16th Brazilian Congress of Mechanical Engineering

Engineering for the New Millennium

INVITED LECTURES

ISBN 85-85769-06-6

Invited Lectures

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

SMART STRUCTURES: EXAMPLES AND NEW PROBLEMS

Inman, D. J Center for Intelligent Material Systems and Structures
Department of Mechanical Engineering
310 NEB, Mail Code 0261
Virginia Tech
Blacksburg, VA 24061
Phone:540 231 4709 fax 231-2903
http://www.cimss.vt.edu , dinman@vt.edu

Abstract: *Smart materials, or active materials, offer a host of new solutions to engineering problems in control and identification. In addition they offer new modeling and analysis challenges. Here we examine the use of smart materials in structural applications: called smart structures. An introduction to smart structures and their use in vibration suppression, flight control and structural health monitoring is presented. A summary of the issues and problems of structural health monitoring using smart materials is presented. The article concludes with some thoughts on open problems in smart structures and structural health monitoring.*

1. Smart Materials and Structures

Evolution processes became recently one of the most important research focuses of applied mathematics and computational modeling. The driving force behind this importance research field is the application it finds mainly in biological, ecological and social systems.

This extraordinary advance, I dare to say, is mainly due to the computational capacity at our disposal nowadays. The progress in the development of hard- and software as well, has enabled us with the possibility of dealing with thousands and hundred of thousands equations that appear frequently in the solution of complex systems. As a matter of fact, computational modeling, considered in its broad spectrum, is growing very rapidly, combining the contribution of natural scientists, mathematicians, engineers and social scientists to the solutions of complex problems. This mixing of people coming from different areas of knowledge is creating new interdisciplinary areas that will probably be of fundamental importance in the near future, if not right now.

This paper presents a model that is intended to help understanding the process of knowledge transfer among groups belonging to a certain segment of the same society or possibly the interaction process between different societies. The first version of this paper was published a long time ago, in 1980 [1]. After that, no effort was made to elaborate the theory or test the theory against some empirical data. A relatively recent paper was encouraging and has shown that similar ideas were being explored. This reference [2] considers both genotype and phenotype factors to study the evolution of social groups. That is, inheritance and social interaction are both shaping the history of the group. We are interested here in a relatively short period of time, and the horizontal aspect, interaction among individuals or sub-groups constitutes the predominant factor.

Despite the limitations of the model, the results are plausible and intuitively consistent. So, if the model doesn't reflect completely the reality nonetheless it allows for some enlightenment that puts in evidence important and some times hidden correlation among concrete facts. If taken judiciously the results could help decision makers to implement important actions to foster the generation and transmission of knowledge. It is never too much to insist that social modeling, where the behavior of human being play a key role should be seen very carefully. It is impossible to model perfectly the behavior of human social groups and frequently we are faced with big surprises. It is a system that admits contradictory outcomes, where adaptation and rebellion walk side by side.

In the next sections it is presented a temptative approach to the process of transfer of knowledge. Certainly the model is a simplification but as a first approximation suggests very interesting correlation, particularly the relative importance of knowledge generation, knowledge transmission and learning speed.

2. Structural Health Monitoring

Structural heath monitoring (SHM), also called damage detection or diagnostics, is the concept of using structural measurements to determine the condition, integrity or state of a structure. A basic goal of SHM is to determine if a structure is in danger of failing or not. Examples of common objectives of SHM are to determine if and cracks exist in the structure, if bolted or welded joints are intact and up to specification, if there are any holes or other structural damage present in the structure. In some sense one might fit this topic in with non-destructive evaluation (NDE) methods. The difference between classic NDE methodology and SHM is that of immediacy. In NDE parts or systems are taken out of service to be inspected and a goal of SHM is to leave the structure in service while the analysis is performed. An example of current SHM is the NASA space shuttle that undergoes a vibration test (modal test) before and after each flight. Differences in modal information are then used to determine damage incurred during the flight. While not quite an “on-line” system it does capture the function of SHM.

The problem of damage analysis can be sub divided into four sub problems or levels of health monitoring (Doebling, et al, 1998):

1) Detect the existence of damage.
2) Detect and locate the damage.
3) Detect, locate and quantify the damage.
4) Detect, locate, and quantify the damage then estimate its remaining service life.

Each level requires more modeling and hence more mathematics in order to solve. The Level 1 problem has numerous solutions listed in the literature and in fact has been implemented in practice. Fewer solutions are available for the remaining topics. Level 4 leads to the emerging area called "Prognosis" which has few solutions and is motivating several large programs in government laboratories. Here we propose to expand this list to include several more possibilities:

5) Combine Level 4 with smart structures to form Self Diagnostic Structures.
6) Combine Level 4 with smart structures and control to form Self-Healing Structures.
7) Combine Level 1 with active control and smart structures to from simultaneous control and health monitoring.

Taking the simplified point of view that health monitoring involves looking for changes in a system's physical parameters (such as mass, damping or stiffness) leads to the observation that much of the mathematics associated with SHM will come from the field of parameter identification. This is clear in the work of Banks et al (1996b) who solve a Level 3 heath-monitoring problem by adapting techniques form identification theory to solve a damage detection problem for simple beams using piezoceramic actuators and sensors, as well as traditional instrumentation.

Obviously, each level problem becomes more difficult to solve. The problem of Level 1 can be solved with out any reference to a structural model by using simple signal analysis. One example of such a solution is given by Cattarius and Inman (2000). In this technique the idea often used in damage detection that a defect will produce a small change in stiffness and/or mass, and hence frequency is used. Early vibration based damage detection methods often looked at frequency response function (FRF) data for a small change in frequency. However small changes in frequency are generally difficult to measure using FRF data. As an alternative, Cattarius and Inman (2000) continuously compared the healthy time signal of the structure under study to the current time signal of the structure. If a small difference in frequency exists between these two signals, then when they are combined they will produce the beat phenomena that serves to magnify small differences in frequencies. This effect is used detect the presence of damage in plates and then in a helicopter blade by using internal piezoceramic materials to both excite and sense the various time histories. This is and example of a time domain procedure that is totally self-contained in a moving structure, capable of self-diagnostics using embedded piezoceramic sensing and actuation at relatively low power costs. The procedure depends only upon subtracting two signals, does not require any modeling of the structure and hence is simple enough for on board use.

Another diagnostic procedure not relying on any mathematical model is impedance based and may also be used in a self-diagnostic configuration by incorporating local, embedded and/or surface mounted piezoceramic sensors and actuators. This approach is a high frequency impedance based method that looks for a shift in electrical impedance measurements as an indicator of damage. In the impedance-based qualitative health monitoring technique, real-time implementation relies on a simple scalar damage index that can be easily interpreted. Using this damage index in conjunction with a damage threshold value, the approach can warn an operator in a green/red light form, whether or not the threshold value has been reached. A damage index that is not affected by environmental effects was established based on statistical analysis and signal processing, making the damage index more stable under all types of environmental conditions (Park, et al, 1999, 2000). Figure 1 illustrates the use of a self-contained impedance based system of a frame with bolted joints. Such a system represents the complexity of a real service structure, yet allows for systematic "damage" in the structure by adjusting the torque on the joints.

Fig. 1 Impedance based damage detection of a piping system using piezoceramics.

In Banks et al (1996b) a method which takes advantage of knowledge of the structure is presented to provide experimental proof that a PZT based diagnostic system can not only determine the existence of damage, as can the previous methods, but is also able to determine the size and location of the damage (Level 2 and 3). The method is based on using a partial differential equation model of a structure that is partially layered with PZT patches which again forms a “self-diagnostic” structure. The algorithm uses a spline-based approximation of the equations of vibration and successfully identifies the existence, size and location of holes in a beam. The method works by estimating functions of the longitudinal direction of the beam corresponding to the damping parameters (both Kelvin-Voigt and air damping), the modulus and the density. Each of these is allowed to be discontinuous in order to allow holes to be included in the solution set of functions.

It is important to note, that while this method works very well, it has the disadvantage of requiring a very detailed model of the structure including a model of the internal damping mechanisms. The parameters are allowed to have some measure of uncertainty as they are estimated in the inverse procedure used to identify the damage. However, the form of the governing differential equation must be known in substantial detail. With this noted, the method very effectively identifies the damage. Furthermore, the results are consistent across several different types and sets of measurements.

Another important feature of the estimation based method is that it does not use modal data, but rather is based on time domain measurements. The experiments where repeated with traditional excitation means (an instrumented hammer) and response measurement (accelerometer, and position probe) as well as with the internal self-sensing actuation scheme offered by the PZT patch. Besides verifying a damage detection algorithm, the tests show conclusively that it is possible to use piezoceramic materials to form self-monitoring devices.

In the following the impedance method is used to illustrate a Level 6 and Level 7 solutions. The main point of using the impedance method here is that the method works off of a very low voltage (~1volt) excitation in the kilohertz range so that it does not interfere with the control law for active vibrations suppression that typically takes place at low frequency. Furthermore, it is shown that this method detects damage while the control law is turned on so that vibration suppression is occurring simultaneously. This works because the high frequency impedance signal is not part of the control bandwidth.

The basic concept of this impedance-based structural health monitoring technique is to monitor the variations in the structural mechanical impedance caused by the presence of damage. Since structural mechanical impedance measurements are difficult to obtain, this non-destructive evaluation technique utilizes the electromechanical coupling property of piezoelectric materials. This method uses one piezoceramic (PZT) patch for both sensing and actuating. The PZT is considered as a thin bar undergoing axial vibrations in response to the applied sinusoidal voltage. Assuming that the piezoceramics' parameters remain constant any changes in the mechanical impedance (Z_s) change the overall admittance of the combined structure and PZT system. Previous experiments have shown that the real part of the overall impedance contains sufficient information about the structure and is more reactive to damage than the magnitude of the imaginary part. Therefore, the impedance analyses is confined to the real part of the complex impedance. The actual health monitoring is performed by saving a healthy impedance signature of the structure, and comparing the signatures taken over the structure’s service life. The impedance measurements are taken with a HP 4194A Impedance Analyzer, however a simple measurement across a resistor would do. A frequency range from 45 kHz to 55 kHz proved to be an optimum for this structure.

To simulate damage on a plate bolted on all four sides to a frame, one or two bolts of the clamping frame were loosened from 25 ft-lb. to 10 ft-lb. Note that at this level of torque the bolt is still very tight with no slip occurring at all. For comparing impedance signatures, a qualitative damage assessment has been developed. The assessment is made by computing a scalar damage metric, defined as the sum of the squared differences of the real impedance at every frequency step. Equation (1) gives the damage metric M in a mathematical form. The used variables include: $Y_{i,1}$ the healthy impedance at the frequency step i, $Y_{i,2}$ the impedance of the structure after the structure has been altered, and n the number of frequency steps

$$M = \sum_{i=1}^{n} \left[\mathrm{Re}(Y_{i,1}) - \mathrm{Re}(Y_{i,2}) \right]^{2} \tag{1}$$

The damage metric simplifies the interpretation of the impedance variations and summarizes the information obtained by the impedance curves. Different damage metric values of the plate are presented in Figure 2. Note the difference in the metric between one bolt loosened and two bolts loosened.

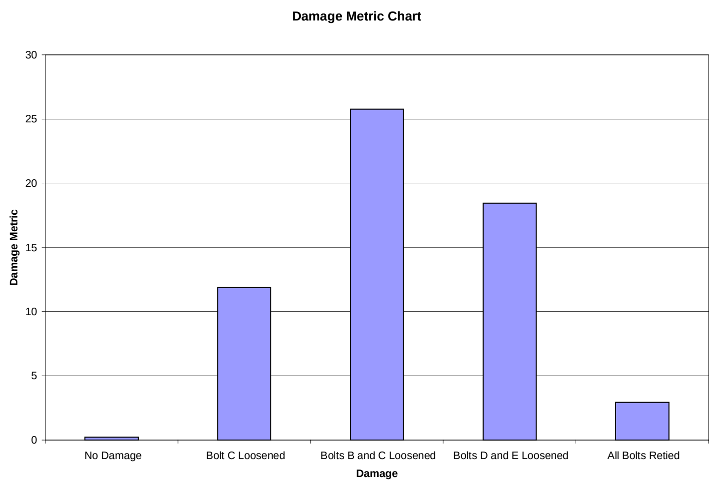


Fig. 2 Damage metric chart of different impacts to a plate.

3. Vibration Suppression

Vibration of aircraft panels is a major source of fatigue and requires the addition of damping material to increase fatigue life and reduce noise transmitted to the interior. The added damping treatments bring a substantial weight penalty. In addition passive damping treatments are limited in frequency range and subject to variation with temperature. In this section the possibility of using active vibration suppression implemented through smart materials to perform damping in structural plates subject to both structural and acoustic loading is examined using an experimental approach. Results are presented indicating a high level of damping available through a single piezoceramic patch serving as an actuator and a fiber optic sensor.

Because modeling and boundary conditions in a real aircraft panels have significant variance, a control method is chosen that is based only on knowing experimentally determined frequency response data for the structure. Positive position feedback (PPF) control based on a measured frequency response to the structure is chosen as the control law (Fanson and Caughey, 1990). This technique combined the technology of optical fiber sensors with piezoelectric actuators to minimize the vibration levels in the test article. This PPF control law uses a generalized displacement measurement from the test article to accomplish vibration suppression. The experiments were performed in a standard test fixture commonly used in industry for evaluation damping materials.

The experiments show that smart damping materials have substantial performance benefits in terms of providing effective noise and vibration reduction at a frequency range that is often outside of the effective range of passive damping materials. Further, judging by vibration reduction per added weight, the test results indicate that the smart damping materials can provide substantial vibration reduction at selected frequencies, without adding any appreciable amount of weight to the substrate structure. For example, smart damping can decrease the vibration peak of a steel panel at 47 Hz by up to 20 dB with an additional mass of only 50 grams. This feature of smart damping materials is particularly useful for applications that involve vehicles, where the constraint requires a particular noise or vibration cancellation at a specified frequency, without adding any weight to the vehicle or requiring any change to the vehicle structure.

Overall, the test results show that the application of smart damping materials and fiber optic sensors can be used for active control of aircraft style panels. The smart damping materials combined with fiber optics can be viewed as a new technology that, once developed and optimized, can extend future life by providing more effective noise and vibration solutions for new or existing aircraft structures.

Fiber optic sensors are widely used as physical parameter gauges in various structural applications, such as strain and vibration sensing and damage detection. The Fabry-Perot (FP) interferometric strain sensors can be classified into two main types: intrinsic FP interferometers (IFPI) and Extrinsic FP interferometers (EFPI). The EFPI sensor is constructed by fusion splicing a glass capillary tube with two optical fibers. Compared to the IFPI sensor, the EFPI-based sensor is relatively simple to construct and the FP cavity length can be accurately controlled. The EFPI sensor can also be easily configured to suit different applications with desired strain range and sensitivity by altering the type of fibers, the capillary tube, air-gap distance and the length of the sensor. In addition, a major advantage of the EFPI sensor is its low temperature sensitivity, which makes it possible to interrogate the EFPI sensor with simple signal processing techniques. An EFPI sensor can be constructed using either single-mode or multi-mode fibers. The single-mode design offers higher accuracy and low insensitivity to unwanted disturbance while the multi-mode design offers higher power coupling efficiency. In our experiment, single-mode fibers are used to deliver and collect the light, and are used as an internal reflector as well.

The test panel used for the active control tests is a 500mm X 600mm, 20-guage, galvanized steel plate with two 72.4 mm X 72.4 mm piezoelectric actuators bonded to its surface. These two locations were chosen for the actuators because the plate has a large amount of strain energy in these regions for the modes that needed to be controlled. In general, the control authority of an actuator is increased when it is placed in a region of high strain energy. Although the test plate has two piezoelectric actuators attached to its surface, it was determined that the center PZT was more effective at minimizing the levels of vibration.

The sensor used for control was a modal domain optical fiber sensor for vibration monitoring. This sensor is based upon a laser that focuses coherent light through a lens into one end of a multimode optical fiber. One end of the fiber was attached to the plate and the other end of the optical fiber passed through a spatial filter and into a photo detector. The output of the photo detector is a variable voltage that is fed into a monitoring unit such as an oscilloscope.
The control system was set up such that the optical fiber sensor sends a signal through a signal conditioner and amplifier into the dSPACE board. Once the signal was processed through the PPF control law it was sent out to an amplifier to drive the PZT actuator. The system's parameters and the overall control system's performance were measured using a two channel HP Dynamic Signal Analyzer. This signal analyzer was used to get the frequency response function between the plate and the clamping frame using two accelerometers, one on the bottom center of the plate and one on the clamping frame.

A series of tests were run to determine the optimum parameters for the active vibration suppression of a representative aircraft panel. These parameters, ranging from the gains assigned to the various amplifiers to the damping ratios of the PPF filters, were based upon the tuned frequency and the number of modes assigned to each PPF filter. The final Simulink Model used for the active control tests uses three PPF filters to provide active damping to most of the modes from 0-400 Hz. However, you can achieve significant active damping of a mode or modes with one PPF filter tuned properly. There were two different means of supplying disturbance energy to the plate for the active control tests. The first method used a 100 lb shaker to excite the plate mechanically, and the second method used a 10" sub-woofer to excite the plate acoustically. Both the shaker and the speaker were driven with a 0-400 Hz periodic chirp input signal. The initial design procedure for the active control system was to create a Simulink model with one PPF filter to control as many modes as possible. This design method involved determining the frequency range at which the system had control authority over multiple modes. The tests results show that the control system with one PPF filter was most effective when tuned to a frequency between 60 and 70 Hz. Figures 3 show some of the best results obtained with the one PPF filter controller. In each of these tests the 100 lb shaker excited the plate mechanically. The active control test was run with a PPF filter damping ratio of 0.02, a Simulink gain of –80,000, a sampling time of 0.0001 seconds, and tuned to a frequency of 60 Hz.

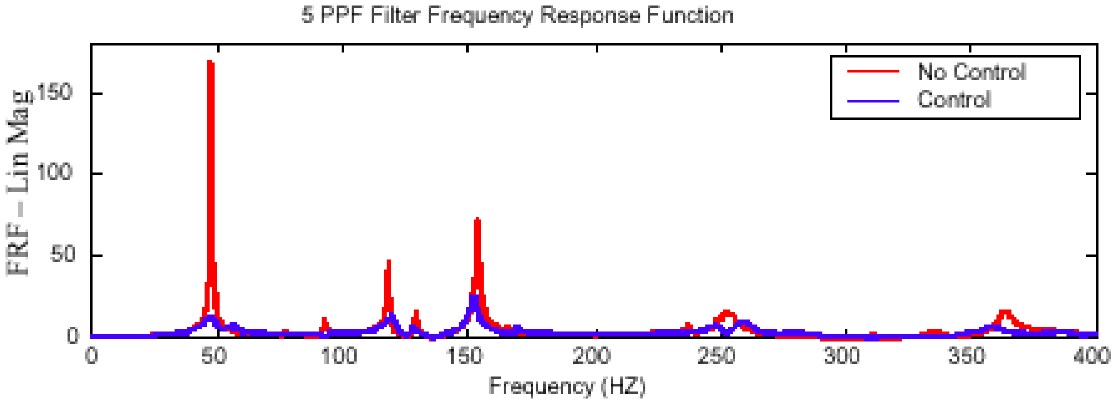


Figure 3. Broadband Vibration Suppression Using Five PPF Controllers

Additional tests were run using pure tone inputs at the resonant frequencies of the plate. These tests were performed to determine the effectiveness of the controller in the suppression of the structure born noise resulting from the vibration of the test plate. Most of these tests were run at the first resonant frequency of the plate (47 Hz) because this mode produced the most audible structure born noise. The test results proved that the controller reduced the noise from the plate significantly. Figure 3 shows the best reduction achieved by the controller with a 47 Hz pure tone input signal.

This experiment represents the first use of a fiber optic sensor in an active vibration suppression system with piezoceramic actuation. The controller is not based on an analytical model, rather it is based on an experimentally determined modal model of the structure. The performance obtained via the active control system is comparable to and in some cases better then that obtained by passive damping treatments. The piezoceramic actuator and fiber optic system weigh much less then a constrained layer damping treatment for similar performance. However, the mass of the control hardware must also be considered. While the control system here was obtained using off the shelf components (dSpace, MATLAB) which are of substantial size and weight, it is possible to put this entire control system, signal conditioning, etc onto a single chip of just a few grams (Inman, 1997).

4. Simultaneous Diagnostics and Control (Level 7)

The problem of simultaneously monitoring the health of a structure and active vibration suppression on the same structure represent opposing goals. The goal of active vibration suppression is to reduce the structure's response as quickly as possible. On the other hand, diagnostic, or health monitoring, generally requires that the response be examined as long as possible. Furthermore, if the same hardware is to be used for both vibration suppression and health monitoring the possibility of signals' interfering with each other presents itself. To solve these problems the impedance method of health monitoring it used. The impedance method uses a high frequency, low voltage, self-sensing signal to determine small changes in stiffness, damping and mass of a structure.

The control law chosen needs to not interfere with the diagnostic measurements and at the same time provide robustness to analysis. In particular the control must not depend on an analytical model and must continue to provide vibration suppression as the panel heats up. In the case of the aircraft panel demonstrated here, the control hardware and diagnostic hardware were required to be the same in order to minimize the amount of hardware in an aircraft. A piezoceramic patch was chosen for the actuator and sensor, building on the concept of a self-sensing actuator (Dosch, et al, 1992). The control law was chosen to be Positive Position Feedback because it can be applied to an experimental model and because its closed loop stability depends only on knowing the systems natural frequencies. Thus, the PPF control law may be applied with hope of success as long as experimentally determined frequencies are available.

Since the task consisted of simultaneous health monitoring and active control with the same actuators for both the control and the health monitoring system needed to be de-coupled. The impedance method is very sensitive to disturbing voltages in the measuring circuit. The controller however creates exactly those disturbances by generating the control signal. A simple capacitor of 390 nF in series with the impedance analyzer blocked efficiently the control signal from the impedance analyzer. All health monitoring data was taken while the shaker was exiting the plate with a periodic chirp signal from 0 to 200 Hz. The active controller was also switched on and increased the damping of the first few modes of the plate significantly. The control results are straightforward. First an experimental result is given for the case of constant temperature. Figure 3 illustrates the open loop transfer function along with the closed loop control provided by five PPF controllers. The controller is able to suppress the all the modes in the bandwidth from 0 to 400 Hz.

The ability to perform simultaneous health monitoring and active vibration suppression has been illustrated and experimentally verified. Furthermore, the active control of panel vibrations in the presence of temperature changes has been illustrated in Figures 2 and 3, which were made simultaneously. These experimental results do not depend on any model of the system and hence are applicable to complex structures. The control method and the health monitoring method were performed with the same hardware at the same time. These results indicate that it may be possible to perform simultaneous health monitoring and control on real structures outside of the laboratory setting.

Applications of smart materials to problems of interest to the private sector are now fairly routine to solve. In addition the use of smart materials has been shown to be extremely beneficial in numerous cases. The only thing that holds back the wide spread use of smart material in industry is the lack of good, design-based models that are easily accessible by industrial designers.

5. Self-Healing (Level 6)

This section illustrates the feasibility of creating smart structural bolted connections, which consist of structural members joined together by bolt and nut combinations equipped with piezoceramic and shape memory alloy elements. These combinations can be used to monitor bolt tension and connection damage. When damage occurs, temporary adjustments of the bolt tension can be achieved actively and remotely in order to restore lost torque for continued operation, thus illustrating a self-healing systems. The system is illustrated in Figure 4. The piezoceramic is used to detect the existence of damage (an out of torque bolt in this case) and the Shape Memory Alloy (SMA) is then use to regain the lost torque by applying an increased normal force when activated. The impedance for the various levels of torque is illustrated in Figure 5 which shows that the torque is recovered, at least partially.

