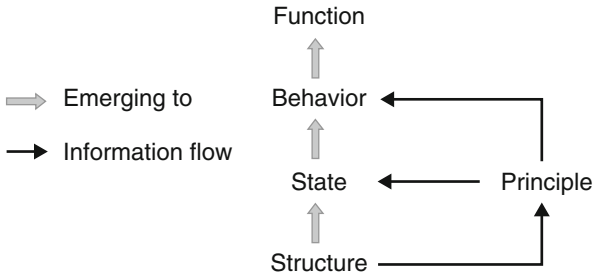
The FBS Model

FBS models a design artifact by distinguishing the following levels of object representation: function, behavior/state, and structure, as shown in Fig. 1. The basis of the FBS model is that the transition from function to structure is performed via the synthesis of physical behaviors. Therefore, behaviors allow characterizing the implementation of a function. As many different views of the FBS model have been developed and researched, the Function Behaviour Principle State Structure (FBPSS) model presented by Zhang et al. (2011) serves as a unifying framework for the different FBS schools of thought. This model is based on the analysis and generalization of the Japanese (Umeda and Tomiyama 1995, 1997), European (Pahl et al. 2007),



Synthesis, Fig. 1 FBS of a crank compression mechanism (Becker 2010)

Synthesis,
Fig. 2 Relation between function–behavior–principle–state–structure (Based on Zhang et al. (2005))



American (Chandrasekaran et al. 1993), and Australian (Gero and Kannengiesser 2004) schools of design theory and methodology.

The FBPSS model uses the following definitions:

- Structure: Is a set of entities and relations among entities connected in a meaningful way. Entities are perceived in the form of their attributes when the system is in operation. For example, in Fig. 1 the structure is represented by an electric motor and a crank mechanism. Here, the two possible entities (structures) are the lengths of the bars *L1* and *L2*.
- States: Are quantities (numerical or categorical) of the behavioral domain (e.g., heat transfer, fluid dynamics, psychology). States change with respect to time, implying the dynamics of the system. For example, in Fig. 1, the states of the structure are represented by the distance *L0* between the electric motor and the piston, the torque *T* of the electric motor, or the displacement of the pistons.
- Principle: Is the fundamental law that allows the development of a quantitative relation of the state variables. It governs behavior as the relationships among a set of state variables. For example, in Fig. 1, two possible principles are electromagnetism ruling the operation of the electric motor and solid mechanics ruling the function of the crank mechanism.
- Behavior: Represents the response of the structure when it receives stimuli. Since the structure is represented by state and structure variables, behaviors are quantified by the values of these variables. In the case presented in Fig. 1, the two behaviors are *Generate torque* and *Convert torque into force*.
- Function: It is about the usefulness of a system. For example, in Fig. 1, one possible function of this system is to compress gas.

Figure 2 shows how these definitions are related. The relationship between state and structure is a one-to-many relation. The behavior is produced as the combination of state sets underlined by a given set of principles to the structure. Behavior and function have a many-to-many relation, which depends on the context and usefulness of the structure.

Classification of Design According to Its Synthesis Process

Within this framework, one can classify top-down steps aiming at determining the structure of an artifact given a functional representation as synthesis processes, while their back reasoning counterpart of determining function characteristics given a known structure as analysis processes.

From this perspective, synthesis processes are classified into three groups according to the type of representations:

- Routine design: One in which the space of functions, behaviors, and structures are known, and the problem consists of instantiating structure variables.
- Innovative design: One in which the functions and behaviors are known, and the design consists of generating new structures that satisfy them.
- Creative design: One in which the functions are known, and the problem consists in

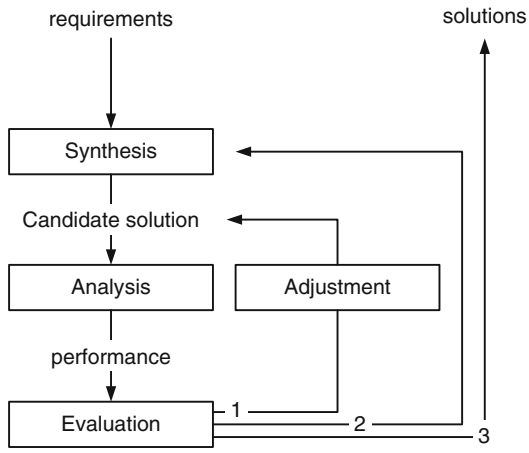
determining the structures and behaviors required to satisfy them.

Furthermore, as nature encompasses a vast variety of behaviors (physical, chemical, human, etc.), synthesis processes can also be classified according to the types of behaviors being targeted:

- Engineering design: Behaviors are characterized by principles stated in the laws of physics. Depending on the discipline of study, engineering design can be further classified into mechanical, electrical, chemical, geological, etc.
- Human-centered design: Behaviors are characterized by physiological, psychological, and emotional human reactions. Two examples are architectural design and industrial design.

Information Flow in Synthesis

Figure 3 shows a well-accepted model of the design process (Schotborgh et al. 2012). According to this model, a candidate solution is generated in a synthesis process. This candidate solution is then analyzed to calculate its performance. Finally, the evaluation process assesses whether the solution is to be adjusted (path 1), rejected (path 2), or accepted (path 3). During the adjustment process, modifications (small) are made to the candidate solution, i.e., without changing the solution principle.



Synthesis, Fig. 3 Generic model of the design process (Based on Schotborgh et al. (2012))

The flow of information through these processes can be classified into three types of information (Webber 2005; McMahon 1994): embodiment, scenario, and performance. Embodiment regards the information that describes the product being designed (e.g., its topology, size, and shape). Scenario regards the information that describes the flow of energy, mass, and signals the embodiment is exposed to. Finally, performance regards the information that determines how the embodiment behaves under a given scenario, and is used to make statements on the quality of the artifact being designed.

The relation between these three types of information varies according to the four processes of the design process model. In the synthesis process, embodiment information is generated (i.e., embodiment parameters are chosen and a candidate solution is formed) such that it meets certain performance parameters for a given scenario. Conversely, in the analysis process performance parameters are quantified or qualified for an embodiment undergoing a given scenario. In the evaluation subprocess, the generated performance parameters are used to determine what follow-up action should be taken (paths 1–3 of Fig. 3). Finally, in the adjustment subprocess small changes to some embodiment parameters can be made in order to improve the performance of the candidate solution.

Cross-References

- ▶ Design Methodology
- ▶ Process
- ▶ Product Development

References

Becker JMJ (2010) From how much to how many. Managing complexity in routine design automation. PhD thesis, Gilde Print, Enschede

Chakrabarti A (2002) Engineering design synthesis: understanding approaches and tools. Springer, London

Chandrasekaran B, Goel A et al (1993) Functional representation as design rationale. IEEE Comp 26:48–56

Gero JS, Kannengiesser U (2004) The situated function–behaviour–structure framework. Des Stud 25:373–391

McMahon CA (1994) Observations on modes of incremental change in design. J Eng Des 5(3):195–209
Pahl G, Beitz W et al (2007) Engineering design: a systematic approach. Springer, London
Schotborgh WO, McMahon C, van Houten FJAM (2012) A knowledge acquisition method to model parametric engineering design processes. Int J Comput Aided Eng Technol 4(4):373–391
Umeda Y, Tomiyama T (1995) FBS modeling: modeling scheme of function for conceptual design. Working papers of the 9th international workshop on qualitative reasoning about physical systems, Amsterdam
Umeda Y, Tomiyama T (1997) Functional reasoning in design. IEEE Expert Intell Syst Appl 12(2):42–48
Webber C (2005) CPM/PDD – an extended theoretical approach to modelling products and product development processes. In: 2nd German–Israeli symposium on advances in methods and systems for development of products and processes, Fraunhofer-IRB-Verlag, Stuttgart
Zhang WJ, Lin Y, Niraj S (2005) On the function-behavior-structure model for design. The 2nd CDEN Conference, Alberta, 18–20 July. CD ROM, 8