A test specimen consisting of two aluminum beams was constructed with a bolted joint. A list of dimensions of the test specimen is given in table 1. The bolted-joint structure was hung vertically by a string. One PZT patch bonded to one of the members was used to measure the electrical impedance. An SMA washer (Intrinsic Devices AHE 0957, Figure 1) was inserted between the bolt and the nut, as illustrated in Figure 2. Initially, the bolt was tightened to 30 ft-lb, and the torque was reduced to 10 ft-lb to introduce a loosening mode of a bolt failure. This damage however can be considered in its incipient stage, which still maintains the integrity of the joint. This level damage corresponds roughly to a quarter turn of the bolt, so that the joint is still very tight, but not up to the design specified torque.

Figure 5 summarizes the basic proof of concept experiment. The impedance monitoring system is activated while the bolt is at its design specified value of torque. The PZT patch applies a 1-volt excitation signal in the kilohertz frequency range that continuously monitors the system by comparing the impedance signal of a healthy system to the impedance of the current system. When the impedance signal changes, the bolt is then indicated to be out of torque. This initiates a current to the SMA washer, resulting in a recovery of the lost torque (Inman et al, 2001).

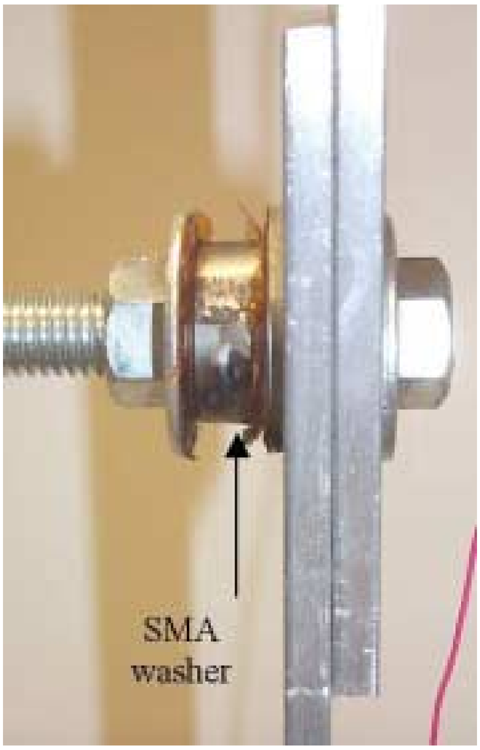


Fig. 4: The proof of concept lap joint assembled with SMA washer and PZT impedance sensor and actuator.

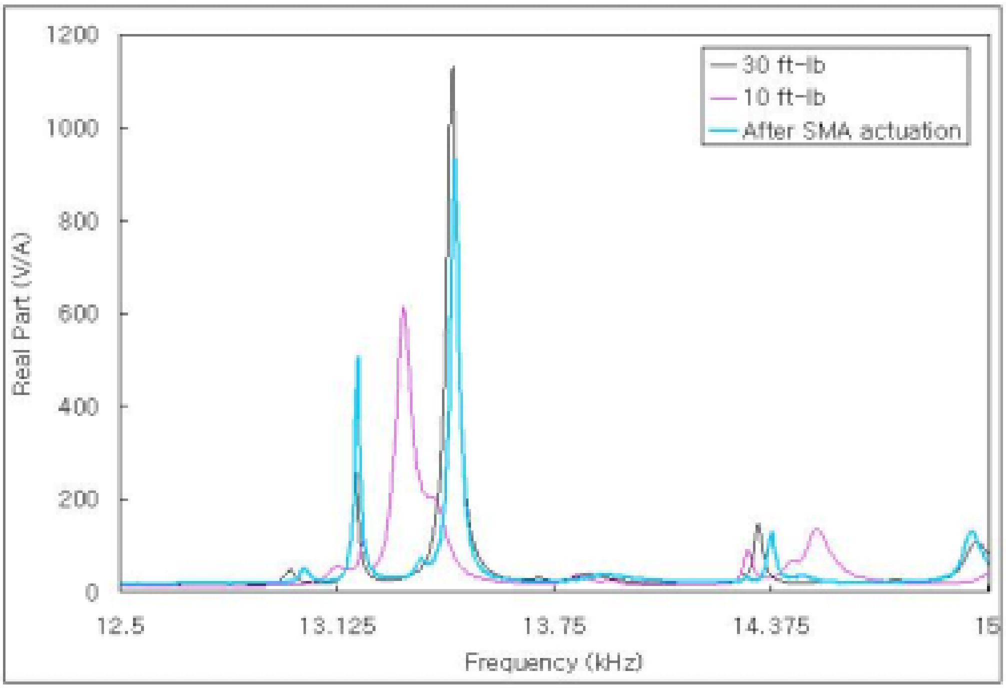


Fig. 3 The impedance showing the deterioration of impedance as the torque is reduced, followed by the recovery of impedance as the SMA is activated.

6. Morphing Aircraft

Aircraft are becoming smaller, lighter, and more flexible, while their performance requirements are becoming more stringent. This is especially true for the newly conceived Uninhabited Combat Air Vehicles (UCAV) or Micro Air Vehicles (MAV). Recently, a great effort has been put forth in designing “smart wings” for uninhabited as well as conventional aircraft. Smart wings as used here refers to the use of various active materials like shape memory alloys, piezoceramics, electrostrictives, MEMS, etc. to bend, twist and change the surface of a wing in flight. This way, it becomes possible to obtain more favorable lift and drag properties, replace flaps or ailerons, prevent aeroelastic instabilities and increase vehicle performance. Figure 6 shows a comparison between a morphing wing (twisting) and traditional flap in terms of required power. Note that at high speed the twisting wing concept is better then moving a flap, pointing to an advantage in using a smart structures approach (Gern, et al, 2000). Morphing wings are a natural application of smart materials (Petit, 2001).

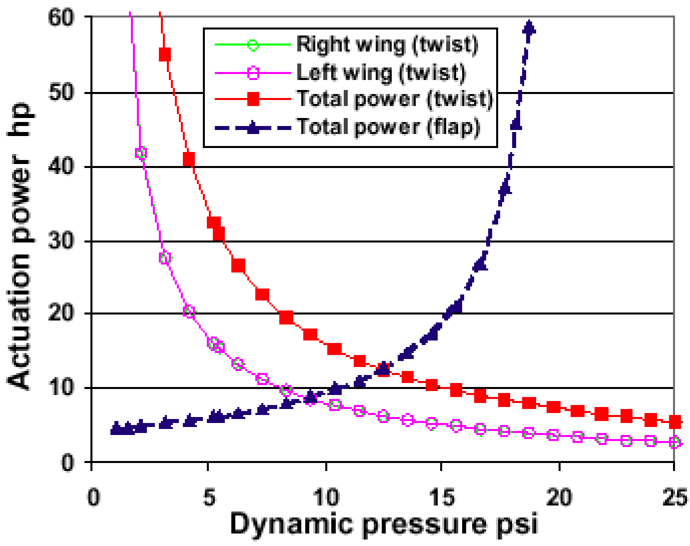


Fig. 6 Actuation power vs. flight dynamic pressure for the morphing wing with twist actuation. The rolling moment created is equivalent to the one obtained by a 5º antisymmetrical flap deflection at reference conditions (950,000in-lbf).

7. Inflated Satellites

The emerging area of inflated satellites is also a natural application of smart materials and structures. Inflated space-based devices have gained popularity over the past three decades due to their minimal launch-mass and launch-volume. With a recent surge in inflatable structure applications, ground testing has become supercritical in understanding the structures' dynamic behavior and in verifying predictive modeling techniques. However, current attempts at ground tests have been difficult because of the unique nature of an inflated structure and the consequent difficulty of sensing and actuation. The extremely flexible nature of the structure produces only local deformation of the skin rather than excitation of the global modes during point excitation. Unfortunately, the global modes are necessary for model verification and parameter identification. Testing of these structures has proved challenging, and development of proper testing procedures still presses onward. In addition, the choice of suitable sensing and actuation materials for use in an inflated structure is limited because of the requirement that the system must be able to fold up prior to deployment.

Fig 7 Ground test of an inflated torus.

In Park, et al (2001) we experimentally investigated if smart materials (for instance, PVDF films) could be used for a modal analysis of a flexible inflated structure. Piezoelectric materials, or smart materials, produce an electric field when a mechanical strain is applied, and vice versa. This effect has already been successfully implemented into vibration testing and control for various structures. Because the sensor and actuator materials must be flexible enough to conform to the toroidal shape, smart materials are most compatible with inflatable structure applications. The variety of sizes, low mass and power requirements, and fast time response offered by smart materials make them an excellent candidate for use in the dynamics and control of inflatable structures.

The use of smart materials in the dynamics of inflatable structures has been confirmed in recent tests. A PVDF patch was used to measure the vibration response and the MFC^{TM} actuator was used as an excitation device in modal tests of an inflatable structure, a torus. The experimentally obtained modal parameters of two different test types, shaker--accelerometer and MFC--PVDF sensor, are in good agreement with each other. The actuators and the PVDF sensors hardly interfere with the suspension modes of a free-free boundary condition of the torus. The experimental results presented validate the usefulness of flexible PVDF sensors for measuring the dynamics of inflatable space structures.

8. Summary

A variety of applications for smart materials to vibration suppression, vibration testing, flight control and structural health monitoring have been presented. In each case the use of smart materials integrated into a smart structure forms a unique solution to an important problem. It is hoped that this lecture will stimulate others to attempt using smart materials to solve other engineering problems.

9. Acknowledgements

The author would like to thank the Flight Sciences Department, Raytheon Systems Company for both suggesting and funding this work and for the excellent technical monitoring of Richard A. Ely. In addition, thanks is due the National Science Foundation (CMS-9713453-001) and the Airforce of Office of Scientific Research (F49620-99-1-0231) for funding work in control, smart structures and health monitoring used in this effort. Dr. Guyhae Park provided the expertise in health monitoring using the impedance-based method and with the inflated structures. Dr. Frank Gern provided figure 6. Thanks are extended to both for their continued support.

10. References

Banks, H. T., Smith, R. and Wang, Y., 1996, *Smart Materials and Structures: Modeling, Estimation and Control*, Wiley

Banks, H.T., Inman, D.J. Leo, D. J. and Wang,Y., 1996, "An Experimentally Validated Damage Detection Theory in Smart Structures," *Journal of Sound and Vibration*, Vol. 191 (5), pp. 859-880.

Cattarius, J. and Inman, D. J., 2000. “Experimental Verification of Intelligent Fault Detection in Rotor Blades,” *International Journal of Systems Science*, Vol. 31 No. 11, pp. 1375-1379.

Caughey, T.K. and C.J. Goh, 1982, “Analysis and Control of Quasi Distributed Parameter Systems,” California Inst. Of Technology, Pasadena, CA, Dynamics Lab. Rept. DYNL-82-3.

Culshaw, B., 1996 *Smart structures and Materials*, Artech House.

Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W., 1998, *"A Review of Damage Identification Methods that Examine Changes in Dynamic Properties,"* Shock and Vibration Digest, 30 (2) pp. 91-105.

Dosch, J. J., Inman, D. J., Garcia, E., 1992, “A Self-Sensing Piezoelectric Actuator for Collocated Control,” *Journal of Intelligent Material Systems and Structures*, Vol. 3, pp. 166-185.

Fanson, J. L., Cauhey, T. K., 1990, ”Positive Position Feedback Control for Large Space Structures,” *AIAA Journal*, Vol. 28, No. 4, pp. 717-724.

Inman, D. J., 1997, “Vibration Suppression Through Smart Damping," Proceedings 5th International Conference on Sound and Vibration, Vol. 1, pp. 115-132.

Inman, D. J., Muntges, D. E., and Park, G., 2001, “Investigation of a Self-Healing Bolted Joint Employing a Shape Memory Actuator,” *Proceedings of the SPIE 8th Annual International Symposium on Smart Structures and Materials*, Newport Beach, California, 4-8 March, Vol. 4327.

Gern, Frank H., Kapania, Rakesh K., and Inman, D. J., “Structural and Aeroelastic Modeling of General Planform UCAV Wings with Morphing Airfoils,” *AIAA Journal*; submitted 2000.

Park, G., Cudney, H., Inman, D. J., 2000 “An Integrated Health Monitoring Technique using Structural Impedance Sensors,” *Journal of Intelligent Material Systems and Structures*;.

Park, G., Kabeya, K., Cudney, H. H. and Inman, D. J., 1999. “Impedance-Based Structural Health Monitoring for Temperature Varying Applications,” *JSME International Journal*, Series A, Vol. 42, No. 2, pp. 249-258.

Park, G, Ruggiero, E. Suasse, M, and Inman, D. J., 2001, “Vibration testing and Analysis of Inflatable Structures using Smart Materials” in review.

Petit, C. W., 2001, “Up, Up and Away”, *U.S. News and World Report,* May 21, 2001, pp. 46-47.

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

INTERDISCIPLINARY STUDIES IN MECHANICS: A DIFFUSION MODEL FOR THE KNOWLEDGE TRANSFER

L. Bevilacqua
Laboratório Nacional de Computação Científica/Ministério da Ciência e Tecnologia
Av. Getúlio Vargas 333 Quitandinha
25651-070 Petrópolis, RJ, BRAZIL.
e-mail: bevi@lncc.br

M.E.P. Bulnes
Laboratório Nacional de Computação Científica/Ministério da Ciência e Tecnologia
Av. Getúlio Vargas 333 Quitandinha
25651-070 Petrópolis, RJ, BRAZIL.

Abstract: *The advance in computing methods, applied mathematics, computational science are changing several aspects of science and technology. One important aspect is the formation of interdisciplinary groups. One of the better ways to link persons from different backgrounds is the development of models. Particularly, social scientists and hard sciences researchers will communicate better if they are interested in a common problem with challenging questions for both sides, all of them contributing to the development of an integrated model. This paper proposes an exploratory model for the process of knowledge transfer among social groups starting from a well known diffusion problem, namely the heat conduction in a straight rod. The growth of a group depends on the creativity not on the cleverness or the capacity of learning quickly. There is no use to provide excellent mechanisms for learning and teaching, accelerating the flow of knowledge between adjacent segments if the creativity remains low. Better means for knowledge transfer are only meaningful if the creativity is high. The extrapolation of a theory good for natural phenomena to explain social phenomena is always a delicate task. One of the most commons criticism is the difficulty involved in the experimental aspect of the investigation. The constraints imposed by the ethical commitment to society and individuals reduce drastically the flexibility of the experiments. Therefore the experimental proof of a theory in social sciences is difficult and requires accurate observations and careful interpretations. Starting from a plausible theory that, at least, doesn't lead to contradictions as compared with observations, it is possible to find consistent guidelines for experimental work, as the identification of key variables and the proper correlation among them.*

1. Introduction

The advance in computing methods, applied mathematics, computational science are changing several aspects of science and technology. One important aspect is the formation of interdisciplinary groups. We believe that one of the better ways to link persons from different backgrounds is the development of models. Particularly, social scientists and hard sciences researchers will communicate better if they are interested in a common problem with challenging questions for both sides, all of them contributing to the development of an integrated model.

This paper is intended to propose an exploratory discussion of a model for the process of knowledge transfer starting from a well known diffusion problem, namely the heat conduction in a straight rod.

The extrapolation of a theory good for natural phenomena to explain social phenomena is always a delicate task. One of the most commons criticism is the difficulty involved in the experimental aspect of the investigation. The constraints imposed by the ethical commitment to society and individuals reduce drastically the flexibility of the experiments. Therefore the experimental proof of a theory in social sciences is difficult and requires accurate observations and careful interpretations. Starting from a plausible theory that, at least, doesn't lead to contradictions as compared with observations, it is possible to find consistent guidelines for experimental work, as the identification of key variables and the proper correlation among them.

Building up multidisciplinary research groups is not easy, but once established they will certainly open new avenues for the progress of scientific knowledge. Modeling could be an effective approach to gather investigators from different fields of science and to keep them working as a team.

I must apologize for the simple way in which the paper has been written. I am sure that several parts could have been skipped, for the audience is aware of the basic concepts. I expect, however, this paper to be eventually read by people from other fields of knowledge less familiarized with mechanics and applied matehematematics. This explains the introductory character of some parts of the paper.

2. Evolution Processes

Evolution processes became recently one of the most important research focuses of applied mathematics and computational modeling. The driving force behind this importance research field is the application it finds mainly in biological, ecological and social systems.

This extraordinary advance, I dare to say, is mainly due to the computational capacity at our disposal nowadays. The progress in the development of hard- and software as well, has enabled us with the possibility of dealing with thousands and hundred of thousands equations that appear frequently in the solution of complex systems. As a matter of fact, computational modeling, considered in its broad spectrum, is growing very rapidly, combining the contribution of natural scientists, mathematicians, engineers and social scientists to the solutions of complex problems. This mixing of people coming from different areas of knowledge is creating new interdisciplinary areas that will probably be of fundamental importance in the near future, if not right now.

This paper presents a model that is intended to help understanding the process of knowledge transfer among groups belonging to a certain segment of the same society or possibly the interaction process between different societies. The first version of this paper was published a long time ago, in 1980 [1]. After that, no effort was made to elaborate the theory or test the theory against some empirical data. A relatively recent paper was encouraging and has shown that similar ideas were being explored. This reference [2] considers both genotype and phenotype factors to study the evolution of social groups. That is, inheritance and social interaction are both shaping the history of the group. We are interested here in a relatively short period of time, and the horizontal aspect, interaction among individuals or sub-groups constitutes the predominant factor.

Despite the limitations of the model, the results are plausible and intuitively consistent. So, if the model doesn't reflect completely the reality nonetheless it allows for some enlightenment that puts in evidence important and some times hidden correlation among concrete facts. If taken judiciously the results could help decision makers to implement important actions to foster the generation and transmission of knowledge. It is never too much to insist that social modeling, where the behavior of human being play a key role should be seen very carefully. It is impossible to model perfectly the behavior of human social groups and frequently we are faced with big surprises. It is a system that admits contradictory outcomes, where adaptation and rebellion walk side by side.

In the next sections it is presented a temptative approach to the process of transfer of knowledge. Certainly the model is a simplification but as a first approximation suggests very interesting correlation, particularly the relative importance of knowledge generation, knowledge transmission and learning speed.

3. An Unidimensional Diffusion Problem

Let us consider a mechanical problem that it is assumed to be similar to the knowledge transfer. That is, it contains the same basic characteristics of diffusion and evolution that is found in the process of knowledge transfer. So we start with a mechanical problem that is artificial in some aspects in order to better adapt to the process of knowledge transfer. As a matter of fact the formulation as it is needs to assume the existence of external sources of energy and mass in order to meet the law of mass conservation and the second law of thermodynamics. We assume that all the physical requirements to make the formulation consistent are satisfied.

On the other hand the physics of the problem is simple enough to be understood by a non expert. We will try to appeal as much as possible to the intuition of the reader. The mathematical deductions are skipped and only the relevant results are presented and commented.

Consider a finite rod, of length $l = 1$, with a non-uniform cross section A, whose area, and consequently the mass per unit length, varies along the rod as a function of x
(Fig 1). Let us assume further that A is also a function of the temperature T. Along the axis of the rod there are heat sources Q whose intensities are also function of x and the temperature T. Consistent with these hypothesis assume:

i) $A(x, T) = A_0(x)(1+\alpha T)$ (1a)

ii) $Q(x, T) = Q_0(x)(1+\alpha T)$ (1b)

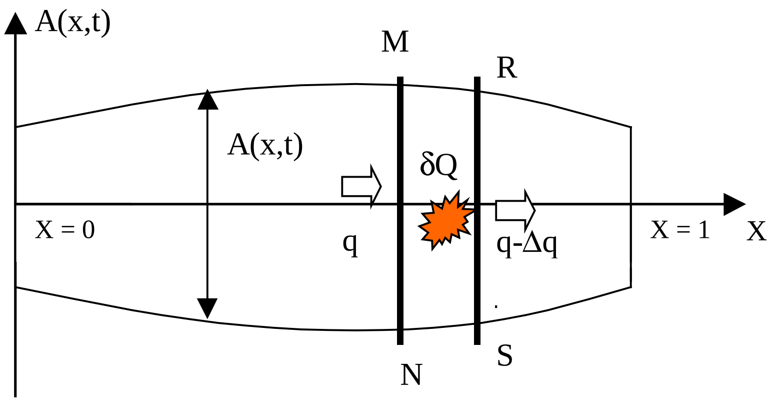


Figure 1. Mass distribution profile.

Of the rod there are heat sources Q whose intensities are also function of x and the temperature T. Consistent with these hypothesis assume:

iii) $A(x , T) = A_0(x) (1+\alpha T)$ (1a)
iv) $Q(x , T) = Q_0(x) (1+\alpha T)$ (1b)

where A is the area of the cross section and Q is the heat source intensity distributed along the x axis, T the temperature and α a constant. Note that the dimension of α is the inverse of temperature, $1/^{o}C$, and therefore αT is a non-dimensional quantity. That is we are postulating that the cross section varies linearly with the temperature and that heat is generated also proportionally to the temperature. It goes without saying that all the energy and additional mass required to meet the laws of conservation are assumed to be fulfilled somehow.

Briefly, we have a rod where heat is generated and mass is added, in both cases proportionally to the temperature and there is an external energy source that allows for the accomplishment of all the physic laws. More precisely we should say that heat and mass vary, since it is possible for both to diminish if the temperature falls down.

The mathematical formulation of the heat conduction can be put into a differential form representing the heat flow balance across an element MN-RS displayed in the figure 1, according to the law of physics. Essentially, given an interval of time Δt, the following relation must hold:

QUANTITY OF HEAT RETAINED WITHIN MN-RS (Δq)	QUANTITY OF HEAT GENERATED INSIDE MN-RS (δQ)	QUANTITY OF HEAT NECESSARY TO INCREASE THE TEMPERATURE OF MN-RS OF ΔT

Now it can be easily shown that following the classical law of heat conduction, and after some manipulations the balance equation written above leads to:

$$\frac{1}{\theta}\frac{\partial}{\partial x}\left(\frac{k}{\theta}\frac{\partial\theta}{\partial x}\right)+\alpha C_0\theta = C_p\frac{\partial\theta}{\partial t} \tag{2}$$

This is the governing equation of heat conduction for a rod with variable cross section and heat sources distributed along the longitudinal axis according to the assumptions i) and ii). It has been written with the main purpose of clarifying the physical meaning of the parameters involved. The variable θ is equivalent to a non-dimensional temperature:

$$\theta = 1 + \alpha T$$

It measures indirectly the energy at each cross section relatively to some reference estate The other parameters have the following meaning:

k: Thermal conductivity. It is the parameter that measures the "material permeability" to heat. The greater the parameter k the easier will be for the heat to flow, from one cross section to the next, along the rod.

C_0: The heat source density along the rod for T=0. It represents the intensity of heat generated inside the rod. The thermal energy generated in the rod is proportional to this parameter.

C_p: Specific heat. It is a kind of thermal inertia or impedance. Large C_p means that the heat necessary to increase the temperature of a given volume of material of 1 degree Celsius is proportionally large.

Since the purpose of this paper is to investigate, whether or not the proposed model is able of giving meaningful response to simple cases of knowledge transfer we will impose some simplifications in the equation (2) at this stage. So suppose that all the parameters, k, $C_{0,}$ C_p and α are constants. Note also that the equation (2) is non-linear but can be easily transformed into a linear one. Taken into account these observations, the above governing equation can be written in the form:

$$\frac{\partial^2 u}{\partial^2 x}+bu = c\frac{\partial u}{\partial t} \tag{3}$$

where:
$u(x,t) = \theta^2(x,t)$ is the new variable

$$b = \frac{2\alpha C_0}{k} \quad and \quad c = \frac{C_p}{k}$$

are new parameters.

This is a much simpler equation whose solution is well known.
The complete formulation of the problem requires however the specification of two more important conditions.

A- TEMPERATURE SPECIFIED AT BOTH ENDS

The so called *boundary conditions* that defines how the temperature or the heat flow behave at both extremities of the rod and the *initial conditions* that defines the temperature distribution along the rod at the initial time t=0. Of course these conditions are not uniquely defined, they depend on the problem. Suppose that for a particular problem the extremities of the rod are kept at constant temperature by some artificial device. Translated into mathematical terms this reads:

$$\text{Boundary conditions}: \begin{cases} u(0,t) = \theta^2(0,t) = (1+\alpha T(0,t))^2 = u_0 \\ u(1,t) = \theta^2(0,t) = (1+\alpha T(1,t))^2 = u_1 \end{cases}$$

That is, in this case the particular conditions of the problem imposes the temperature or the thermal energy level to be kept constant at both ends (x=0 and x=1) equal to u_0 and u_1 at x=0 and x=1respectively. Note that the new variable *u(x,t)* is related to the initial variable, the temperature *T(x,t)*, according to the transformation previously defined, in order to give the proper expression for the boundary conditions. Similarly the initial condition that specifies the temperature distribution along the rod trough the same transformations ends up with the following relation:

$$\text{Initial contidtion}: \quad u(x,0) = \theta^2(x,0) = (1+\alpha T(x,0))^2 = f(x)$$

where the function *T(x,0)* stands for a prescribed temperature distribution along the rod at time t=0 which defines automatically the function $f(x) = u(x,0)$ related to the new variable $u(x,t)$.

In summary the complete formulation of the problem reads:
Solve the equation:

$$\frac{\partial^2 u}{\partial^2 x} + bu = c\frac{\partial u}{\partial t} \tag{3}$$

Subjected to the conditions:

$$\text{Boundary conditions}: \begin{cases} u(0,t) = u_0 & (3\text{-a}) \\ u(1,t) = u_1 & (3\text{-b}) \end{cases}$$

$$\text{Initial conditions}: \{ u(x,0) = f(x) \tag{3-c}$$

This is a very well known problem in the theory of partial differential equations. We are not going into the details of the solution. The main purpose of this paper is to discuss the dependence of the solution on certain parameters and on the boundary conditions. The reader interested in the mathematical treatment of partial differential equations of the evolution type can refer to [3], [4] . So some of the steps in the sequel will not be proved in order to keep the attention concentrated on the main focus of this discussion.

It can be shown that the solution of the equation (3) together with the subsidiary conditions (3-a, b, c) can be stable or unstable depending on the value of the parameter b in (3). That is, there is a critical value $\mathbf{b_{crit}}$ for b above which the solution grows without any limit as $t \to \infty$. We may write more concisely:

For $b > \mathbf{b_{crit}}$ $\Longrightarrow$ $u(x, \infty) > M$ M: an arbitrary positive real number
For $b < \mathbf{b_{crit}}$ $\Longrightarrow$ $u(x, \infty) < u_{lim}(x)$

Where $u_{lim}(x)$ is a curve that depends on b, c and the subsidiary conditions. That is, for b less than the critical value the solution is bounded as $t \to \infty$. Recall that "b" is proportional to the heat source intensity C_0 and inversely proportional to the conductivity k or the capacity to transmit heat from one cross section to the next. So, for a given value of C_0 if k is sufficiently small, b will be greater then the critical value and the solution will grow steadily with time. The same is true for a fixed value of k and C_0 sufficiently large. This means that there is no use in increasing the transmission capacity if there is no adequate energy generation. On the contrary, if the heat source distribution is weak, in order to have a growing solution, the transmission has to be kept below a certain upper bound.

The figure 2 shows the dependence of the solution u(x,t) on the initial conditions, for a value of "b" smaller than the critical. It is clear from the figure that in both cases the solution tends to a limit curve.
After a sufficiently large time interval t_1 the solutions for both initial conditions become undistinguishable and very close to the limit curve. Note that even for an initial condition with a high energy level (I.C.- 1) as shown in the figure the solution will decay till the limit curve is reached. The heat sources are not powerful enough to promote the growth of the temperature, it is to say of the growth of the mass. But if the initial condition is below the limit curve (I.C.- 2) then the heat generation will be enough to increase the mass towards the limit curve.

Note that the stability of the solution doesn't depend on the parameter C_p at all. This parameter controls the speed of growth or decay, but doesn't come into play in the stability of the solution.

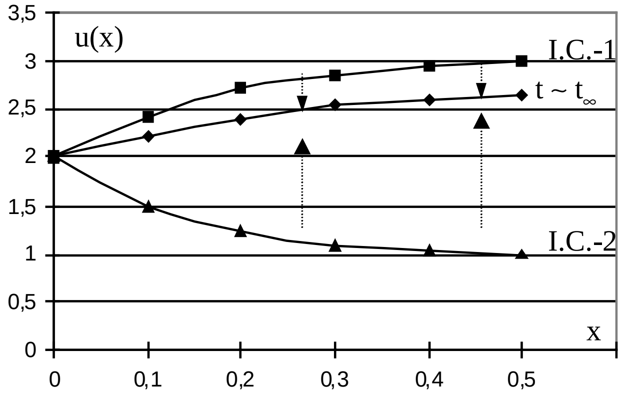


Figure 2. Evolution of the profile u(x,t) for different I.C. Temperature specified at both ends. $b < b_{crit}$.

B- HEAT FLUX SPECIFIED AT BOTH ENDS

Another possible definition of the boundary conditions has to do with the heat flux at the ends of the rod. Instead of specifying the temperature it is also possible to specify the rate of heat flow with respect to time at the ends. That is, the quantity of heat coming in or leaving out the rod through the ends per unit time. It is well known that the heat flux depends on the temperature gradient. So specifying the heat flux is the same as specifying the temperature gradient. From the previous definition of the independent variables coming into play in our problem we may then write:

$$\frac{\partial u(x,t)}{\partial x} = 2\alpha(1+\alpha T(x,t))\frac{\partial T(x,t)}{\partial x}$$

from which follows:

$$\frac{\partial T(0,t)}{\partial x} = \frac{1}{2\alpha(1+\alpha T(0,t))}\frac{\partial u(0,t)}{\partial x} \qquad \text{for x=0}$$

$$\frac{\partial T(1,t)}{\partial x} = \frac{1}{2\alpha(1+\alpha T(1,t))}\frac{\partial u(1,t)}{\partial x} \qquad \text{for x=1}$$

or recalling the definition of u(x,t) in terms of T(x,t) we get:

$$\frac{\partial T(0,t)}{\partial x} = \frac{1}{2\alpha\sqrt{u}} \left.\frac{\partial u}{\partial x}\right|_{x=0} \tag{4- a}$$

$$\frac{\partial T(1,t)}{\partial x} = \frac{1}{2\alpha\sqrt{u}} \left.\frac{\partial u}{\partial x}\right|_{x=1} \tag{4- b}$$

If we specify the left hand side terms in (4-a, b) then the equivalent boundary conditions for the differential equation in terms of u(x,t) is a combination of the rate of variation of u(x,t) and the value of u(x,t) itself. For sake of simplicity and since this is an exploratory analysis we will prefer to specify the derivatives of u(x,t). In this case we will have a mixed boundary condition for the variable T(x,t). The problem to be solved is therefore:

$$\frac{\partial^2 u}{\partial^2 x} + bu = c\frac{\partial u}{\partial t} \tag{5}$$

Boundary conditions

$$\frac{\partial u(0,t)}{\partial x} = 2\alpha(1+\alpha T)\left.\frac{\partial T}{\partial x}\right|_{x=0} = D_0 \tag{5-a}$$

$$\frac{\partial u(1,t)}{\partial x} = 2\alpha(1+\alpha T)\left.\frac{\partial T}{\partial x}\right|_{x=1} = D_1 \tag{5-b}$$

Initial condition

$$u(x,0) = (1+\alpha T(x,0))^2 = f(x) \tag{5-c}$$

So the solution of the system becomes simpler but the boundary conditions in terms of T(x,t) need some explanation. The boundary conditions can be considered as "valves" which control the heat flux at both ends. Given D_0 , for instance, this means the "valve" at x = 0 will let the heat flow in or out, depending on the sign of D_0 , minus or plus respectively, at a time rate that is inversely proportional to the temperature.

It is a sort of control on the flux that reduces the flow rate for high temperatures as if it was advantageous to retain the flux going out for high energy levels at the ends, or in the opposite case, that prevents the admission of heat coming in, also if the temperature is high.

For the system (5) plus (5- a, b, c) the stability analysis is more difficult. The critical value of b doesn't determine so neatly the stability regions as in the case of the previous boundary conditions. We can only say that for b greater than the critical value given before the solution increases monotonically as $t\to\infty$. That is:

For $b > \mathbf{b_{crit}}$ $\Longrightarrow$ $u(x, \infty) > M$ M: an arbitrary positive real number

But if b is smaller than the critical value the solution is not always bounded. It will depend on the boundary conditions. That is, for the present case the boundary conditions play a decisive role in the stability of the solution together with the parameter b. We say that stability is sensitive to the boundary conditions. One important characteristic of this system is that for unstable solutions, that is if u(x,t) grows without limit, it tends to homogenize along the x-axis approximating a horizontal straight line that continues to increase in time. All the sections tends to the same value as $t\to\infty$. The figure 3 displays this behavior clearly.

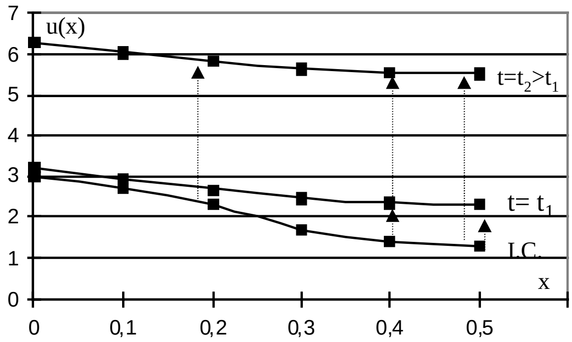


Figure 3. Evolution of the profile u(x,t), b and c constants. Heat flux specified at the ends, $b<b_{crit}$.

As a matter of fact stability occurs only under very special conditions. This discussion is beyond the scope of this paper. But deviating from the previous case, where the temperature was supposed to be specified at the ends, it is possible to have instability in the opposite sense. That is, under certain conditions the solution can decrease indefinitely, $u(x,t) \to -\infty$ as $t \to \infty$. The figure 4 shows this type of instability. For both cases all the parameters are equal, the difference being in the boundary and initial conditions. For the case I.C.- 1, the flux at the ends is very high and the continuous flow of heat to the exterior of the rod is not compensated by the heat source inside the rod. The temperature decreases very rapidly at the ends and reaches values less than 0 in a region close to the ends which is meaningless in terms of T(x,t). For the case I.C.-2 however, the heat flux to the exterior of the bar is moderate and the heat source inside the rod compensate this loss and the solution grows.

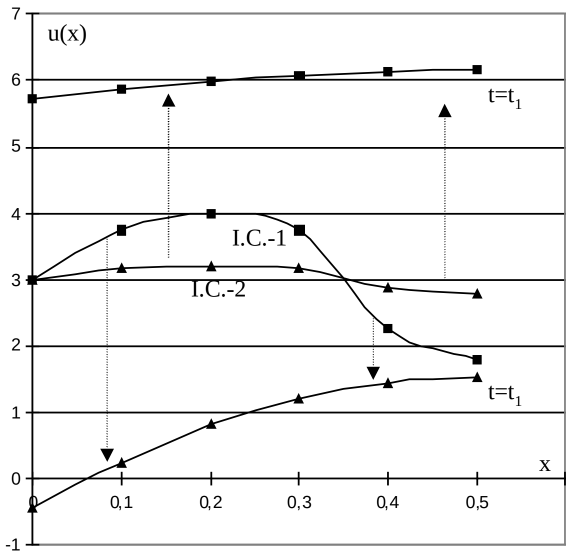


Figure 4. Evolution of the profile u(x,t) for two different I.C. b and c constants and equal for both cases $b < b_{crit}$.

It is important to remark that in order to have real values for T(x,t) the function u(x,t) must be positive. If u(x,t) becomes negative the solution is imaginary. We interpret imaginary solutions as vanishing of mass, that is the destruction of the segment of the rod where $u(x,t) < 0$. We remark that for homogeneous boundary conditions that is for $D_0 = D_1 = 0$ the solution always increases. This an expected result, since if the ends are insulated the energy produced by the heat source inside the rod will keep the mass growing indefinitely.

4. Model Generalization for Variable Conductivity

We present now, briefly, a generalization of the model for the case where k is a function of x. For $k = k_0\, p(x)$ where $0< p(x) < 1$ is a continuous function of x, the governing equation reads:

$$\frac{\partial}{\partial x}\left(p(x)\frac{\partial u}{\partial x}\right) + b^{*} u = c^{*}\frac{\partial u}{\partial t} \qquad (6)$$

To see the influence of the variation of the conductivity with the space variable x we will solve some simple problems with the temperature specified at the ends and the initial condition a given function of x as shown in the figure 5.

Boundary conditions: $u(0,t)= u_0 = 4$ $u(1,t) = u_1 = 4$

Initial condition: $u(x,0) = f(x)$ (as shown in the figure 5)

Consider three rods with the following mechanical properties, symmetric with respect to the middle section:

Rod 1 : b*= 6 and c*= 4 both constant along the rod ; p(x) = 1

Rod 2 : p(x) = 1 constant; b* , c* varying with x as shown in the table 1 below.

Table 1

x	.0	.1	.2	.3	.4	.5
b*	8	7	6	3	1	1
c*	2	2	4	6	8	8

Rod 3: p, b* , c* varying with x as shown in the table 2 below.

Table 2

x	.0	.1	.2	.3	.4	.5
b*	8	7	6	3	1	1
c*	2	2	4	6	8	8
p	1	.9	.75	.6	.4	.2

The respective solutions for a given initial condition are shown in the figure 5. The solution for the first rod evolves more rapidly. This is due to the fact that the specific heat or impedance c* in the middle region is lower for the first rod. The speed of growth of the solution depends essentially on this parameter. Comparing the second and the third rod it is seen that the conductivity decreases from the ends towards the center for the third rod while remains constant for the second. This variation of k produces a delay in the mass evolution at the middle section. At the regions close to the ends however where the third rod displays a relatively bigger conductivity the evolution is more rapid for the third rod.

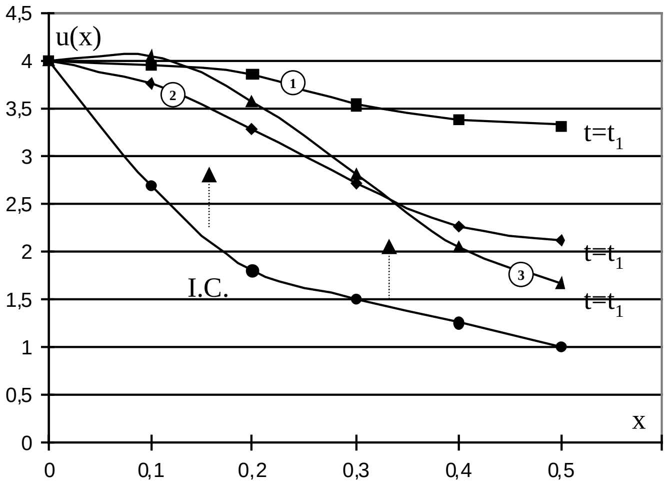


Figure 5. Comparison of the evolution profile u(x,t) for different b(x), c(x) and k(x).
Curve 1 – b, c and k constants; Curve 2 – b(x) and c(x) varying with x, k constant; Curve 3 – All parameters, b(x), c(x) and k(x) varying with x.

Consider now a fourth rod with the following properties, also symmetric with respect to the middle section:

Rod 4

Table 3

x	.0	.1	.2	.3	.4	.5
b*	8	7	6	3	1	1
c*	2	2	4	6	8	8
p	.2	.4	.6	.75	.9	1.

Compare the solutions for this two rods, for the same initial condition (Fig.6). It is remarkable the speed of evolution near the ends displayed by the profile corresponding to the fourth rod. Noting that the only difference between the two rods is the conductivity variation along the x - axis, it is possible to say that the difference between the two solutions is due to the conductivity distribution. Apparently, despite the fact that the conductivity in the last case near the end is low, the high values of the conductivity in the middle section allows for a better exchange of heat along the rod.

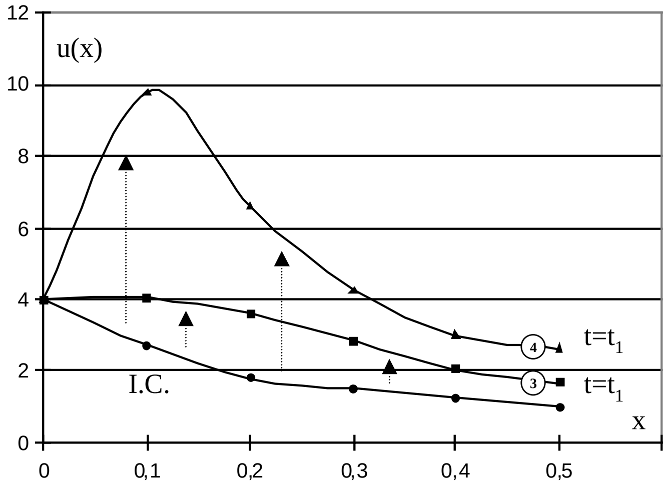


Figure 6. Comparison of the evolution profile for different variation of k(x) with x.
Curve 4 – k(x) maximum at the center x=5
Curve 3 – k(x) maximum at the extremitis x=0, x=1

5. A Model for Knowledge Transfer

Is it possible to transfer the model presented in the last sections to the process of knowledge transfer ? This question has no definitive answer yet. Certainly the material presented in the previous sections is rather simple to represent such a complex phenomenon as the transfer of knowledge. Nonetheless some interesting hints emerge from the comparison between the heat transfer problem and what we could expect from the process of knowledge transfer.

It is plausible to make the following parallels:

i) The temperature could be considered as a measure of the intellectual productivity or some other activity depending on the knowledge capacity.

ii) The conductivity is the means available to transfer the knowledge and the willingness of the individuals in the group to teach and learn.

iii) The heat generation is obviously the creativity of the individuals in the group, or the average creativity distributed among the individuals according to some criteria.

iv) The specific heat can be compared with what we will call cognitive impedance. It is related to the learning capacity. The lower the impedance the quicker the individual learn.

v) The cross section or the mass per unit length can be associate to the number of individuals belonging to a sub-group in the knowledge chain, or better in the knowledge application or use chain.

Consider a "knowledge chain" constituted by the various segments of a group having in one end the scientific production and in the other the final industrial production. The intermediate segments are applied science, technological development, basic engineering, design, fabrication process, management, etc. If we compare now the results obtained in the previous sections with the correlation above we may come to the following conclusions:

A. The growth of the group depends on the creativity not on the cleverness or the capacity of learning quickly.

B. There is no use to provide excellent mechanisms for learning and teaching, accelerating the flow of knowledge between adjacent segments if the creativity remains low. Better means for knowledge transfer are only meaningful if the creativity is high.

C. If the creativity remains below a certain level for a given capacity of transferring knowledge, no matter how clever the individuals are, the growth will be always limited if the production is constant at both ends, there is no control on the flux of knowledge at the ends. Keeping the production constant at both ends is equivalent to freezing the number of individuals active at both ends.

D. The cleverness of the individuals or the average cleverness of the group, is effective only for the growth rate with respect to time, it has nothing to do with limits of expansion.

E. In general the specification of the flux of knowledge at both ends is more adequate to induce the progress of the group than the production specification. But if the flux is very high in the sense that a great quantity of information leaves the group at both ends in a short period of time, it is possible for the group to collapse.

F. A variable mechanism of knowledge transfer between segments is an important tool to control the evolution profile of the group. Further investigation is needed.

G. If the heat flux to the exterior of the rod through the ends is interrupted, at both ends, and there is creativity, not matte how small, the group will always grow.

These statements are plausible and give some indications on how to improve the model. It is striking the conclusion that what determines the unstable behavior towards an unbounded progress is the creativity and in some cases the boundary conditions, but not the cleverness, defined as the capacity to learn rapidly.

6. Conclusion

The parallel between a classical diffusion problem and the transfer of knowledge process can be established in our opinion. Of course the theory exposed here needs to be expanded to consider other variables and other interactions. The results are encouraging and we think that this approach deserves further efforts.

It is difficult, but not impossible, to device some field experiments or data collection
To measure some of the parameters. Finally this type of interdisciplinary study needs to incorporate people from different areas of knowledge.

We expect this paper to open the discussion about transfer and knowledge production, in order to help the decision makers to find better policies for our educational and research system.

7. Acknowledgments

This research has been partially supported by the “Fundação Carlos Chagas Filho” – FAPERJ – through the program “Cientistas dos nosso Estado” .

8. References

[1] L. Bevilacqua, 1980, "Transferência de Tecnologia: Meta ou Mito"; Anais I SIBRAT, Salvador, Baía, pp. 619-626
[2] L. Cavalli-Sforzza et alli,, 1993, "Evolution of Social Groups "; Research Report, Santa Fe Institute.
[3] F. John,1990, "Partial Differential Equations"; Springer Verlag, New York.
[4] M. Braun, 1993, "Differential Equations and their Applications", Springer Verlag, New York.