System

Kosmas Alexopoulos, Konstantinos Efthymiou and George Chryssolouris
Laboratory for Manufacturing Systems and Automation (LMS), Department of Mechanical Engineering and Aeronautics, University of Patras, Patras, Greece

Synonyms

Arrangement; Organization; Schema

Definition

A system is an integrated composite of people, products, and processes that provides a capability to satisfy a stated need or objective.

Theory and Application

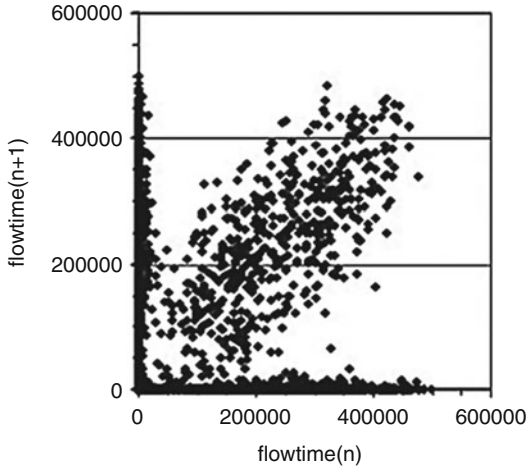
Introduction

The transdisciplinary study of systems which aims to identify and determine principles, applicable to all types of systems independent of the scientific field, is the field of the system theory. System theory as a minimum includes parts/wholes, system/boundary/environment, structure/process, emergent properties, hierarchy of systems, positive and negative feedback, information and control, open systems, holism, and observer. The term "system theory" is originated from the work "general system theory" of von Bertalanffy who first recognized the application of the aforementioned concepts across various disciplines (Bertalanffy 1968). System theory is associated with a vast number of theories of specific disciplines such as cybernetics, soft systems and problem structuring methods, critical systems and multimethodology, systems biology, sociology and sociocybernetics, systems psychology, complexity theory, system dynamics, and system engineering (Mingers and White 2010). The current study emphasizes mainly on the latter three disciplines since they are more related to production systems problems than the other disciplines.

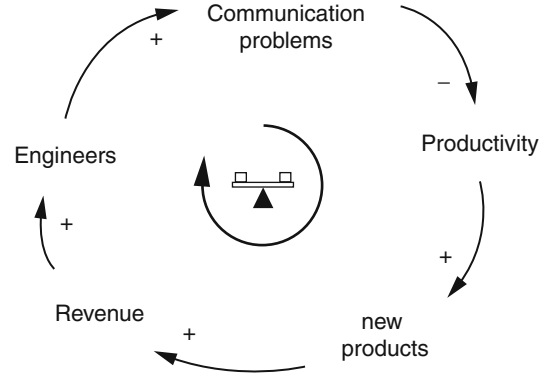
Complexity Theory

The analytical tradition of thinking dominates the majority of scientific theories and models until the beginning of the twentieth century. The analytical tradition is mainly based on two pillars, reductionism and determinism. Reductionism is the process of studying a system by decomposing it into its constituent elements and their fundamental properties. Determinism claims that every change can be represented as a trajectory of the system through state space, following fixed laws of nature. These laws completely determine the trajectory toward the future (predictability) as well toward the past reversibility. Complexity theory studies systems from a different perspective, in a more holistic manner based on the idea that a system is more than just assembling a set of components together. A complex system is considered to be composed of many parts and exhibits emergent behavior. This emergent behavior cannot be inferred directly from the behavior of the components. Nonlinear dynamics and chaos theory follow the notions of complexity theory, and during the last years, various approaches of production systems analysis have been proposed (Fig. 1).