HOW TO HAVE A BETTER INTEGRATION BETWEEN UNIVERSITIES AND COMPANIES

Salej, Stefan D.
FIEMG – Federação da Industrias do Estado de Minas Gerais.
Brazil

MECHANICAL CIRCULATORY ASSIST: FROM DISPLACEMENT TO ROTARY BLOOD PUMPS

Reul, Helmut
Helmoltz-Institute for Biomedical Engineering
RWTH Aachen - Germany

ROBOTICS IN SURGERY

Slade, Alan
Medical Engineering Research Institute
University of Dundee - Scotland

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

APPLICATION OF NON INVASIVE METHODS IN MEDICINE AND IN ENGINEERING

Tomasini, Enrico
Dipartimento di Meccanica
University degli Studi di Ancona - Italia

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

MICROFLUID DYNAMICS

Lysenko, Olga
Bielorussian Academy of Science
Bielorussia

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

The Role of Materials' Design in Engineering Design Process of Products and Their Elements

Leszek A. Dobrzański
Institute of Engineering Materials and Biomaterials, Silesian University of Technology, Konarskiego St. 18A, 44-100 Gliwice, Poland
ldobrzan@zmn.mt.polsl.gliwice.pl

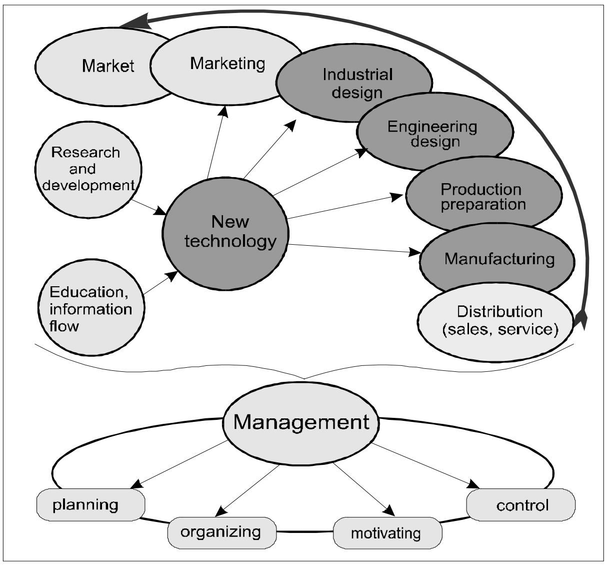
Abstract: *An important role of materials' design is stressed in the paper as of one of indivisible elements of engineering design of products and their elements. The particular attention is paid to the interrelation of materials' selection with their working conditions. Significance of materials' costs is stressed in conjunction with their properties, indicating that only in cases when the material's cost features a dominating portion of the total product cost, this selection criterion plays a main role. A particular attention is paid to the significance of materials' design and the need of dissemination of pertinent knowledge among many engineers carrying out engineering design, as the dynamical civilisation progress is forecast and observing that clearly the materials' development used to be one of the progress determining factors in the history to date. On the bases of observation of human civilisation development it was stated that civilisation development depends on materials development at a large extend. The paper presents the historical development of engineering materials since the beginning of the humankind and also the development and the role of heat treatment of metallic engineering materials is presented on this background. The particular attention is paid to the role of phase transformations occurring during heat treatment in forming of structure and properties of this group of materials, classifying also phase transformations occurring during heat treatment of the most common groups of metal alloys. The dominating development of surface treatment is stressed among the contemporary trends of the development of heat treatment technologies, advances in part body treatment and changes occurring in the heat treated metallic materials.*

Keywords: development of engineering materials, engineering design, engineering materials, heat treatment, history of civilization, manufacturing process, materials design, phase transformations

The general characteristics of the engineering design process of products

The manufacturing consists in making products from the raw materials in various processes, using various machines and in operations organized according to the well-prepared plan. Therefore, the manufacturing process consists in a proper use of resources like: materials, energy, capital, and people. Nowadays, manufacturing is a complex activity uniting people working in various professions and carrying out miscellaneous jobs using diverse machines, equipment, and tools, automated to a various extent, including computers and robots. The process of product manufacturing witch aim is to fulfill human needs is preceded by design. Design of the product is located between its marketing and manufacturing in the process of its introducing into the market (fig. 1) and

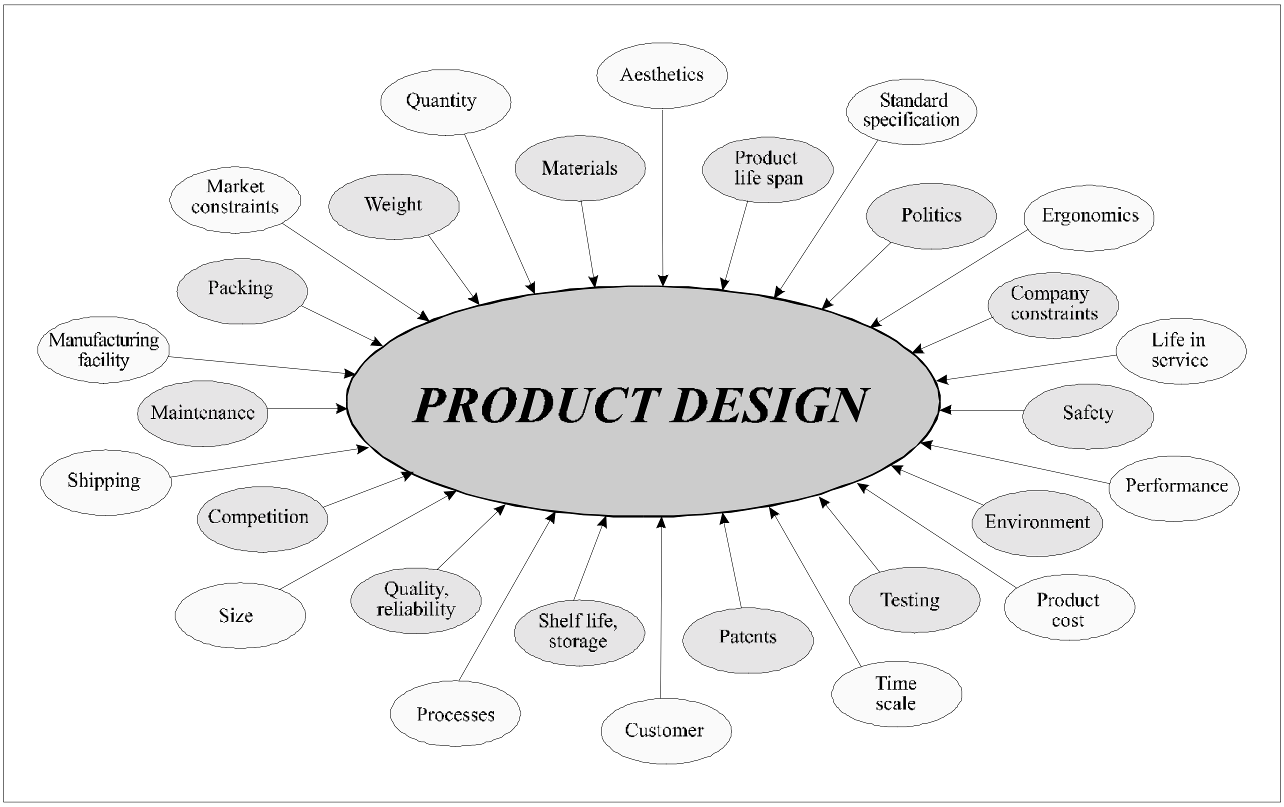
Fig. 1. Relations between factor connected with introducing a product to market



it is not a separated activity, as it influences all other phases of this process, on which it is simultaneously dependent. The first product design phase pertains industrial design, connected with the general description of the product's functions and with working out of its general conception, comprising only its outer shape, colour, and eventually, some general assumptions referring to linking its main elements. The succeeding phases include the engineering design and next production preparation. Engineering design is a complex activity requiring taking into consideration many diversified elements (fig.2). The engineering design merging in itself three equally important and indivisible elements, (fig.3) i.e.:

- structural design, whose goal is to work out the shape and geometrical features of products satisfying human needs,
- materials' design to ensure required physical and chemical, as well as technological properties of engineering materials, ensuring the expected life of the product or its elements, and
- technological design making it possible to impose the required geometrical features and properties to the particular product the elements, and also to ensure their correct mating after assembly, accounting for the production volume, its automation level and computer support, and also with ensuring the lowest possible costs of the product.

Fig. 2. Elements for consideration in a product design specification (prepared according to C. Nevay and G. Weaver)



The designed product has to meet the parameters pertaining fully to its functionality, and also requirements connected with its shape and dimensional tolerances, moreover, the design has to include the list of used materials, manufacturing methods, and other necessary information. Each product shape version imposes some requirements pertaining the material properties that can meet them, to which one may include the relationships between stresses resulting from the product shape and its load, and the material strength. The change of the manufacturing process may change the material's properties, and some product-material combinations may be infeasible using some technological processes. The first engineering design stage (fig.4) consists in a conception development, connected with a general specification of the available materials and processes. In the succeeding general project design stage, the shapes and approximate sizes of its elements are determined using the engineering analysis methods. The designer decides at this stage about a general class of the materials to be used and a type of a manufacturing process, e.g., selecting plastic forming or casting to make an element from the nonferrous metal alloys. Material properties should be defined more precisely at this time. At the detailed project design stage, material and the technological process are finally selected. One, appropriate material is chosen and only several variants of the technological process at the very most. It is connected simultaneously with making a decision pertaining dimensional

Fig. 3. Relationships between elements of engineering design i.e., structural design, materials' design and technological design (prepared according to G.E. Dieter)

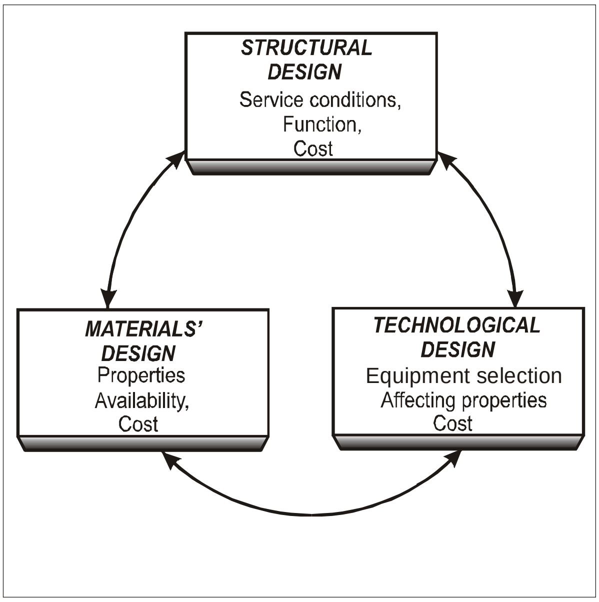
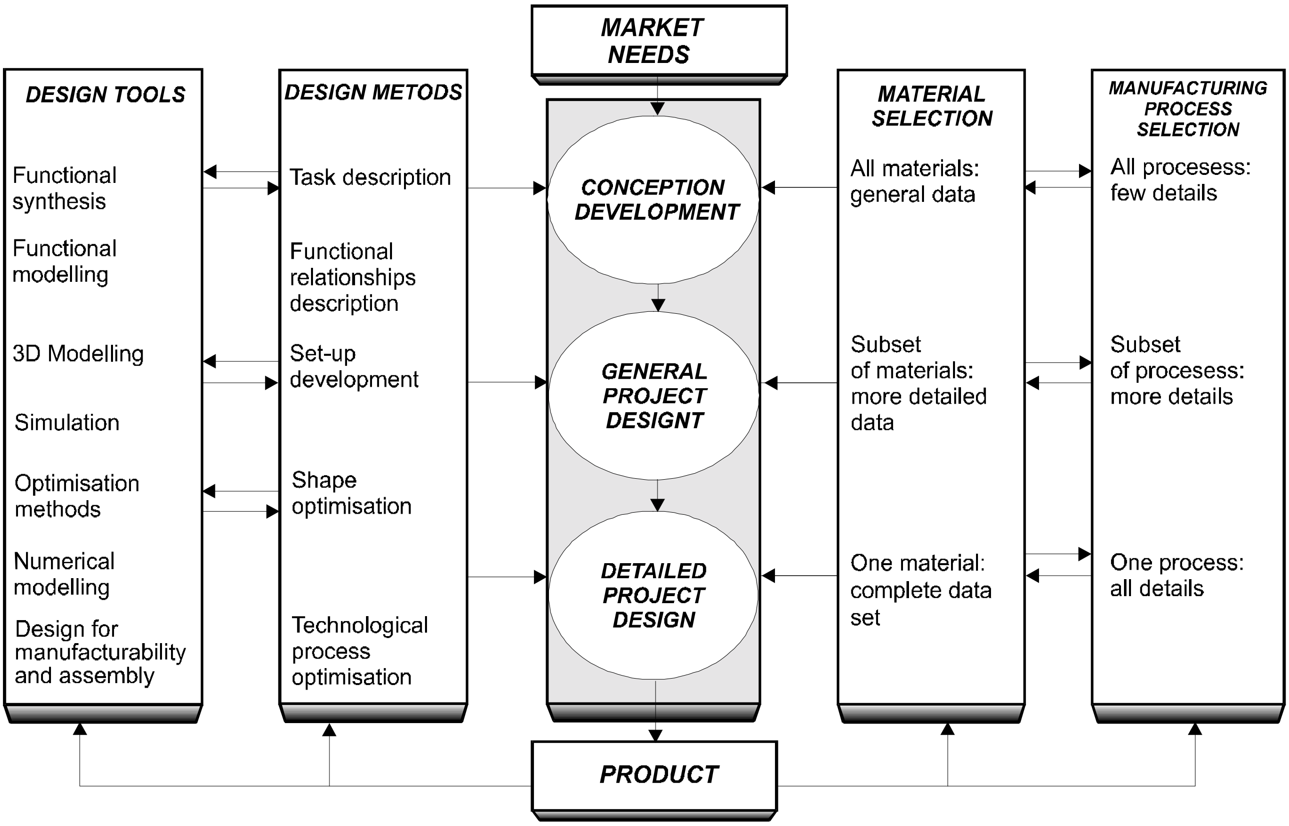


Fig. 4. Stages of engineering design (prepared according to the general idea of M.F. Ashby)



tolerances, stress state optimisation, and with selection of the best manufacturing process, employing the quality engineering methodology and costs simulation. The designer should be well versed in detailed materials' properties, depending on the significance of the designed element. Independently from the approach taken to design the product elements, one has to account for the effect of notches and stress concentration, as factors increasing their vulnerability to failures. Cyclic loading, service in high or low temperature, as well as the presence of media causing the general corrosion or cracks resulting from stress corrosion, all feature specific hazards that have to be considered in the material selection process. Relationships of the common failure types and mechanical properties of materials are shown in Table 1. The technological process selected and a possibility of using certain technologies, influencing engineering design, deciding materials' selection and a sequence of operations, and also dimensions, dimensional tolerances, mating elements' design and other aspects. Therefore, the design process requires accounting for many factors connected with the technological process. The main criterion for the selection of the product manufacturing processes is a maximisation of a product quality with a simultaneous minimisation of costs of its elements. Working properties of a product are obtained only when the right material is used, manufactured in the properly selected technological process, imparting both the required shape and other geometrical features, including dimensional tolerances of particular elements, making the final assembly of the product possible, and also forming the required material structure, ensuring the expected mechanical, physical and chemical properties of the product.

Main factors deciding material selection in engineering design of products

Variety of materials available nowadays, makes it necessary to select them properly for the constructional or functional elements, tools and eventually other products or their elements. This selection should be carried out basing on the multi-criterial optimisation, basing also on lot of properties of materials. A selection of a proper material along with the appropriate technological process is vital, as it ensures the longest product life with the lowest costs, considering that one has to account for more than 100.000 engineering materials possible and available on the market, and yet, an average engineer has detailed knowledge about practical applications of some 50-100 engineering materials. One can specify four stages in the material selection process for the chosen elements:

- specifying, basing on the criterial properties, whether the element is to be made from metal alloys, polymer, ceramic, or composite materials,
- in case of metal alloys, specifying whether the element will be made using plastic forming or casting, and in case of polymer materials, specification if the thermo-plastic or thermo-setting polymer will be used,
- selection limited to a strictly defined category of materials, e.g., in case of metal alloys, a specification whether the element is going to be made from the heat-treatable alloy constructional steel, heat-resistant steel, or from the aluminium alloy for plastic forming, and in case of polymer materials, specifying which of the thermo-plastics or hardening plastics will be used, e.g., polyester or polycarbonate.
- selection of the suitable material, with its designation or symbol.

Two first stages of the material selection process are commonly connected with the conceptual phase of the engineering design, that may also considered by the third stage, however, it is connected most often with the general design. Stage the fourth of materials selection pertains the phase of the detail design.

In case of a selection of materials for new products, one should observe the following procedure:

- define the functionality of the product and describe its functions with the required properties like, rigidity, strength and corrosion resistance, and economical coefficients, like cost or availability,

Table 1. Relationships between the failure type and materials' properties

Material properties / Failure type	Tensile strength	Yield point	Compression yield point	Shear strength	Fatigue properties	Workability	Break energy	NDT temperature	Modulus of elasticity	Creep rate	Brittle resistance	Resistance to corrosion embrittlement	Electrochemical potential	Hardness	Coefficient of thermal expansion
Loss of load capacity		●		●											
Buckling			●						●						
Creep										●					
Brittle fracture							●	●			●				
Low-cycle fatigue					●	●									
High-cycle fatigue	●				●										
Contact fatigue			●												
Fretting			●										●		
Corrosion													●		
Stress corrosion	●											●	●		
Electro-chemical corrosion													●		
Hydrogen embrittlement	●														
Wear														●	
Thermal fatigue										●					●
Corrosion fatigue					●								●		

● - this material property is useful for assessment of the given failure type

- define the requirements pertaining its manufacturing, defining the number of required elements, their size and complexity, required dimensional tolerances, finishing, general quality level, and general material manfacturability,
- compared the required properties and parameters with the extensive materials databases, the best of the computerised ones, to select preliminarily some materials that might be applicable, usually basing on the review of only some selected properties of the analysed materials with the extremum properties,
- inspect closer the preliminarily selected material, acquired from the retail network and used in the particular product, its cost, producibility and availability in the form and sizes necessary for the application,

Fig. 5. Strength and relative cost for a unit volume for various materials (prepared according to M.F. Ashby)

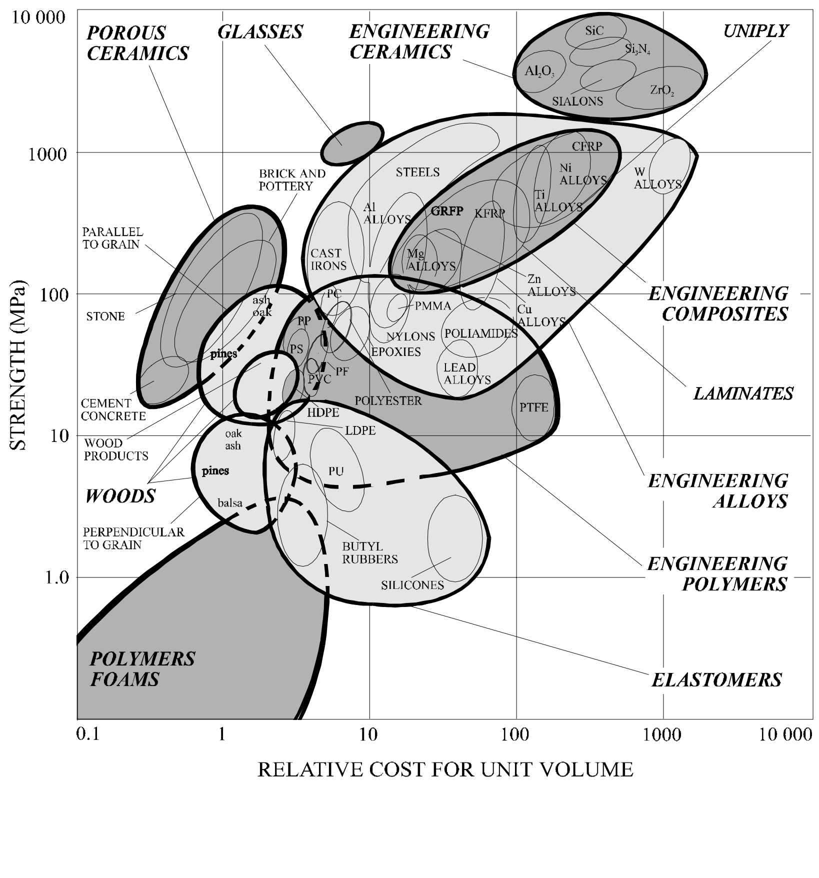


Fig. 6. Strength and specific energy consumption of various materials (prepared according to M.F. Ashby)

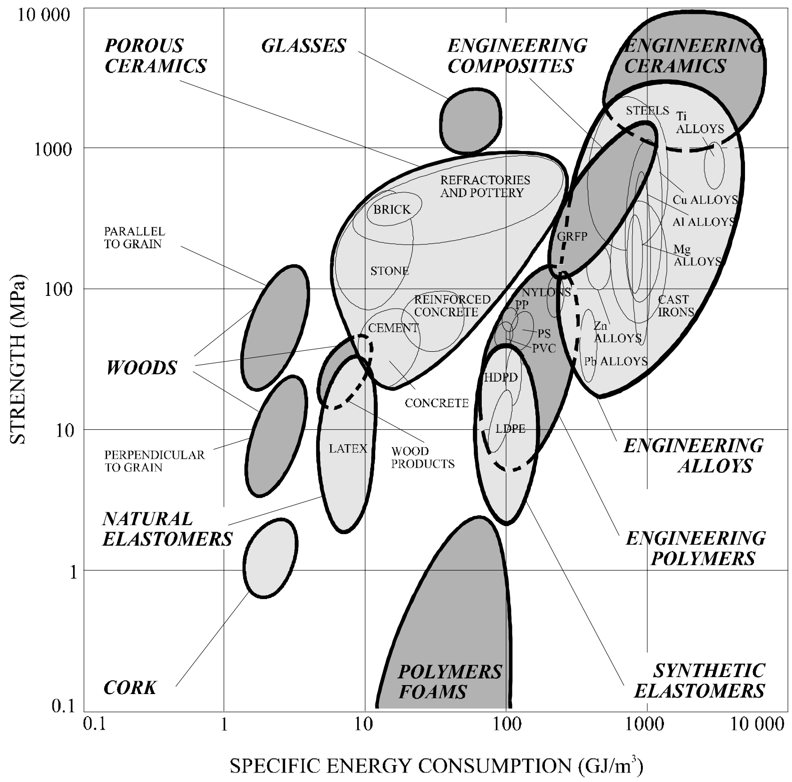
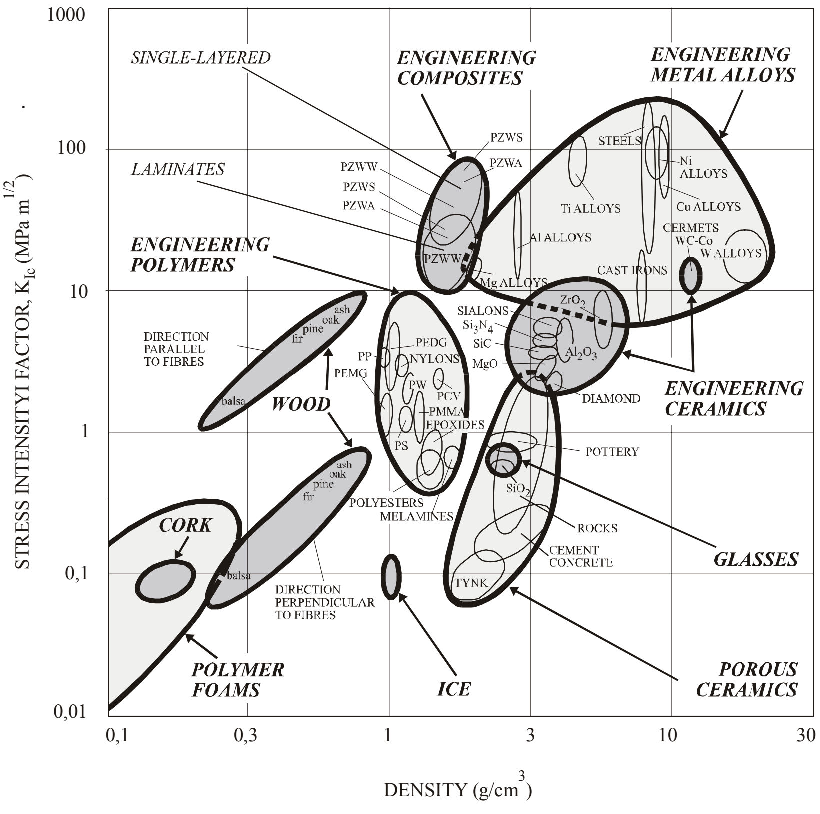


Fig. 7. Stress intensity factor and density of various materials (prepared according to M.F. Ashby)



- complete the design data, defining the minimum number of properties describing the particular material, and in case of more specific applications, like for the space or nuclear technology, carry out investigations according to the well developed programme, to obtain design data with the high statistical accuracy.

To change the materials in the existing product one has to use the following procedure:

- characterise the material used currently in its particular form, requirements of the technological process, and costs,
- specify, which of the properties have to be improved to extend the product functionality, with the particular attention paid to failure analysis premises,
- look for the alternative materials and technological processes, using the advantages review method,
- set up a short list of materials and applicable technological processes and compare the costs of the manufactured elements, using the engineering value analysis, e.g., to which every material, element, and technological process should be subjected, in which a negative answer confirms the value of the particular factor,
- develop results obtained in the previous step and indicate the alternative material, defining its critical properties, with the specification or investigation of materials for the specific applications, like in the case of materials selected initially.

To the main factors deciding materials selection one may include: functional requirements and limitations, mechanical properties, form, availability, timeliness of deliveries, alternatives, possibility of manufacturing, resistance to corrosion and degradation, stability, producibility, specific properties, aesthetical issues, economical criteria (material prices, costs of acquiring, processing and operation), opportunity criteria (fashion, political preferences, private links). Generalising, one has to reckon the following criteria as the most important ones for material selection: material quality and modernity, price, availability (timeliness of deliveries). Timeliness of deliveries features a criterion that is more important than price sometimes. The important criterion of materials selection is the manufacturing cost for the high quality element. The following factors have to be referred to in a selection of the best materials, when manufacturability of the elements is considered: material type and chemical composition (type of alloy, polymer, ceramic, or composite material), material form (bar, pipe, wire, strip, sheet, plate, powder, etc.), size (dimensions and dimensional tolerances), heat treatment condition, anisotropy of mechanical properties, surface treatment, quality (structure, non-metallic inclusions, etc.), production volume, producibility (machinability, weldability, castability, etc.), recyclability, material cost.

The material costs factor must be certainly taken into consideration at a selection of the material for various products. Figure 5 shows a rough comparison of the relative costs of the unit volume of material (defined as the ratio of 1 kg of the material and 1 kg of the bar made from the medium-carbon steel, multiplied by the material's density) versus strength. There is a possibility of attaining the economical effects when manufacturing highly processed products of the highest possible technical level, and also, that in these cases the specific cost of material may be even 100,000 times lower than the specific cost of the entire product. The apparent saving is undoubtedly unjustified in these cases. On the contrary, just then one has to employ materials having the best properties and reliability from all available ones. Fig. 6. illustrates the relation of strength and the specific energy consumption of materials (defined as the product of energy required to make the material, i.e., obtaining the raw materials, their refining, and shaping of the produced material, related to 1 kg of the material, and its density). This coefficient expresses indirectly the influence of the material's manufacturing process on degradation of the environment. The specific energy consumption shows linear dependence with the material's strength. It is expected that the levels of economical profitability for employment of particular raw materials be determined by their half-periods of depleting of their respective stocks. The character of ductility and fracture toughness changes (measured by the stress intensity factor) differs from changes of strength (fig. 7.). The highest ductility is demonstrated by metals and their alloys.It seems that their common use is owed to the compromising merging of the highest possible ductility with the very high strength. Composite materials demonstrate similar properties. However, the definite brittleness of the engineering ceramics features a serious limitation for its use.A comparison of other mechanical, technological, and working properties of various engineering materials, as well as an analysis of the ecological aspects connected with the use of various engineering materials was presented in previous publications.

The development of the engineering metallic materials and their heat treatment throughout the ages

Development of materials and accompanying growth of the production forces governed notably progress of human civilization (fig.8). Only when proper materials had become available practical applications of many inventions has been made possible.

Fig. 8. Diagram presenting the significance of various epochs of the human civilisation development, with dates of introduction of new materials (prepared according to M.F. Ashby)

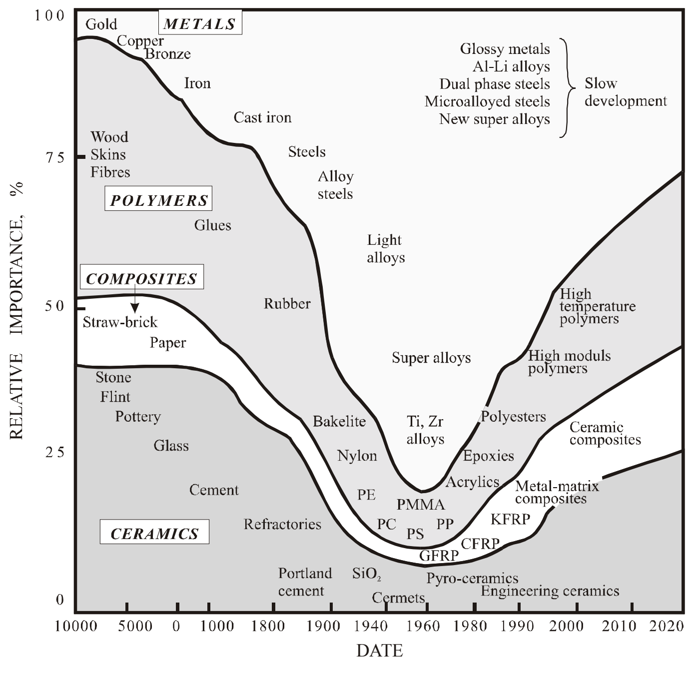


Fig. 9. Fragment of a picture on the western wall of the Redmere tomb in Thebes (c. 1,450 B.C.) showing metal processing scenes in ancient Egypt (charge preparation, melting, casting, finishing)

The pre-historic man might use only the natural raw organic and inorganic materials, including, e.g., leather, wood, rock, flint, that he processed into the useful articles making it possible for him to get food, improving his safety and living conditions. Up to now it is hard to decide which metal was processed by humans first. It is sure that metals occurring in their raw native states, like gold (Near East, Caucasus, Egypt – Nubian desert, the Eastern desert), silver (north-east Asia Minor, the district associated with Hittites), copper (Asia Minor, Armenia, Elam, from which the Sumerians received it as early as 3,500 B.C., Eastern Alps, Egypt up to c. 2,000 B.C., Cyprus) and iron obtained from meteorites (Greenland, utilized by Eskimos for more than a century). The significant progress was made only after mastering the methods of obtaining metals from their ores. This has happened – most probably by chance – c. 4,000 B.C., in pottery manufacturing, during glazing with the hot flame and pulverized minerals. Since that time, apart from the attempts to find the metals in their raw native forms, utilization of ores is carried out including two independent processes, i.e., a separation of metals from other chemical elements with which they are chemically bounded and working-up of metals into useful articles. The working properties of articles made from metals were inferior to those of stone tools and a landmark was manufacturing of copper alloys – bronzes, with arsenic at first, and next with tin. Traces of getting of copper ores and its processing were found in settlements of farmers and herdsmen in valleys of Kura and Araxes rivers in Transcaucasia, in the eastern Anatolia, Assyria and Mesopotamia, as early as in 4,000 and 5,000 B.C. In 4,000 B.C. the Cu-As bronze was used quite consciously, not by chance, that was later – at the turn of 3,000 and 2,000 B.C. replaced by the Cu-Sn bronze. At the cemetery of the Sumerian kings of the Ur state that reined at c. 2,600 B.C. decorative articles were get out made from bronze containing both As and Sn. Also in Central Asia by Indus river and in Europe, as

Fig. 10. Picture on a bowl from Troia (V century B.C) showing a Troian forging

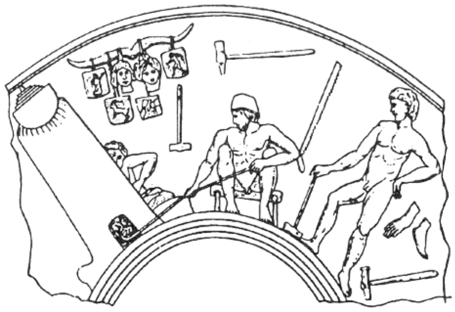


Fig. 11. Fragment of a picture in the temple No 100 in Thebes erected during the reign of Tuthmosis III (1,490-1,436 B.C.) showing gifts in the form of the copper ingots from Crete submitted to the Egyptian Pharaoh

was used earlier than Sn as a bronze addition. The civilization of the Bronze Age initiated in the middle of the third millennium B.C. in Asia Minor and Egypt (fig.9), embraced the entire Mediterranean basin and Southern Europe by c. 2,000 B.C. (fig.10), and at about three centuries later also Central Europe. Metal blooms became the objects of commercial exchange, took over the role of money, and became the objects of accumulation (fig.11). Works of art and craft, weapons, ornamentations, and cult objects, remaining from the Bronze Age attest to the good knowledge of properties of copper and bronze of their contemporary makers, from various – even located far away from each other - geographical regions. Many engineering works appeared also at that time, e.g., the 3 km long pressure water line erected at c. 180 B.C. by the King of Pergamum, Eumenes II, supplying water by the cast bronze pipes from the Hagios-Georgios mountain. As early as in the Bronze Age weapons and tools began to be made from iron, better suited for this purpose.

A sword with a golden grip made from the meteoritic iron dated from 3,100 B.C. was found in the archeological excavations in the city of Ur in Mesopotamia, and also a dagger with the iron blade was found in the Tutankhamen tomb from 1,350 B.C (fig.12). The first steel was obtained

– most probably in China c. 2,220 B.C. by the long annealing of the iron pellets mixed with chunks of charcoal with no air access, which turned out to be just the first known case of employing cementation. This process was used a little bit later in India in manufacturing weapons. Obtaining iron from its ore by the direct reduction with the charcoal took place in the middle of the second millennium B.C. south of Caucasus, beginning of the Iron Age in the Middle East was c. 1,200 B.C., in the countries of the Mediterranean basin c. 1,000 B.C., in Central Europe c. 750 B.C., and in the 4th-2th century B.C. it was in common use already. Making of iron axes (7,000 B.C.), saw and wood-working tools, as well as scissors used for shearing sheep fleece (500 B.C.), anvils and dies (200 B.C) are among many achievements of the craftsmen's art. A method of refining of the drained cast iron purged repeatedly with oxygen in Huainan was described about 120 B.C. The newly introduced iron, first from meteorites, and then obtained by ore reduction, was processed in a similar way as copper and bronze, forming the products on the required shape by cold forging and applying soft annealing in fire. Manufacturing of iron tools with the satisfactory properties, e.g., with the edges sharp enough, required cementation to 0.15÷1.5% and eliminating of slag.

Fig. 12. Dirk with an iron blade found in the Tutenkhamon's tomb in Egypt (c. 1,350 B.C.)

Invention of cementation, apparently by the "Chalybes" of Asia Minor, a subject-tribe of the Hittite empire, at 1,400 B.C. solved the problem only in part. Manufacturing of iron consisted in a process for steeling wrought-iron bars by repeated hammerings and heatings in direct contact with charcoal to diffuse carbon into the surface of the metal. If the smelting process is sufficiently elaborated, then some iron ores yield steel directly. This phenomenon was exploited c. 500 B.C. in the central European region of Styria and Carinthia. Nevertheless, it was still not known that steel required to be further hardened by quenching the hot article in cold water, which effect upon copper or bronze was to make it softer. The cementation process was followed – some two centuries later - by the tempering process. As late as in the last millennium B.C. the ancient iron metallurgy reached its height, and even then it could not provide cast iron that was virtually unobtainable with the small furnaces and low temperatures of that period. At the turn of the 8th and 7th century B.C. Homer described quenching of steel in water, which is recognized as the first literary description of this process. Just these discoveries gave rise to the contemporary heat treatment.

The unquestionable achievement of the craftsmen's art remains up to our time the Damascus steel, attesting to their high skill of the thermo-mechanical treatment. During the conquest of India by Alexander the Great in 327 B.C., swords made from this steel were used, with very good properties, which were later forgotten, albeit began to be widespread in Europe in the 3rd century B.C. This steel was first made by Hindu tribes by forging together the sintered bars from the steel containing 1.2÷1.8%C at a temperature of about 750°C, annealing and repeated forging, which ensured its high hardness and elasticity, although it was not quenched, with the corrugated lines visible on its surface, originating due to its natural etching throughout ages in natural conditions. Weapon made from this steel with good properties appeared in ancient Rome, samurai swords were also known in Japan at that time. This production was improved in the 4th-11th centuries A.D. by Arabs, mostly near Damascus, and then in Central Asia, Syria, and Persia, to appear again in Europe at the turn of the 16th and 15th centuries A.D., as the Damascus steel this time, that was – most probably – in addition quenched and tempered. Table 2 presents the historical development of the heat treatment technology.

Table 2. Historical development of the heat treatment technology

Period	Type	Geographical region
c. 2,220 B.C.	Cementation of iron in the presence of charcoal	China
8th/7th century B.C.	Quenching of steel by heating to the red heat and immersing in water	Mediterranean basin
c. 530 B.C.	Soft annealing with the subsequent forging of the steel bars welded together	India
3rd century B.C.	Annealing of castings to obtain malleable cast iron	China
100 B.C.	Nitriding or carbonitriding of steel by burning of soya in the presence of the red hot steel	China
1st century A.D.	Employment of olive oil for quenching of steel	Rome
11th century A.D.	Manufacturing of Damascus steel	Europe
15th century A.D.	Annealing of castings	France
1924	Gas nitriding	Germany
1930	Bath and glow discharge quenching	Germany
1938	Heat treatment – heating with the electron beam	Germany
c. 1965	Laser heating in heat treatment	Europe, USA

The 6th century A.D. in China and in the 10th centuryA.D. in Hartz Mountains in Europe brings an invention of obtaining iron in its liquid state, which took place before developing the iron blast furnace. In Medieval times, in many regions, including also Europe, metals processing methods were developed, along with the necessary technological equipment, among others for armour making and for minting (fig.13).

Inventions of Bessemer (1856), Siemens (1863), Martin brothers (1865), and Thomas (1877), created the fundamentals of the modern, mass production of steel with the engineering methods (table 3). In the second half of 19th century A.D. and in the first fifty years of the 20th century A.D. most of the steel groups known now were worked out as regards their chemical composition and technology, and the 20th century A.D. witnessed emerging of the metal alloys theory from the hands-on practice. In 1722, Reaumur presented a schema of the internal structure of steel. For the first time in the world he investigated this structure using the light microscope. In 1799 Clouet and de Morreau found out that iron obtains its hardness due presence of carbon. The first systematic studies on fusibility and crystallisation of alloys were begun by Rodberg (1831). Sorby, in 1864, was the first one to carry out observations of the etched steel, and this work was continued by Martens (1878). The first alloy microscopic structures were obtained by Osmov and Werth (1885), and the metallographic investigations of the microstructure were begun by Le Chatelier in 1890-1905. In 1867 Matthiessen explained presence of impurities and intermetallic compounds in metals, and Guthrie in 1884 gave the definition of the eutectic alloy.

Fundamentals of the knowledge about the phase transformations in iron alloys were created by Chernov (1868) working on the carbon-iron phase equilibrium diagram, Abel, who in 1888 has found out occurrence of the F_3C cementite in iron alloys, as well as Osmov (1895-1900) who discovered martensite as a separate phase in the quenched steel (table 4). Later results of works of Bain and Davenport (1929) concerned kinetics and mechanism of transformations of the super-cooled austenite, creating fundamentals of the theory of heat treatment of iron-base alloys. The exceptional significance had discoveries of Wilm (1906), as well as of Guinier and Preston (1935) concerning processes of supersaturation and ageing. Development of the physical chemistry, physics and chemistry of the solid body, as well as of physics of metals, electron theory of metals

Fig. 13. Striking coins from the 12th century A.D. – Norman carving (currently a symbol of Journal of Materials Processing Technology published by Elsevier B.V., The Netherlands)

Table 3. Landmarks of materials knowledge development

Scientists' Names	Year	Studies
Reaumur	1722	schema of the internal structure of steel
Clouet and de Morreau	1799	iron obtaining its hardness due presence of carbon
Rodberg	1831	first systematic studies on fusibility and crystallisation of alloys
Osmov and Werth	1855	first microscopic structures of alloy
Bessemer Siemens Martin brothers Thomas	1856 1863 1865 1877	fundamentals of modern mass production of steel with the engineering methods
Sorby Martens	1864 1878	observations of the etched steel
Matthiessen	1867	impurities and intermetallic compounds in metals
Guthrie	1884	definition of the eutectic alloy
Le Chatelier	1890-1905	metallographic investigations of microstructures

Table 4. Essential stages of the phase transformations in iron alloys

Scientists' Names	Year	Studies
Chernov	1868	carbon-iron phase equilibrium diagram
Abel	1888	F_3C cementite in iron alloys
Osmov	1895-1900	martensite as a separate phase in the quenched steel
Wilm Guinier and Preston	1906 1935	supersaturation and ageing processes
von Laue Bragg brothers	1912 1913	X-ray diffraction
Bain and Davenport	1929	kinetics and mechanism of transformations of the super-cooled austenite, fundamentals of the theory of heat treatment of iron-base alloys
Knoll and Ruska	1931	electron microscope

and quantum mechanics, theory of defects of the crystal structure, physics of plastic deformation and cracking, as well as of grain boundaries, developed in the 20th century A.D., have provided the comprehensive cognitive basis for development of the contemporary heat treatment theory of metal alloys. Research methods introduced in parallel turned out to be very useful, mostly discovery made by von Laue (1912) and employed by Bragg brothers (1913) of the X-ray diffraction, designing by Knoll and Ruska (1931) of the electron microscope, and also development of spectral methods, mostly WDS and EDS, which – equipment and research methods alike – were improved and modified in the next decades and their use for the contemporary materials science and engineering has been invaluable.

The historical review presented above indicates that only after three millennia of the practical use of iron and its alloys and five thousand years of using copper and its alloys, and also of other metals, the roles were learned of the chemical composition and phase transformations occurring during heat treatment, in forming of the structure and properties of these alloys. Science and technology developed in time as different and separate activities. The science used to be a field of speculation practiced mostly by philosophers, while the technology was a matter of practical concern to craftsmen of many types. The fields of interest of scientists and technologists remained different in the ancient cultures. This situation began to change as late as during the medieval period of development in the West when both technical innovation and scientific understanding began to interact being stimulated by the commercial expansion and a growing rapidly urban culture. Looking at this process from the contemporary perspective we can see that it was not just a one-way influence of science on technology. It was a synergy of these two activities as technology created new tools and machines with which the scientists were able to achieve better insight and understanding of phenomena.

The role of phase transformations in technological processes of heat treatment of metallic engineering materials

Nowadays, a selection of engineering materials for the technical applications is also closely connected with - not only - a selection of the appropriate material, but also with ensuring its required structure, in case of metal alloys mostly by the heat treatment methods. It should be noted that working properties of a product are obtained only when the right material is used, that has to be manufactured in the correctly selected technological process. This process should ensure both the required shape and other geometrical features, including the dimensional tolerances of particular parts that will make the final assembly of the product possible. Nevertheless it is also an issue of forming the required material structure, ensuring the resulting expected mechanical, physical, and chemical properties of the product (fig.14). Forming of the required structure of metal alloys due to their heat treatment is connected with phase transformations occurring in them in solid state.

Fig. 14. Relationships among some factors connected with material, processes and functions of a product

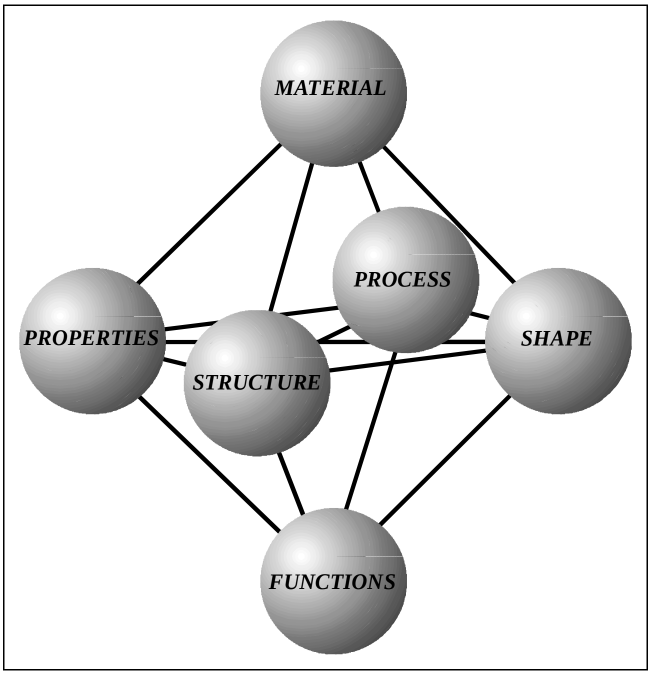


Table 5. Classification of phase transformations (prepared according to J.W. Christian)

Heterogeneous transformations

- Athermal growth (moving phase boundary)
  - Coherent boundary
    - Coherent martensite
    - Mechanical twinning
  - Semi-coherent boundary
    - Martensite
    - Low-angle borders
- Heat activated growth
  - No mass flow (controlled by processes on grain boundaries)
    - Polyform transformations growth from the gaseous state (vapour)
    - Grain growth during recrystallization
    - Ordering
  - With mass flow
    - Continuous transformations
      - Controlled by near-border processes
        - Growth of the crystal from vapour
        - Release from solutions
      - Release from solutions controlled by diffusion
    - Discontinuous transformations (diffusion + near-border processes)
      - Eutectoid transformation
      - Discontinuous cellular disintegration
- Growth controlled by heat flow

Phase transformations are phenomena causing changes of structure and physical and mechanical properties of matter. They include processes like vaporising and condensing from the gaseous state into the liquid or solid state, melting and transition of substance from the liquid state into the crystalline or amorphous state, and changes of the arrangement of atoms or other particles of the matter in solid state, connected with the allotropic transformations of pure chemical elements or with phase transformations occurring in metal alloys or other substances. All the above mentioned transformations may occur solely in the direction of minimising the Gibbs free energy of a system, and the difference of the Gibbs free energy in the input and output states of the system is a driving agent of each of these transfor- mations.

The classification of phase transformations according to growth processes, proposed by Christian and accepted by the International Committee for Phase Transformations is presented in Table 5. Generally speaking, phase transformations can be divided into homogeneous and inhomogeneous. The flow of each transformation can be divided into the nucleation and growth of the transformation products. Transformation order disorder and a spinodal decomposition of the solid solution are reckoned as the homogeneous transformations. The inhomogeneous transformations are divided into the diffusion ones of the nucleation and growth type and to the diffusionless ones of the martensite type, albeit in some cases both transformation mechanisms may proceed simultaneously. The product of the diffusion transformation may differ from the input phase by the state of matter, lattice structure, without a change of a chemical composition and may occur as a mixture of various phases with a differentiated lattice structure and chemical composition. This group of phase transformations includes crystallisation of metals and alloys, allotropic transformations, precipitation processes in supersaturated solid solutions, cellular growth – including the discontinuous precipitation and the eutectoid transformation. The martensitic transformations occur solely in solid state and are diffusionless and consist in a coordinated displacement of a large group of atoms over small distances in respect to a certain lattice plane of the output phase. The growth of phases during the transformation depends on a structure and a mobility of the phase boundaries. The coherent and also – to some extent – the semi-coherent boundaries are active without thermal activation as opposed to the non-coherent boundaries.