Chaos and nonlinear dynamics techniques are used for the assessment of complexity in



System, Fig. 1 Shortest process time phase portrait (Chryssolouris et al. 2004)

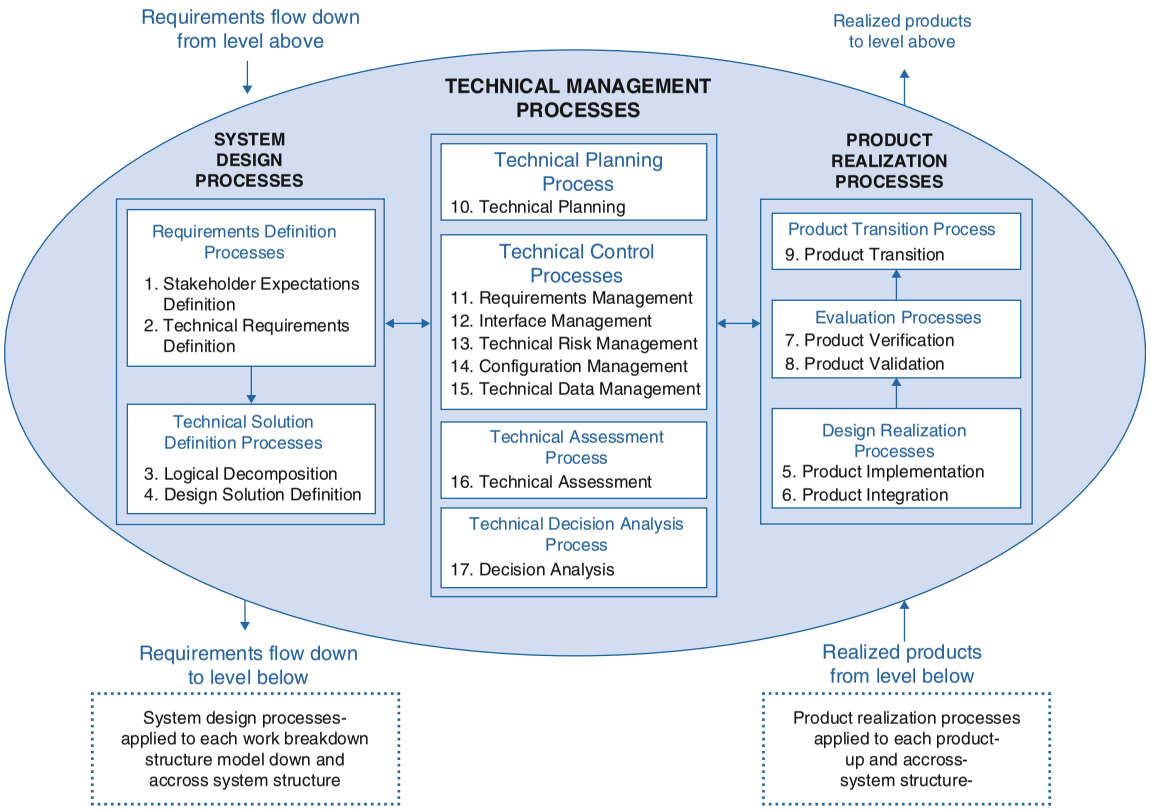


System, Fig. 2 System dynamics simple example modeling

production systems complexity (Papakostas et al. 2009). Phase portraits and time delay plots have been utilized in order to examine the scheduling of simple manufacturing system. Based on this type of analysis, new dispatching rules have been proposed presenting promising results in terms of time performance characteristics. Time delay plots, that is, Poincaré maps, have also been utilized for studying the effect of buffer size on the performance of manufacturing system. The adaptability to demand of a steel construction industry under different operational policies and parameters has been studied, utilizing maximal Lyapunov exponents and bifurcation diagrams. Similarly, maximal Lyapunov exponents also along with Fourier analysis and fractal dimensions have been used for examining the chaotic behavior of a production system based on buffer index time series. A simulation-based method along with regression analysis and nonlinear dynamics analysis has been proposed for the analysis of complex systems. The aim of this simulation-based methodology is the determination of the sensitivity of a manufacturing system to workload changes, the measurement, and the control of the system's complexity (Chryssolouris 2006). Sensitivity analysis is broadly performed in order to identify the chaotic behavior of the system, by introducing small perturbations in system's initial conditions.