Table 6. Examples of employing various phase transformations in different heat treatment processes used for diverse metal alloys

Phase transformation type	Heat treatment process type	Group of materials	Additional heat treatment
Martensitic (in some cases also bainitic)	Quenching, quenching and tempering, thermal hardening	Special-purpose construction steel Steels for operation at elevated temperatures for pressure equipment Special-purpose machine steels Machine carburising steel Special-purpose corrosion-resistant high-chromium martensitic steels martensitic steels for valves Special-purpose corrosion-resistant tool steels High-speed steels Alloyed cold-work tool steels Alloyed hot-work tool steels Special-purpose bearing steels	High-temperature tempering High-temperature tempering High-temperature tempering Carburising and tempering Tempering High-temperature tempering Low-temperature tempering High-temperature tempering Tempering High-temperature tempering Low-temperature tempering
Pearlitic	Normalising	Constructional fine-grained weldable quality steels Special-purpose constructional steel resistant to atmospheric corrosion	Preceding austenitisation Preceding austenitisation
Precipitation	Ageing, precipitation hardening	Special-purpose corrosion-resistant Cr-Ni, Cr-Ni-Mn austenitic steels Special-purpose corrosion-resistant Cr-Ni ferritic-austenitic steels Special purpose austenitic steels Heat-resisting and creep-resisting ferritic-austenitic and austenitic steels Alloys of Cu and Al and other non-ferrous metals with a variable solubility in the solid state Creep-resisting Ni alloys	Preceding supersaturation Preceding supersaturation Preceding supersaturation Preceding supersaturation Preceding supersaturation Preceding directional crystallisation
Recrystallisation	Recrystallisation annealing	Steels for flat products for cold-working Special-purpose corrosion-resistant high-chromium and ferritic steels Single-phase alloys of Cu, Al, and other non-ferrous metals	Preceding cold-working Preceding cold-working Preceding cold-working
Dynamic recrystallisation and precipitation	Regulated thermo-mechanical treatment, rolling, controlled recrystallisation	Fine-grained weldable constructional steels with microadditions of Al, V, Ti, Nb, Zr, N, B Steel with low yield point	Preceding hot-working Preceding hot-working
Reverse martensitic transformation	Heating	Alloys with the matrix elastic memory like TiNi, CuZn, and other with the B2 lattice and L20 order	Preceding martensitic transformation
Nano-crystallisation	Annealing	Fe, Ni or Co allys with additions of B, C, P, Si, Ge, Hf, and Zr, magnetically soft	After the preceding vitrification during cooling with the rate exceeding > $10^4 \cdot K \cdot s^{-1}$

Mechanisms and kinetics of phase transformations depend significantly from the forced conditions and are the reason for the differentiated phase composition and morphology of phases, deciding structures of solids, both during changes of the state of matter from liquid into solid as well as in the solid state. Phase transformations are employed to forming the required properties of elements from metal alloys, using the heat treatment methods, making it possible to form structure of these alloys. Moreover, processes may occur in solids causing the significant mutual displacements of atoms and changes of their neighbours without changes of the crystalline structure (plastic deformation, recovery and recrystallisation, grain growth), differing from phase transformations, as change of the arrangement of atoms is not connected in this case with the Gibbs free energy change but is a result of stresses or a pursuit of the system to decrease its internal or surface energy. The second case has an application in heat treatment (recrystallisation annealing). Examples of the employment of phase transformations in heat treatment, mostly for steel, are presented in Table 6.

Because of the steady growth of importance of ceramic, composite, and polymer materials among the engineering materials, employment of metal materials decreases relatively. This becomes the main reason for the decrease of the relative portion of heat treatment in manufacturing technologies, from c. 20% not so long ago – in 1992, down to c. 10% in 1998, albeit in the absolute scale the growth of investment expenditures on heat treatment of machine and equipment elements, as well as tools, is evident. The development of the heat treatment technology from the beginning of the last century has been connected with the steady growth of the importance of surface treatment, at the cost of the decreasing portion of body treatment (fig.15). A current state of the art and a significance of the particular heat treatment technologies are presented in Table 7. From all heat treatment development trends, the heating evolution is certainly the most important one. Electrical and fuel heating devices are still used. In spite of the higher investment expenditures one should expect the growing trend to use electrical energy in heat treatment and limiting the fuel energy, both due to eco-logical reasons and because of the easier automation and robotisation of the technological processes. The highly efficient methods of electrical heating offer the additional possibilities of increasing the heating rate, and also of depositing or implanting chemical elements or compounds on the charge surface.

Fig. 15. Changes of portions of the heat treatment technology (prepared according to J.M. Belot and D. Ghiglione): 1- body treatment, 2 - skin hardening, 3 - nitriding, 4 - solid carburising, 5 - liguid carburising, 6 - gas carburising, 7- vacuum and glow carburising

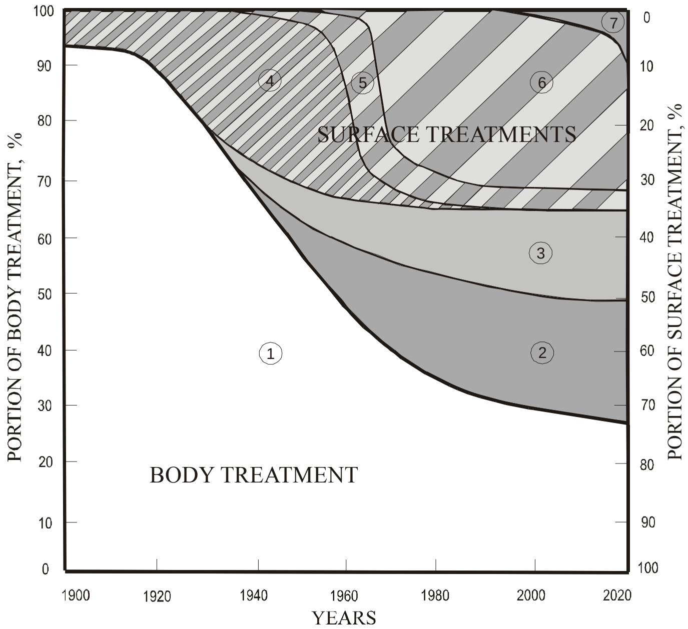


Table 7. Development and use of various heat treatment technologies (prepared according to T. Burakowski)

Heat treatment technologies	Furnace technologies	Common	Atmospheric	—	●
			Powder	↓	●
			Paste	↓	●
			Air	↓	●
			Bath	—	●
			Inductive	↑	◕
		Modified	With the protective atmosphere	↑	◕
			Fluidised	↑	○
			Vacuum	↑	◑
	Heater technologies	Non-beam technologies	Bath	—	◕
			Inductive	↑	◑
			Glow discharge	↑	◔
		Beam technologies	Laser	↑	○
			Electron	↑	◔
			Plasma	—	◕
Development rate	○ very fast	◔ fast	◑ medium	◕ slow	● no development
Use	↑ increasing		↓ decreasing	— stable	

Taking into account both requirements of the sustainable development and automation along with an introduction of information technologies to technological processes one can state that the current development of the ordinary heat treatment of engineering metallic materials includes the following trends:

- search for the new, efficient ways of heating and new insulation materials, minimising energy consumption and use of machined materials due to interaction of temperature and atmosphere, and – to a somewhat less extent – being a burden to the environment,
- search for and introducing the new efficient cooling media, featuring a lesser burden to the environment, that would not interfere with the subsequent technological operations,
- introducing integration of the heat treatment operations to increase the efficiency of the heat treatment process and integration of the heat treatment processes with other manufacturing processes, also to increase the overall efficiency of the manufacturing process and to ensure the required structure and properties of products.
- selection of the right production scale, optimised taking into consideration the economical issues and burden of environment.
- search of new materials and making rational use of phenomena occurring during their heat treatment to ensure the desired, most advantageous properties of products or their elements made from them.

The role of modern and advanced materials in future development forecast

Analysing the contemporary development trends of various material groups, it is realized - which is evident - that the mass share of the ultramodern products (like space technology products or even biomedical materials) in all products manufactured by people, is not large, albeit it is continuously growing. So far it is not possible that polymers would conquer the surrounding reality (they only seem to be ubiquitous), because of their relatively low wear resistance and other types of wear, low corrosion resistance, and also because of their service temperature range that does not exceed 300÷400°C. Porous ceramics belongs to the construction domain, although glass has also many household applications, but additionally it is used in car manufacturing, moreover, some grades - especially glass ceramics - are used even in machines. The main materials in machine technology, car, ship, household, and tool industries, as well as in many others, but important also in construction, are still metals and their alloys, albeit in many cases engineering ceramics, and also some composites, compete successfully with these materials. Main industrial applications of some engineering materials groups are given in table 8.

One might attempt to present a vision of the future and evaluate the development trends of various engineering materials. Certainly, they are connected with forecasts pertaining development of various fields of activity and manufacturing processes. Table 9 presents these trends, basing on visions proposed by eminent bodies consisting of scientists and futurologists. Many people, even today, do their work at home, without leaving it. Houses will have to be organised and furnished in a totally other way within several years' time span. Towns, transportation and telecom systems will be organised differently than nowadays. Towns and transport system will be organised in another way, including novel urban transportation systems connecting the sky-high buildings, electrically powered cars, robotised safety systems and municipal wastes' utilization systems. Health protection system will be based on diagnoses made at hor non-allergic nutrition, an early detection of serious illnesses and their disallowing, but a implanting of artificial organs - heart, and of a new generation of biomaterials. Future agriculture, forestry and fish industry will be based on genetic engineering achievements, mastering farming new plants, employing other processes than photosynthesis, and also comprehensive robotisation. Mining and manufacturing industries will be based on a total robotisation of processes of industrial recycling of water and air, on a development of the ultra-microprocessor technology, and also on the high-throughput power transmission systems employing organic materials substituting copper. The Earth protection systems fighting climatic effects, implementing recovery from damages caused by torrential rains, fighting droughts, and exploitation of the tropical forests, as well as decreasing the ozone layer discontinuity effects, will undergo significant changes. Systems for surveillance of oceans and seas and monitoring their contamination, and for observations of earthquakes will be developed, moreover, robots will be introduced to underwater service. Space technique employing solar energy will make space flights more common and will give rise to novel technologies and setting up of space factories for market production, setting up of lunar observation bases and to space journeys to Mars.

Table 8. Common industrial applications of several material groups

Industrial Applications / Material groups	Building industry	Transportation	Maritime economy	Power industry	Space technology	Armaments industry	Defense and aviation	Chemical industry	Petroleum production	Household	Electronic industry and telecommunication	Materials production	Health protection	Packings	Consumer goods production	Communication electro-technology	Other industry branches
Steel and other iron alloys	●	●	●	●	●	●	●	●	●	●		●		●	●		●
Nonferrous metals	●	●		●	●	●		●			●	●			●		●
Aluminium	●	●					●			●				●		●	
Nickel				●			●										
Titanium							●										●
Copper	●									●							
Zinc														●			
Lead	●	●															
Tin														●			
Heat-resistant alloys and super alloys		●		●	●	●			●			●					
Ceramic materials and glasses	●	●		●	●	●	●	●			●	●			●	●	●
Concrete	●	●															
Composites		●			●	●									●		
Polymer-matrix composites		●	●				●									●	
Inorganic compounds						●		●									
Polymer materials	●	●								●		●		●	●	●	●
Semiconductors					●						●				●		
Biomaterials													●				
Wood	●											●					

Even if not everything, according to these forecasts, will come true or be slightly delayed, one has to take into account that nearly all of the forecasted projects will require the relevant manufacturing technologies and above all - relevant materials. Many of these materials are already available nowadays, some of them should be developed soon according to the outlined requirements. It is good to realise that many venturous projects will be made possible if these new materials are made. One can specify two main priorities in this area:

- continuous improvement of the existing materials and technological processes,
- development of materials and technological processes ensuring environment protection or improving conditions of human life or extending its time span.

Table 9. New technologies and constructional and organisational solutions foreseen to be introduced in the 21st century (prepared according to M.H. Van de Voorde)

2006 • advanced data storage • hybrid vehicles common **2007** • computerised self-care • computer sensory recognition • modular software **2008** • parallel processing computing • information superhighway • genetically produced food • personal digital assistants • half of all household waste recycled **2009** • intelligent agents • ubiquitous computing environments • spread of electronic banking/cash • broadland networks • holistic health care **2010** • alternative energy sources • 'green' environmental methods • spread of expert systems **2011** • mass customisation • electric cars are common organic energy sources **2012** • computer-integrated manufacturing • machine learning • computer language translation • farm chemicals drop by one-half	**2013** • gene therapy • online publishing • major diseases cured • half of autos recyclable **2014** • optical computers • ceramic engines • aquaculture • computerised vision implants **2015** • precision farming • factory jobs drop to 10% • spread of neutral networks • alternative/organic farming • superconducting materials • industrial ecology • hydroponic production **2016** • nanotechnology • recycled goods • sophisticated robots • energy efficiency • fossil fuels cut greenhouse gas • fuel-cell electric cars • intelligent transportation systems • materials composites **2017** • high-speed maglev • biochips • fuel cells **2018** • new materials from space	• one-half of goods sold electronically • automated highway systems • cloning/organ replacement **2019** • private space ventures • synthetic body parts • telecommuting **2020** • farm automation • urban greenhouses • genetic engineering • hydrogen energy • fission power **2022** • artificial foods **2023** • clustered communities **2025** • hypersonic air travel **2026** • intelligent materials • fusion power **2027** • self-assembling materials **2028** • permanent moon base **2037** • manned Mars mission **2042** • stellar exploration **2049** • extraterrestrial contact **2062** • near-light speed achieved

All the above mentioned facts require, of course, a dissemination of knowledge about an important role of a correct selection of materials among engineers of various non-materials science specialisations, especially these deciding products design and the technological processes of their manufacturing.

It is forecasted that in the forthcoming time technique development will force following trends in the development of engineering materials which will make it possible to use them in the engineering design of many new products expected on the market:

- development of modelling the relationships among chemical composition, structure, parameters of the technological processes, and service conditions of the engineering materials, using the modern IT tools, including the Artificial Intelligence methods, to improve the methodology of the engineering design processes, including improvement of the engineering materials' selection and the most suitable technological processes,
- development of the pro-ecological manufacturing technologies with the possible lowest harmful environmental impact and/or influencing the environment and atmosphere, as well as decreasing the degradation of the environment to date and deployment of the relevant materials and technologies,
- development of surface engineering and related technologies in order to increase significantly the competitiveness of products and technological processes, to reduce the hazard to the environment, as well as deployment of fast and inexpensive welding technologies making it possible to introduce the competitive products and manufacturing processes,
- development and deployment of the industrial applications of the "intelligent" materials and automatically supervised technological processes,
- development of manufacturing technologies making it possible employing the existing high-temperature superconductors in market products, and development of materials for the cellular telephony and telecom industry needs, including materials for opto-electronics.
- introduction of the new heat resistant and high-temperature creep resisting materials for service at elevated and high temperatures, especially for the space, aviation, automotive, power generation, and electronic industries,
- introduction of new generations of biomaterials that will render it possible to extend the range of possible medical interventions and implanting the artificial organs and limbs to improve the level of treatment of diseases and injuries.

Summary

This paper underscored the very significant role of materials' design in the design and manufacturing processes of new, demanded products, having the highest attainable quality and performance at the optimal and reasonably set possibly lowest costs level. The engineering design processes cannot be set apart from the materials' design, being more and more often computer-aided, nor from the technological design of the most suitable manufacturing processes. The review of the multi-millennia long history of human civilisation indicates that the significant increase of the level of life and production is connected most often with introduction of new material groups, having properties better and better adjusted to the real customers' requirements that get more sophisticated nearly each day, and also relevant to them technological processes. The reasons given, make it possible to forecast that the future of the market and products on, it with the required properties, are inseparably connected with the development of Materials Engineering and Materials Processing Technologies.

Acknowledgements

The author expresses his sincere thanks to Prof. Maria Helena Robert from the University of Campinas (Brasil) for her inspiration to prepare this paper, and to Mr D. Pakuła, MSc from the Silesian University of Technology in Gliwice (Poland) for his technical assistance in the preparation of this paper.

References

Adamczyk J., Metaloznawstwo teoretyczne vol. I, Wyd. Pol. Śl., Gliwice, 1999.
Ashby M.F., Materials Selection in Mechanical Design, Pergamon Press, Oxford - New York - Seoul - Tokyo, 1992.
Belot J.M. & Ghiglione D., Trait. Therm., 307, 1998, p. 20.
Brady G.S., Clauser H.R., Materials Handbook, McGraw-Hill Inc., New York - St. Louis- San Francisco - Auckland -Bogotá - Caracas - Hamburg - Lisbon - London - Madrid - Mexico - Milan - Montreal - New Delhi - Paris - San Juan - São Paulo - Singapore - Sydney - Tokyo - Toronto, 1991.
Burakowski T., Przegląd Mech., 59, 2000 p.10.
Christian J.W., The Theory of Transformations in Metals and Alloys, Perg. Press, Oxford, 1975.
Davis J.R., Mills K.M. (eds.), ASM Handbook - Properties and Selection: Irons, Steels, and High - Performance Alloys, ASM, Materials Park, Vol.1, 1990.
Derry T.K. & Williams T.I., A Short History of Technology, Dover Publications, Inc., New York, 1993.
Dieter G.E. (ed.), ASM Handbook - Materials Selection and Design, Metals Park Vol.20, 1997.
Dobrzanski L.A. (ed.), Zasady doboru materiałów inżynierskich z kartami charakterystyk, Wydawnictwo Politechniki Śląskiej, Gliwice, 2001.
Dobrzanski L.A., Metaloznawstwo z podstawami nauki o materiałach, WNT, Warszawa 1999.
Dobrzanski L.A., Proceedings of 11th ADM International Conference on Design Tools and Methods in Industrial Engineering, Sesion B, Palermo, 1999, p.31.
Dobrzanski L.A., Proceedings of 12th ADM International Conference on Design Tools and Methods in Industrial Engineering, (CD-ROM), Rimini, 2001.
Dobrzański L.A., Procedings of 15th Brasilian Congress of Mechanical Engineering COBEM'99, (CD-ROM), Sao Paulo, Brasil, 1999.
Dobrzański L.A., Proceedings of 8th Seminar of IFHTSE, Dubrownik-Cavtat, 2001, p.1.
Dobrzański L.A., The role of materials selection in the design and manufacturing processes of contemporary products and their elements, Proc. of the Workshops EUINTEGRATION, vol.2, Gliwice - Leganes - Dublin - Guimaraes - Manchester - Orbassano, 1999.
Liščić B., Tensi H.M., Luty W., Theory and Technology of Quenching, Springer Verlag, Berlin - Heidelberg - New York - London - Paris - Tokyo - Hong Kong -Barcelona - Budapest, 1992.
Maciejny A., Od metalografii do nauki o materiałach i inżynierii materiałowej, Politechnika Śląska, Gliwice, 1999.
Totten G.E., Howes M.A.H., Steel Heat Treatment Handbook, Marcel Dekker, Inc., New York - Basel - Hong Kong, 1997.
Van de Vorde M.H., Contemporary research and development trends in materials engineering area – invited lecture, AMME'98 International Scientific Conference, Gliwice – Zakopane 1998.
Waterman N.A., Ashby M.F., The Materials Selector, Vol. I-III, Chapman & Hall, London - Weinheim - New York - Tokyo - Melbourne - Madras 1997.

THE Ti+TiN, Ti+Ti(CXNX-1), Ti+TiC PVD COATINGS ON THE ASP 30 SINTERED HIGH SPEED STEEL

Leszek A. Dobrzanski
Silesian University of Technology
Gliwice – Poland

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

ABRASION IN WEAR AND MANUFACTURING PROCESSES

Professor Ian M Hutchings

University of Cambridge, Institute for Manufacturing, Department of Engineering, Mill Lane, Cambridge CB2 1RX, UK
imh2@cam.ac.uk

Abstract. Abrasion by hard particles is responsible for wear in many practical situations, but can also be employed usefully in the manufacturing processes of grinding and polishing. This lecture presents a brief overview of abrasion, an historical survey of polishing and then discusses abrasion test methods.

Keywords: abrasion, wear, grinding, polishing, wear test

1. Introduction

Abrasion is responsible for wear in many practical applications; indeed Eyre (1981) has suggested that about 50% of all industrial wear problems can be ascribed to abrasion. Abrasion is also widely used in manufacturing, to produce smoothed and shaped surfaces. Abrasive wear is defined by ASTM (1999) as 'wear due to hard particles or hard protuberances forced against and moving along a solid surface'. This paper examines abrasion processes over a wide range of length scales, and discusses some useful applications of abrasion, as well as exploring our understanding of polishing and grinding from an historical viewpoint. Some recent developments in laboratory-scale abrasive wear testing are also reviewed.

2. Abrasive wear

Abrasive wear occurs in the presence of small hard particles and relative sliding motion. The particles may either be fixed to a counterbody as illustrated in Figure 1(a), which leads to 'two-body abrasion' or be free to slide and roll as in 'three-body abrasion' (shown in Figure 1b). For many purposes it is useful to describe the motion of the particles against the wearing surface, giving a classification into either 'grooving (or sliding) abrasion' or 'rolling abrasion' (Trezona et al. 1999). Grooving abrasion causes pronounced linear scratches on a metallic surface, whereas rolling abrasion results in multiple indentations and produces a surface with no significant directionality. Abrasive wear is also sometimes classified as either 'low stress', where the abrasive particles remain relatively undamaged, or 'high stress', in which the particles experience extensive fracture.

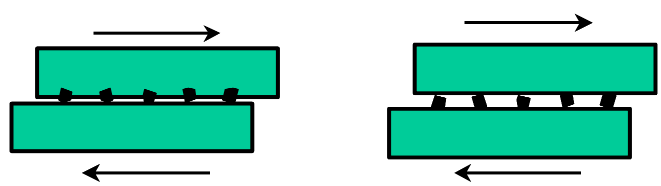


Figure 1. (a) 'two-body' or grooving abrasion, in which the particles are fixed to one surface; (b) 'three-body' abrasion, in which the particles are free to roll or slide between the surfaces.

The rate of wear is influenced by the hardness of the particles. If they have a lower hardness than the surface then wear will be much lower than with harder particles. For a solid particle to indent or damage a surface by scratching, its indentation hardness must be greater than that of the surface by a factor of at least ~1.2. The commonest natural abrasive is crystalline silica (quartz) which has a hardness (Vickers or Knoop) of ca. 8 to 10 GPa (800 to 1000 kgf mm^{-2}). Since even a hard martensitic steel will have a hardness less than 1.2 times this, it is clear that steels and non-ferrous metals will be especially susceptible to abrasive wear by natural quartz. Materials which contain hard phases, such as cermets and hard-facing alloys, often also contain softer phases. Ceramic materials, however, may be sufficiently hard and homogeneous to resist indentation by most common abrasive particles, and these materials can therefore provide useful wear resistance in the form of bulk materials or coatings. Because they are so hard, even harder abrasive particles such as diamond, silicon carbide or alumina must be used to grind or polish them.

The fundamental mechanisms responsible for abrasive wear can involve both plastic flow and brittle fracture. Plastic flow can occur alone, but both can sometimes take place together, even in materials like inorganic glasses which are brittle on a large scale. In the simplest models for abrasive wear it is assumed that either plastic flow or brittle fracture provides the dominant wear mechanism.

For 'grooving' abrasion of a ductile material by hard particles an equation identical to the Archard wear equation for (non-abrasive) sliding wear can be derived (Archard 1953, Rabinowicz 1965):

$$Q = \frac{KN}{H} \tag{1}$$

Here Q is the volume removed from the surface per unit sliding distance, N is the normal load on the contact (assumed to be distributed between all the abrasive particles) and H is the indentation hardness of the surface which is being worn. K is dimensionless, and has a value which is typically between 10^{-2} and 10^{-3} for the grooving abrasive wear of metals.

Equation (1) suggests that the volume of material removed by wear should be proportional to the sliding distance, and also to the normal load on the particles; this type of behaviour is often observed. Wear rate also varies inversely with hardness for pure metals, but more complex behaviour is seen for alloys.

Models for abrasive wear by brittle fracture are usually based on the crack patterns associated with a hard, sharp indenter (Hutchings 1992). Lateral cracks which grow towards the surface from beneath the indentation, or from beneath the plastically-formed groove in grooving, are responsible for material removal. Both fracture toughness K_c and hardness H play important roles in controlling the occurrence and extent of lateral cracking, and they therefore control the wear rate. Several different models have been proposed, which suggest that the wear rate will be inversely proportional to both H and K_c, raised to powers of about 1/2. They do not take into account the influence of microstructure on wear rate in brittle materials such as bulk polycrystalline ceramics or hard coatings, nor are they applicable to the wear of hard materials by rather softer abrasive particles.