System Dynamics

System dynamics is a computer-aided approach to policy analysis and design. It applies to dynamic problems arising in complex social, managerial, economic, or ecological systems, literally any dynamic systems characterized by interdependence, mutual interaction, information feedback, and circular causality (Richardson 1999). System dynamics was developed by Forrester in Massachusetts Institute of Technology, and the main characteristics of this method are the existence of a complex system, the dynamic behavior of the system, and the existence of the closed loop feedback. This feedback describes the new information about the system condition that will yield the next decision (Forrester 1961). The application range of system dynamics is quite wide including domains such as strategy and corporate planning, public management, business process development, energy and the environment, dynamic decision-making, software engineering, and supply chain management. In particular, a variety of system dynamics models have been developed for inventory management and planning activities of production systems. Moreover, the coordination of supply chain over a single period and delivery planning model within a just-in-time setting has been studied with the help of system dynamics (Fig. 2).



System, Fig. 3 The systems engineering engine (NASA 2007)

Systems Engineering

Systems engineering is a methodical, disciplined approach for the design, realization, technical management, operations, and retirement of a system. A "system" is a construct or collection of different elements that together produce results not obtainable by the elements alone (National Aeronautics and Space Administration 2007). The elements, or parts, can include people, hardware, software, facilities, policies, and documents, that is, all things required to produce system-level results. The results include system-level qualities, properties, characteristics, functions, behavior, and performance. The value added by the system as a whole, beyond that contributed independently by the parts, is primarily created by the relationship among the parts, that is, how they are interconnected. NASA identifies three sets of common technical processes: (a) system design, (b) product realization, and (c) technical management. System design processes consist of the definition of the technical requirements based on stakeholder expectations and the conversion of the technical requirements into a design solution that will satisfy the stakeholder expectations. The second process, this of product realization, aims to create the design solution for each product (e.g., by the product implementation or product integration process) and to verify, validate, and transition up to the next hierarchical-level products that satisfy their design solutions and meet stakeholder expectations as a function of the applicable life-cycle phase. Final, the objectives of technical management processes are the following: (a) to establish and evolve technical plans for the project, (b) to manage communication across interfaces, (c) to assess progress against the plans and requirements for the system products or services, (d) to control technical execution of the project through to completion, and (e) to aid in the decision-making process (Fig. 3).

Cross-References

- ▶ Complexity in Manufacturing
- ▶ Factory
- ▶ Manufacturing System

References

Bertalanffy L (1968) General system theory: foundations, development, applications. George Braziller, New York

Chryssolouris G (2006) Manufacturing systems: theory and practice, 2nd edn. Springer, New York

Chryssolouris G, Giannelos N, Papakostas N, Mourtzis D (2004) Chaos theory in production scheduling. CIRP Ann Manuf Technol 53:381–383

Forrester JW (1961) Industrial dynamics. The MIT Press, Cambridge, MA. Reprinted by Pegasus Communications, Waltham

Mingers J, White L (2010) A review of the recent contribution of systems thinking to operational research and management science. Eur J Oper Res 207:1147–1161

National Aeronautics and Space Administration (2007) NASA systems engineering handbook, NASA/SP-2007-6105 Rev1, NASA Headquarters Washington, DC, 20546

Papakostas N, Efthymiou K, Mourtzis D, Chryssolouris G (2009) Modeling the complexity of manufacturing systems using non-linear dynamics approaches. CIRP Ann Manuf Technol 58:437–440

Richardson GP (1999) System dynamics. In: Gass S, Harris C (eds) Encyclopedia of operations research and information science. Kluwer, Boston

Systematics

- ▶ Cladistics for Products and Manufacturing