The range of length and time scales involved in mechanical processes can be illustrated on a chart, as in Figure 2. With logarithmic scales on both axes, this allows a very wide range of phenomena to be compared over length scales which extend from interatomic distances (ca. 0.1 nm) to the size of the Earth (ca. 10 Mm), and from the time of one atomic vibration in a solid (ca. 0.1 ps) to a human lifetime (ca. 2.4 Gs). In this diagram, sliding speeds can be plotted as straight lines of slope 1, and areas can be used to show the regimes of lateral distance and time (or sliding speed) over which particular tribological processes occur.

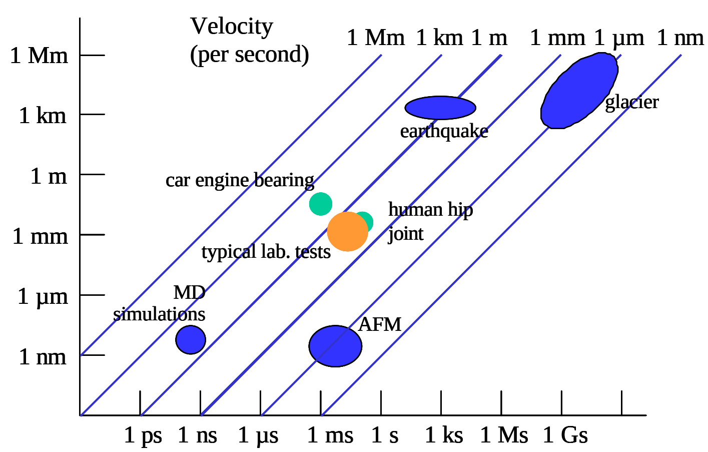


Figure 2. Chart representing abrasion processes over a wide range of length and time. The distance and time scales are logarithmic; the diagonal lines represent constant sliding velocities.

The range of sizes and sliding speeds involved in typical machine components occupies quite a small area in Figure 2: from millimetres to metres, and from millimetres per second to, at most, hundreds of metres per second. Some specific tribological elements are shown in Figure 2: the main bearing of an automobile engine, a synovial joint in the human body (the hip joint), and the range of conditions used in typical laboratory abrasion tests. Neither engine bearings nor human synovial joints should normally experience abrasive wear, but it can be important in both under certain conditions. For example, contamination of oil or grease by hard grit particles can lead to serious abrasion in both plain and rolling element bearings (Kahlman and Hutchings 1999, Williams and Hyncica 2001). In artificial hip joints, the presence of fine isolated scratches on an otherwise smooth metallic femoral head can cause rapid abrasion of the polymer acetabular cup in which it slides (Dowson et al. 1987).

Abrasion does not only occur on an engineering scale, but also in geological processes. Both earthquakes and glaciers involve sliding over large length scales, but at rather different speeds. The fragments of rock which are produced on the fault plane by an earthquake will undoubtedly lead to large-scale abrasion, but observation and analysis of these phenomena are not straightforward. As a glacier slides it can drag boulders along its path, pressing them against the underlying rock. Typical linear features of plastic grooving are produced and are visible once the glacier has melted away, and there are often also crescent-shaped crack patterns on the rock surface, features associated with fracture (Hambery and Alean 1992).

Scanning probe microscopes can be used to study abrasion damage on a much smaller scale, and also to produce it. Scratching by a sharp indenter has been used, for example, to study the behaviour of silicon surfaces on a submicrometre scale (Bhushan et al 1995), while recently multiple indenters have been used to 'plough' furrows across solid surfaces and thus produce precisely controlled grooves as a novel method of device fabrication (Li et al. 2000). Metallic conductor tracks as fine as 0.15 µm wide have been demonstrated.

3. Grinding and polishing in manufacturing

Grinding and polishing, which both involve carefully controlled abrasion, have been used and developed over many hundreds of years to produce smooth surfaces, either plane or with specific and well-controlled shapes. In order to optimise these processes, a good understanding is needed of the mechanisms involved. Sir Isaac Newton (1704) studied the polishing process for glass, in connection with the manufacture of lenses and mirrors. He wrote that 'the smaller the particles are, the smaller will be the scratches by which they continually fret and wear away the glass till it be polished; but, be they never so small, they can wear away the glass no otherwise than by grating and scratching it, and breaking the protuberances; and therefore polish it no otherwise than by bringing it to a very fine grain, so that the scratches and frettings of the surface become too small to be visible.' Two hundred years later, Lord Rayleigh (1899) concluded that the process of grinding, in which local fracture of the glass could be detected readily with an optical microscope, was distinctly different from polishing. He suggested that 'in the process of grinding glass surfaces, the particles of emery, even the finest, appear to act by pitting the glasses, *i.e.* by breaking out small fragments.. (whereas)…during polishing…... it seems probable that no pits are formed by the breaking out of fragments, but that the material is worn away (at first, of course, on the eminences) almost molecularly.'

It is now known that Rayleigh was more correct than Newton: for scratches in glass which are narrower than about 1 µm, brittle fracture does not occur as there is not enough energy available to grow the cracks, and a groove is formed by plastic deformation (Puttick 1979). However, it is also clear that the polishing of glass involves tribochemical processes and is not purely mechanical.

A great deal of attention has also been paid to the polishing of metals, with contributions by many distinguished researchers. Rabinowicz (1968) and Samuels (1982) have both presented good historical reviews, and have critically discussed the evidence for various mechanisms. It is now well-established that the polishing of metals is caused by plastic deformation on a very small scale, and often also involves some additional tribochemical activity. In spite of this, Beilby's proposal (Beilby 1903, Beilby 1914, Beilby 1921) that the polishing of metals was associated with the formation of an amorphous surface layer continued to receive support for much of the 20th century.

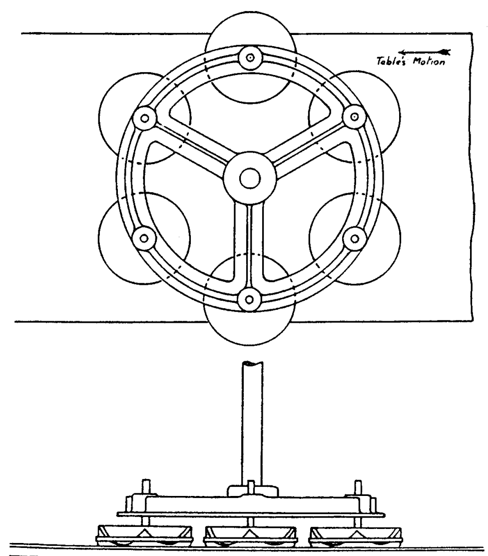


Figure 3. Plan and side views of a polishing machine, with six independently rotating heads, used to flatten and polish a sheet of glass (from Preston 1927)

An early and influential researcher into the grinding and polishing of glass was F W Preston, who was proposed an empirical model for these important manufacturing processes, which he used to optimise the production of flat glass sheets. Before the float glass process was introduced in the 1950s, flat glass was produced by abrasive processing. Rotating grinding and polishing heads were pressed against the glass, as shown in Figure 3. In this machine, six independently rotating heads carried abrasive-charged felt pads. Preston (1927) wrote: 'Consider a piece of glass and a polishing felt pressed together with a pressure of p and moving with a velocity relative to one another of v. Then there is good experimental evidence to believe that the amount of glass polished off in time t is proportional to pvt.'. This relationship is now often referred to in the polishing community as the 'Preston equation', but is exactly the same as equation (1), the Rabinowicz equation for abrasive wear which was proposed several decades later. Most later models for the lapping and grinding of glass have used the Preston equation as their basis (Buijs and Korpel-Van Houten 1993). Abrasive stock removal, surface profiling and finishing are core manufacturing processes for certain glass artefacts, and play a major role, for example, in the production of glass cathode ray tube screens. It has been estimated that more than 30% of the production costs of these items is accounted for by the grinding and polishing processes (Geltink-Verspui 1998).

Another application of grinding and polishing in which it is essential to predict material removal rate accurately is in the production of mirrors for astronomical telescopes. This was a major reason for Newton's interest in the processes at the end of the seventeenth century. The four 8.2 m diameter mirrors for the VLT (very large telescope) project at the European Southern Observatory at Paranal in Chile provide a modern example. This terrestrial instrument is expected to provide resolution which is even better than the Hubble space telescope. The mirrors are 175 mm thick discs of low-expansion glass-ceramic which are ground and polished under computer control to achieve the required local material removal rate and final dimensions (Dierickx et al.1996). The final mirror completed in December 1999 has been claimed to be the 'best astronomical mirror in the world', with a precision of form of 8.5 nm over its surface.

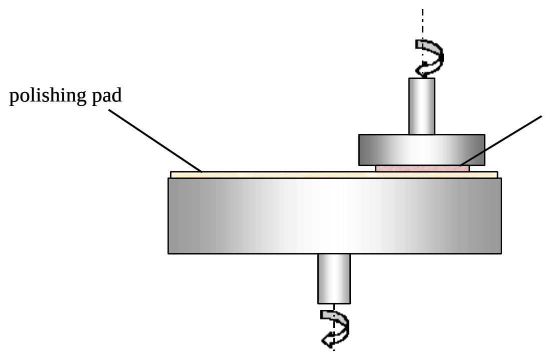


Figure 4. Schematic diagram showing the process of chemical-mechanical planarization (CMP), as applied to semiconductor wafers

Chemical mechanical polishing or planarization (CMP) is another precision polishing process for which a detailed model is required. CMP is rapidly becoming a key step in the manufacture of semiconductor devices. It involves a combination of abrasive polishing with chemical etching. Figure 4 shows the mechanical action, which is broadly similar to that shown in Figure 3 for the polishing of flat glass sheets. Current semiconductor devices (2001) have a typical linear dimension (e.g. conductor width) of 0.15 µm: this is predicted to fall to 0.13 µm in 2002, 0.10 µm in 2005 and as low as 70 nm in 2008. The wafer surface must be extremely flat so that such fine features can be reproduced by UV photolithography with very small depth of focus. CMP plays a central role in achieving this flatness, as well as in new processes for producing multilayer circuits, such as the 'damascene' technique. Much attention has been given to modelling CMP recently, in order to predict and control the material removal rate. Even now, the simple Preston equation is usually used as an empirical starting point (Luo et al. 1998).

4. Abrasion test methods

There are several possible reasons for the experimental study of abrasive wear: to investigate fundamental wear mechanisms; to measure and compare wear rates of different materials under identical conditions; to simulate and optimise related manufacturing processes (such as free-abrasive grinding, lapping and polishing); and as quality assurance tests for coatings and other treated surfaces. Grindstones, as used to produce sharpen knives and other tools, were used for early abrasion testing. Figure 5 shows a grindstone which was described by the Russian chemist and poet Lomonosov in 1752 (Menshutkin 1952). He listed it among eight instruments required for a university course in experimental chemistry, stating that students should 'investigate substances, especially metals, by long continued abrasion'.

During the early years of the twentieth century, various different abrasive wear tests were developed. One of the few to be standardised was the dry sand rubber wheel test (ASTM 2000). In this method a stream of dry silica abrasive particles flows between a flat specimen and a rotating rubber-rimmed wheel. The specimen is pressed against the wheel at a fixed load by a dead weight. This test can be modified to provide better control of the experimental conditions, by feeding the abrasive on to the top of the wheel and by incorporating a load cell to measure the tangential force, as shown in Figure 6. With this apparatus the influence of the test conditions on the abrasion process has been investigated in some detail (Stevenson and Hutchings 1996).

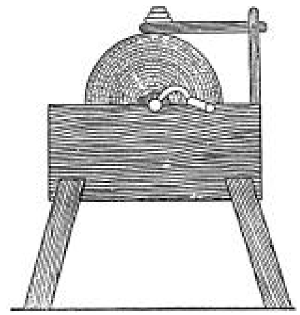


Figure 5. Grindstone proposed by Lomonsov for use as a laboratory abrasion test in St Petersburg, Russia, in 1752 (Menshutkin 1952).

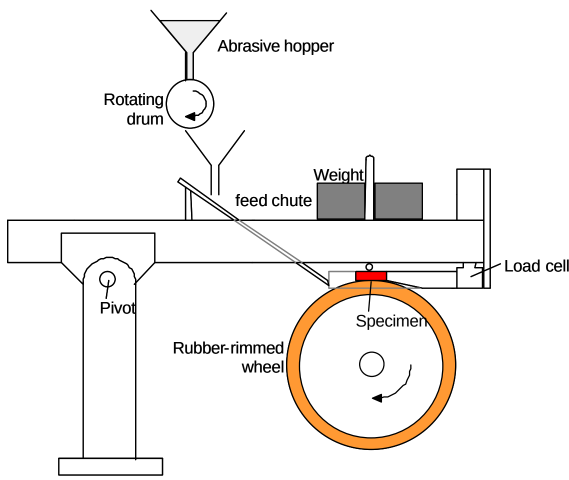


Figure 6. Apparatus for the dry sand rubber wheel abrasion test (after Stevenson and Hutchings 1996).

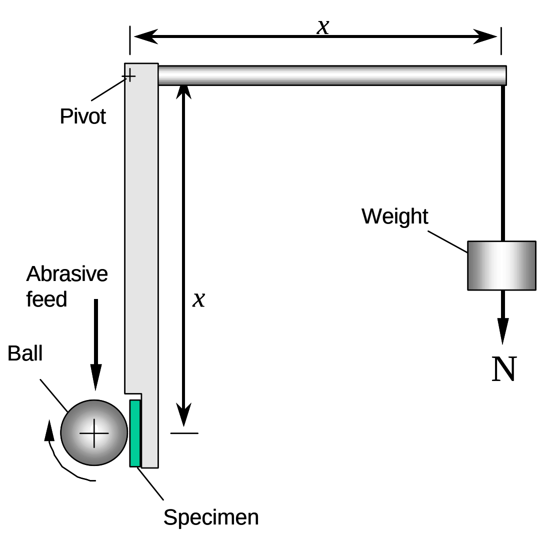


Figure 7. Apparatus for the micro-scale abrasion test: a steel ball rotates against the surface of the specimen, and abrasive particles are supplied to the contact zone as a suspension in a liquid.

Recently there has been considerable interest in the ‘micro-scale abrasion’ or ‘ball-cratering’ test. Its history has been reviewed by Rutherford and Hutchings (1997). In this method a sphere (often a hard steel ball with a diameter of about 25 mm) or a spherically-crowned disc rotates under a constant load against the surface of the specimen. The abrasive particles are supplied directly to the contact region suspended in a liquid. Unlike the rubber wheel test, the method results in a wear scar with a shape which conforms accurately to that of the ball and therefore has a well-defined, spherical geometry. This allows the amount of wear to be determined accurately from simple optical measurement of the wear scar, which is relatively large, typically 1 to 3 mm in diameter. The worn region is much shallower than in conventional abrasive wear tests, typically 3 to 30 µm deep. If the specimen is coated, the method can also provide wear measurements for both coating and substrate material. For the test to yield useful results, the conditions must be closely controlled and the data provided from the test must be analysed in appropriate ways. Significant advances in both these aspects have been made recently (Allsopp et al. 1998, Hutchings 2000). Apparatus in which the ball is driven by a motor and the sample is loaded against it by a dead weight is commercially available and is shown in Figure 7.

By varying the load and the concentration of abrasive particles in the suspension, the particle motion can be changed from two-body, sliding abrasion to three-body rolling abrasion (Trezona et al. 1999). Under the conditions used for abrasion testing the hydrodynamic film produced by the motion of the sphere in the liquid is much thinner than the size of the abrasive particles and the load is supported almost completely by the particles.

For some applications the micro-scale abrasion test has significant advantages over more conventional methods of abrasion testing. It involves a very shallow depth of penetration, and also tests a small area of the sample. It can therefore be applied to specimens, and for purposes, where conventional tests could not be used, and is attractive as an effectively non-destructive quality assurance method for coatings. The method offers potential for development into a standard tribological test. A modified version, in which the ball is replaced by a cylindrical disc, has also been proposed which can be used to compare the abrasivity of different particles (Kelly et al. 2001).

5. Conclusions

Abrasion is responsible for many practical instances of wear. It is also usefully employed in the manufacturing processes of grinding, lapping and polishing. A firm understanding of the underlying tribological principles is required to control abrasive wear, and also to realise the potential of these manufacturing processes. There are still many challenges in these areas. The Preston/Rabinowicz equation, which provides a starting point for modelling abrasion processes, remains a very simple and largely empirically-based relationship. It is still not possible to provide a better model based on a description of the deformation and removal of material by the action of individual particles, and more accurate models usually just involve more complex empirical functions. There are particular problems in the effective description of particle properties, especially shape, and also of the properties developed near the surface of the workpiece material during steady-state abrasion. Both tribochemical processes and chemical influences on the fine-scale fracture of brittle materials are undoubtedly important in chemical-mechanical polishing and in the grinding of glasses; these too cannot readily be modelled from a fundamental basis. A role is likely to remain for many years for the careful experimental investigation of abrasion processes, and the further development of well-controlled experimental methods will play a vital part in this.

6. References

Allsopp, DN, Trezona, RI and Hutchings, IM, 1998, The effects of ball surface condition in the micro-scale abrasive wear test. Tribology Letters, vol. 5, 259-264.

Archard, JF, 1953, Contact and rubbing of flat surfaces. J. Appl. Phys. vol. 24, 981-988.

ASTM, 1999, Standard terminology relating to wear and erosion, ASTM Standard G40-99, American Society for Testing and Materials, West Conshohocken, PA, USA, 1999.

ATM, 2000, Standard test method for measuring abrasion using the dry sand./rubber wheel apparatus, ASTM Standard G65-00, American Society for Testing and Materials, West Conshohocken, PA, USA.

Beilby, GT, 1903, Surface flow in crystalline solids under mechanical disturbance, Proc. Roy. Soc. Lond, vol. A72, 218-225;

Beilby, GT, 1914, Transparence or translucence of the surface film produced in polishing metals. Proc. Roy. Soc. Lond., vol. A89, 593-595

Beilby, GT, 1921, Aggregation and flow of solids, Macmillan, London

Bhushan, B, Israelachvili, JN and Landman, U, 1995, Nanotribology: friction, wear and lubrication at the atomic scale. Nature, vol. 374, 607-616.

Buijs, M and Korpel-Van Houten, K, 1993, A model for lapping of glass. J. Mater. Sci., vol. 28, 3014-3020.

Dierickx, P, Enard, D, Geyl, R, Paseri, J, Cayrel, M and Beraud, P, 1996, The VLT primary mirrors: mirror production and measured performance, ESO report.

Dowson, D, Taheri, S and Wallbridge, NC, 1987, in Wear of Materials 1987, K C Ludema (ed.), ASME, New York, 415-425.

Eyre, TS, 1981, Wear mechanisms. Powder Metallurgy, vol. 57, 57-63.

Geltink-Verspui, MA, 1998, Modelling abrasive processes in glass, PhD dissertation, TU Eindhoven, The Netherlands

Hambery, M and Alean, J, 1992, Glaciers, Cambridge University Press

Hutchings, IM, 1992, Tribology: friction and wear of engineering materials, Arnold, London

Hutchings, IM, 2000, Applications of the micro-scale abrasion test for coatings and bulk materials. in Proc. 26th Leeds-Lyon Symposium on Tribology, Dowson, D et al. (eds.), Elsevier Tribology Series vol. 38, 37-42.

Kahlman, L and Hutchings, I M, 1999, Effect of particulate contamination on grease-lubricated hybrid rolling bearings, Tribology Transactions, vol. 42, 842-850.

Kelly, DA and Hutchings, IM, 2001, A new method for measurement of particle abrasivity. Wear (in press).

Li, SP, Lebib, A, Peyrade, D, Natali, M and Chen, Y, 2000, Microplow-row lithography and fabrication of submicrometer magnetic structures. Applied Physics Letters, vol. 77, 2743-2745.

Luo, Q, Ramarajan, S and Babu, SV, 1998, Modification of the Preston equation for the chemical-mechanical polishing of copper. Thin Solid Films, vol. 335, 160-167.

Menshutkin, BN, 1952, Russia's Lomonosov, Princeton University Pres.

Newton, I, 1704, Opticks

Preston, FW, 1927, The theory and design of plate glass polishing machines. Journal of the Society of Glass Technology, vol. 11, 214-256.

Puttick, KE, 1979, Energy scaling, size effects and ductile-brittle transitions in fracture. J Phys. D: Appl. Phys., vol. 12, L19-L23.

Rabinowcz, E, 1965, Friction and wear of materials, John Wiley,.

Rabinowicz, E, 1968, Polishing. Scientific American, vol. 218(6), 91-99.

Rayleigh, Lord, 1899, Polish. Proc. Royal Institution of Great Britain vol. 16, 563-570.

Rutherford, KL and Hutchings, IM, 1997, Theory and application of a micro-scale abrasive wear test. Journal of Testing and Evaluation, vol. 25, 250-260.

Samuels, LE, 1982, Metallographic polishing by mechanical methods, ASM, Metals Park, OH, USA.

Stevenson, ANJ and Hutchings, IM, 1996, Development of the dry sand/rubber wheel abrasion test. Wear vol. 195, 232-240.

Trezona, RI, Allsopp, DN and Hutchings, IM, 1999, Transitions between two-body and three-body abrasive wear: influence of test conditions in the micro-scale abrasive wear test. Wear, vol. 225-229, 205-214.

Williams, J A and Hyncica, A M, 2001, Mechanisms of abrasive wear in lubricated contacts, in Hydraulic Failure Analysis: Fluids, Components and System Effects, ASTM STP 1339, Totten, GE, Wills, DK and Feldmann, D (eds.), Amer. Soc. for Testing and Materials, W Conshohocken, PA, USA, 13-30.

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

COMPARISON OF WEAR RESISTANT MMC AND WHITE CAST IRON

Hans Berns
Ruhr University, IA 2-140, D-44780 Bochum / Germany
berns@wtech.ruhr-uni-bochum-de

Abstract: In this overview the microstructure of conventional white cast iron (WCI) and new metal matrix composites (MMC) are compared. In contrast to casting the hot isostatic pressing (HIP) of powder mixtures offers more freedom to design specific properties like toughness and wear resistance. Both may be enhanced by a proper MMC composition. Experimental results are so convincing that special industrial applications appear to be feasible as well as cost-effective and some have been initiated. New MMC for high temperature wear, for corrosive environments and for cold forging tools are presented. It is shown how the cost of hard particles in MMC may be reduced by an in situ transformation of ferroalloy particles .

Keywords. *white cast iron, metal matrix composite, MMC, wear resistance, fracture toughness*

1. Introduction

For decades white cast irons (WCI) have been the work horses of wear protection in the mining and cement industry as well as in road construction where abrasion by mineral grains prevails. They are of hypo- to hypereutectic composition and consist of hard carbide phases (HP) embedded in a hardenable metal matrix (MM). Both constituents develop from melt and vary in crystallographic structure and hardness depending on alloying and heat treatment.

Recently wear resistant metal matrix composites (MMC) have been developed, which allow to select hard particles (HP) like carbides, borides and nitrides and a metal matrix (MM) independently of each other and design microstructures of superior properties. In contrast to the solidification of castings in sand moulds close to phase equilibrium, the powder metallurgical (PM) production of MMC may to stay away from this condition.

In general the two constituents of wear resistant materials serve different purposes: The HP are to impede wear by grooving or indenting mineral particles while the MM is meant to provide sufficient toughness. Both properties depend on the amount, size and distribution of HP as well as on the hardness and fracture toughness of both constituents and the bond between them. Because of the comparatively high thermodynamic stability of the HP, material properties like hot hardness or corrosion resistance are expected to rely mainly on the MM: Berns (1998).

It is the aim of the present study to compare the microstructure and resulting properties of two groups of materials and to explain why the higher costs of MMC may well pay off.

2. Microstructure

White cast iron: The chemical composition and cooling rate determine the morphology of the as-cast microstructure. About 2 to 4 mass% of carbon control the amount of HP, while Cr, V, Nb influence their structure and hardness. In addition Mo and Ni assist the hardenability of the MM. The rate of solidification is given by the mould material, the cross-section of the casting and the location within. The faster the cooling, the smaller are the primary and eutectic phases. However, they are precipitated from the melt and are coarse compared to those formed in solid state. For simplicity let us call primary and eutectic carbides ″coarse“ and secondary ones ″fine“. Thus the coarse carbide structure is completed at solidus temperature, i.e. around 1150 °C at the latest. It stays about unchanged during further cooling or heat treatment.

In hypoeutectic irons a eutectic shell solidifies around primary dendrites of metal matrix. In a metallographic section this morphology resembles a carbide net (Fig. 1a). A near-eutectic composition leads to a more homogeneous distribution of eutectic carbides with occasional primary dendrites or carbides depending on local seggregation. In a hypereutectic alloy the solidification starts with the growth of primary carbides, which are subsequently surrounded by a eutectic net (Fig. 1b). Due to the higher temperature of formation the primary carbides are larger than the eutectic ones. In a sand casting of e.g. 50 mm thickness, the latter are in a 10 µm range, while some of the former tend to solidify in a needle-like shape with an aspect ratio of 10 and more, which in thicker parts may extend to the mm range. As soon as more than one eutectic appears in high alloy irons the constitution becomes more complicated and the above terminology is no longer applicable.

Starting from an Fe-C melt a grey solidification was turned into a white one by increasing the Mn/Si ratio or by chill casting to locally obtain a wear resistant surface while the remainder offered good machineability. The eutectic then consists of orthorhombic Fe_3C, which reaches a hardness of about 1000 HV 0.05 and tends to grow as an interconnected skeleton with the metal matrix included therein. In this case MM is somewhat misleading because actually the HP resembles the matrix of the eutectic in which the MM is embedded. In practical application the composition is near-eutectic, because primary carbides add to the brittleness of the eutectic skeleton and primary dendrites lower the resistance to grooving wear. These ″unalloyed“ irons usually develop a perlitic MM and a hardness of ≈ 550 HV 30.

Next NiCr irons emerged, which were designed to harden during cooling in the mould. Here the carbides pick up some Cr, turn into M_3C of equal morphology and become harder. Ni and Cr increase the hardenability. Ni reduces the Ac_1 and the hardening temperature of the MM and lowers its DBTT. As a result of the martensitic and bainitic transformation the hardness increases to ≈ 700 HV 30. Because Ni supports grey solidification the Cr content has to counteract this tendency and the composition is adjusted to the cooling rate, i.e. to the cross-section.

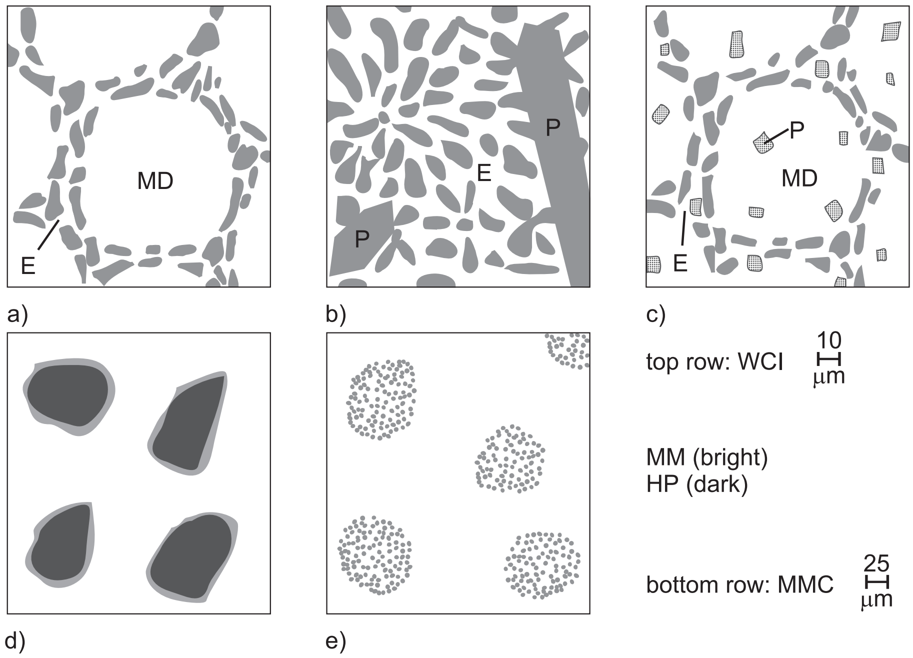


Figure 1. Microstructure of wear resistant materials (schematically), (a) hypoeutectic, high chromium WCI consisting of metal dendrites (MD) surrounded by a net-like eutectic (E) with HP of type M_7C_3, (b) as (a) but hypereutectic with primary carbides (P) surrounded by a eutectic E, (c) as (a)but with primary MC carbides (P) by alloying with V, Nb(Ti), (d) MMC of crushed HP dispersed in a MM, (d) MMC of HP clusters dispersed in a MM (double dispersion).

The following development targeted the increase of carbide hardness and the improvement of carbide morphology by adding from 8 to 26 mass% Cr at carbon contents between 2 and 3.5 mass%. The resulting M_7C_3 carbides stem from hexagonal Cr_7C_3 but may actually contain more Fe than Cr which lowers their hardness to the range of 1250 to 1650 HV 0.05. At the lower end of the Cr/C ratio of the alloys the M_7C_3 may be surrounded by an M_3C layer. The upper limit of this ratio is given by a decrease in the hardenability of the matrix. In contrast to a eutectic M_3C skeleton the M_7C_3 eutectic consists of individual HP embedded in a continuous MM. This change in morphology improves the forgeability of alloys on the low carbon side and the resistance to crack initiation in service at all carbon levels. Because of a high Cr-C-content the as-cast MM may consist of up to 100 % austenite, which is decomposed by annealing before hardening and tempering. Raising the hardening temperature increases the hardness of martensite but also the austenite/martensite ratio. Yet a macrohardness of ≈ 850 HV 30 is within reach.

These high chromium white cast irons may be further improved by adding Nb, V, (Ti) to precipitate primary cubic MC carbides of superior hardness (2000 to 3000 HV 0.05). They are of compact shape and are therefore used to replace elongated primary M_7C_3 in hypereutectic CrMo irons. In hypoeutectic ones they help to protect the MM dendrites within the eutectic net (Fig. 1c). They are also a means of increasing the Cr content of the MM up to the level of passivation, i.e. create corrosion resistant irons. For elevated temperature service secondary hardening is a means of aging the as-quenched martensite in the MM by a precipitation of CrMoV-carbides during tempering at 500 to 550 °C. The partial exchange of C by B leads to a higher HP hardness but is used preferably in hardfacing welding. A number of WCI are given in Table 1.

Metal matrix composites: The MMC of Table 2 consist of coarse HP dispersed in an MM (Fig. 1d) and were manufactured by hot isostatic pressing (HIP) to reach a fully dense state. The HP were crushed to an average size of ≈ 80 µm, which is roughly an order of magnitude above the size of eutectic HP in white cast irons. Agglomerated HP are excluded for lack of inner strength. The HP hardness (HV 0.05) was chosen above that of the hardest minerals, namely flint which is a type of quartz (≈ 1200) and corundum (≈ 2060). The following HP were used: Cr_3C_2 (≈ 2250), CrB_2 (≈ 2500), eutectoid WC/W_2C (≈ 2600) (Table 2). Atomized powders of hardenable bcc steels or a fcc Ni alloy were selected for the MM of MMC as well as for reference specimens without coarse HP. After HIP, pure HSh and CWS contain about 30 vol% of fine globular carbides in the order of 1 μm in size. About 15 vol% of these fine HP are dispersed in HSS. Although they are predominantly of eutectic origin they fall into the size range of fine secondary HP because of the rapid quench during atomisation and subsequent globulization during HIP. The Ni superalloy is precipitation hardened by γ'-Ni_3Al for elevated temperature service.

To obtain a dispersion of the HP in the MM the two powders have to obey a size ratio which depends on their volume ratio (Fig. 2). Relatively fine HP would tend to surround the MM powder grains like satellites forming a brittle net-like structure, which has to be avoided. As the HP and MM are not in thermodynamic equilibrium the HIP conditions decide on the amount of interdiffusion between these partners. It is appropriate to keep the pressure high (100 to 200 MPa) and the temperature low (≤ 1100 °C) to achieve a good bond but not too much HP dissolution. At best thin layers of diluted phases like M_7C_3 around Cr_3C_2, MB/M_3B around CrB_2 or $M_6C/M_{12}C$ around WC/W_2C are formed (Fig. 1d).

Table 1. Grades of wear resistant white cast iron: WCI 1 = near-eutectic alloy with M_3C carbides and a perlitic MM, WCI 2 = hypoeutectic steel and WCI 3 to 6 near eutectic alloys with M_7C_3 carbides of increasing Cr content, hardened and tempered at ≤ 250 °C, WCI 7 = as before, but tempered at 515 °C to enhance secondary hardening for elevated temperature service.

designation	mean composition mass%	coarse HP vol%,	type,	HV0.05	retained austenite,vol%	hardness HV30	W_{ab}^{-1} flint[1)]	10^4 Al_2O_3[1)]
WCI 1	Ni2Mo0.5C3.4	48	M_3C	1130	0	500	7.4	3.1
WCI 2	Cr12C2	15	M_7C_3	1450	25	780	4.5	2.7
WCI 3	Cr9Ni5Si2C2.9	32	M_7C_3	1250	15	740	5.4	3.1
WCI 4	Cr8Ni5Si2C3.2	38	M_7C_3	-	20	710	8.4	3.3
WCI 5	Cr20Mo1C2.8	30	M_7C_3	1500	10	790	10	3.8
WCI 6	Cr26Mo1C3.2	33	M_7C_3	1650	≤3	790	17	4.1
WCI 7	Cr20Mo6Si3Ni2C3.4	28[2)]	M_7C_3	1700	25	730	10	3.3

[1)] 80 mesh abrasive paper [2)] plus M_6C eutectic of 900 HV

Table 2. Materials used for PM steels and wear resistant composites: FDS = forging die steel, CWS = cold work tool steel, HSS = high speed steel, HSh = as before but of high carbon content, all hardened and tempered, NiA alloy solution annealed and aged at 750 °C.

materials		properties after HIP and heat treatment			
MM	powder composition mass%	hardness HV0.05	HV30	$W_{ab}^{-1} \cdot 10^4$ flint[1)]	Al_2O_3[1)]
FDS	Ni1.7Cr1.1Mo0.5V0.1C0.56		660	2.7	2.3
CWS	Cr13V4Mo1C2.3		740	4.1 (9.1)	2.9 (4.3)
HSS	W6Mo5V3Cr4C1.3		880	5.0 (10)	3.1 (3.6)
HSh	W6Mo7V6Co11Cr4C2.3		680	4.8	2.8
NiA	NiCr20Al4Si3		350	1.8 (2.3)	1.8 (2.1)
HP	powders (≈ 80 μm)				
Cr_3C_2		2250			
CrB_2		2500			
WC/W_2C		2600			
MMC		30 vol% HP			
FDS + CrB_2			650	34	4.7
CWS + WC/W_2C			1040	391(1735)	21 (62)
HSS + WC/W_2C			1030	313	19
			830	69	9.8
NiA + WC/W_2C			570	32(455)	6.7 (28)
NiA + Cr_3C_2			590	42	4.5

[1)] abrasive paper of 80 mesh (180 μm) and in brackets 220 mesh (67 μm)

3. Properties

Toughness: This property is important on a microscopic level in the wear surface (microcracking along the rims of wear grooves) and on a macroscopic one (resistance of wear parts to brittle fracture). There are a number of ways to measure the toughness, i.e. the fracture energy, of which the notched impact test does not differentiate sufficiently. Slow bending of smooth specimens is often used for hard materials. However, in the present ones the fracture starts with cracking of the HP and the mechanical properties depend on the length of these cracks, that means on the size of the HP. This is schematically shown in Fig 3a: The bending strength decreases as the particle size is raised. Therefore it seems to be reasonable to measure the fracture toughness, which - at a given HP content - is improved by a growing HP size corresponding to a larger HP spacing (Fig. 3b). At a given size of the stressed zone in front of the crack tip, small HP are bound to fracture leading to microcracks ahead of the main crack, which is thereby destabilized. The larger spacing between coarse HP may locally confine the stressed zone to the tough matrix and increase $K_{Ic.}$ Pre-cracked specimens were fractured in three-point-bending and the K_{Ic} values are compared to K_{Id} results of impact tests with sharply notched (EDM cut) specimens reported by: Zum Gahr, Doane (1980). In Fig. 4 the fracture toughness is plotted over the HP content and in Fig. 5 over the macrohardness.

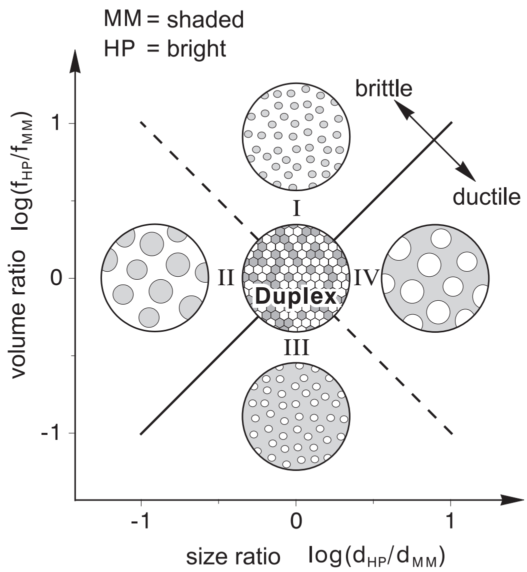


Figure 2. Microstructure of MMC derived from mixtures of near-globular HP and MM powders of volume content f and size d. I, II = brittle dispersion of MM in HP, III, IV = ductile dispersion of HP in MM.

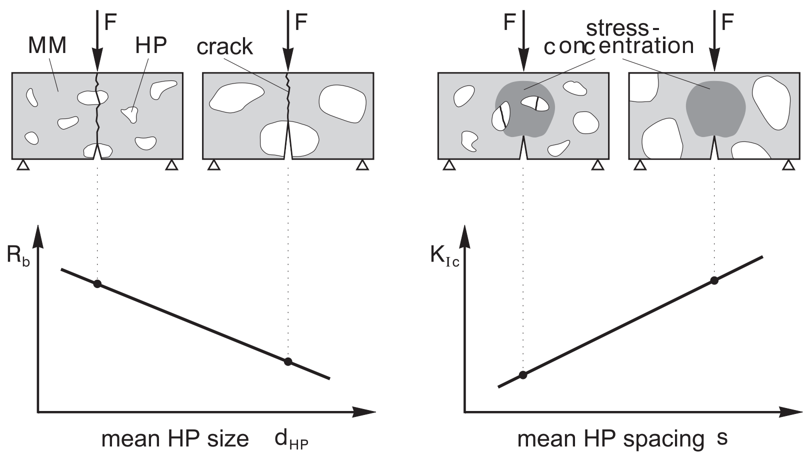


Figure 3. Schematic representation of bending strength R_b and fracture toughness K_{Ic} in dependence of the HP size and spacing.

While these macroscopic values were measured in bending, the indentation method by Palmqvist and others was used to derive the fracture toughness of HP and abrasive particles (AP). The eutectoid structure of WC/W_2C gives an exceptional high product of K_{Ic} and H_{HP} (Fig.6).

Wear resistance: To be most effective reinforcing HP have to be harder and tougher than the AP and at least as large as the groove width of abrasion. If a wear system changes from erosion-like conditions to severe grooving wear, as in mineral extraction and handling, large HP of superior hardness become important. In our pin-on-plate test the end face of a specimen 6 mm ∅ was moved back and forth in parallel trails at a speed of 4.8 mm/s under a normal pressure of 1.32 MPa over fresh abrasive paper while slowly rotating. The wear resistance $W_{ab}^{-1} = \rho \cdot L \cdot A/\Delta m$ was derived from the density ρ, the total length L of the wear path (< 200 m), the contact area A and the mass loss Δm. Results of WCI are given in Table 1 and those of PM matrices and of MMC in Table 2. WCI and MMC with different contents and types of HP are compared in Fig. 7.

4. Comparison

To weigh up different wear resistant materials one has to consider their expected performance in service as well as the feasibility of manufacturing and cost effectiveness.

Performance: The target is a high wear resistance combined with sufficient toughness for service under conditions of severe grooving wear by natural minerals of which quartz is among the hardest. Therefore 80 mesh flint offers a good guideline and finer or harder abrasives are not considered any further. Comparing FDS and HSh the 30 vol% of fine HP (1 µm range) of the latter increase W_{ab}^{-1} by a factor of only 1.8 (Table 2). The same amount of coarse HP (≈ 80 µm) in FDS + 30 vol% CrB_2 yields, however, a factor of 12.6. The factor between FDS and WCI 5 containing 30 vol% of coarse eutectic M_7C_3 (10 µm range) is 3.7 and thus in between the previous two. The best two of the MMC are superior to the best two of the investigated WCI by a factor of 30 to 40 (Fig. 7).

A fine HP size and high hardness as revealed by the PM steels HSS and HSh lead to a poor fracture toughness (Fig. 5), which is not compensated by a superior wear resistance (Table 2). For HSS the product of K_{Ic} and W_{ab}^{-1} is 55 (omitting 10^4). In comparison this product is 950 for the MMC FDS + 30 vol% CrB_2 and 225 for the near-eutectic WCI 4. In all, the performance in respect to wear resistance and toughness is considerably improved in the order of hard steels → white cast irons → MMC.

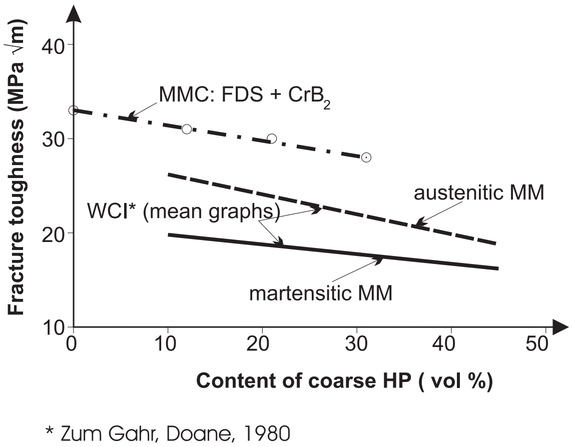


Figure 4. Fracture toughness in dependence of the HP content: WCI with (mass%) 12 to 26Cr, 2.4Mo, 1.4 to 3.9C containing M_7C_3 carbides, tested in the as-cast, predominantly austenitic state (400 to 550 HV50) and after hardening from 900 °C, deep freezing and tempering at 200 °C to a predominantly martensitic state (700 to 900 HV50): Zum Gahr, Doane, 1980. MMC as listed in Table 2.

Feasibility: Compared to casting the size and shape is more difficult to control during HIP which entails more machining. In general there is a risk of cracking during manufacturing of hard materials. WCI may crack in the mould or during hardening. The latter may also happen to an MMC especially if it is combined with an unsuited substrate. However, an even dispersion of coarse HP in an MMC is supposed to improve the crack resistance over an uneven dispersion or interconnected array of HP in WCI. While the microstructure of WCI depends on the cross section, that of MMC does not. If wear parts have to be partially machined, the very hard HP in MMC require appropriate machining conditions. Joining the MMC to an easily machinable substrate may ease the problem.

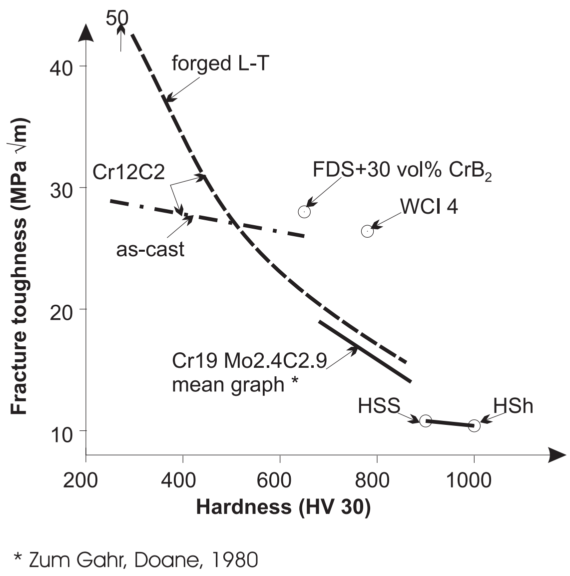


Figure 5. Fracture toughness in dependence of hardness. The near-eutectic white cast iron Cr19Mo2.4C2.9 and the MMC FDS+CrB_2 contain about 30 vol% of coarse HP, the PM high speed steel HSh about the same amount of fine ones. There are about 15 vol% of HP in the hypoeutectic Cr12C2 (coarse) and in the HSS (fine). The specimens were hardened and tempered to the hardness level given.

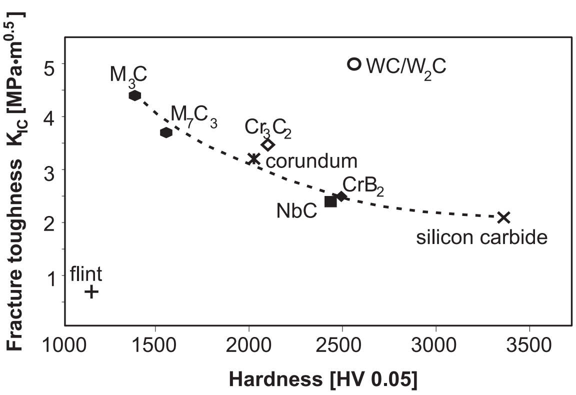


Figure 6. Fracture toughness of HP and AP derived from cracks initiated by Vickers indentations.

The manufacturing costs of PM-HIP are considerably above those of sand casting and require a superior performance of MMC to make this new technology cost-effective. In some applications a change of worn-out parts causes a shut down of a whole plant and these in-service costs are reduced by a better performance. A high reliability and availability is a key goal of production lines e.g. in crushing and compaction of minerals. HIP-cladding of tough steel parts adds to the fracture resistance in service, reduces the costs by limiting the MMC volume and is prone to raise the performance above the WCI level. Crusher rings are a good example of feasible MMC products, if performance, manufacturing and costs are jointly considered. HIP facilities of up to about 1.5 m diameter and a loading weight of about 15 t are available in the market.

5. MMC for special applications

During the solidification of a WCI casting the HP and MM are formed interdependently, which limits the possibilities of controlling their individual composition, size, distribution and properties. An MMC allows more variance of microstructural design to achieve specific properties besides a good abrasion resistance.

High temperature wear resistance: Secondary hardening of CrMoWV irons (e.g. WCI 7, Table 1) is used to improve the wear resistance in elevated temperature service. However, the better hot hardness of the MM is accompanied by eutectic HP. In contrast a matrix of higher hot hardness (HSS, NiA) may be combined with coarser, harder and more evenly dispersed HP (WC/W_2C, Cr_3C_2) to form MMC of superior high temperature wear resistance. This was confirmed in ring-on-disc experiments with loose abrasives of 80 µm in between at temperatures up to 900 °C: Berns, Franco (1997), Berns, Koch (1999) The contact pressure was set at 0.83 MPa and the rotational speed at 28 mm/s over a total length of 50 m to give the resistance to sliding abrasion W_{sa}^{-1}. While in air wear and oxidation interact, the latter is excluded in argon atmosphere. Protective layers of disintigrated AP, wear debris and oxides were formed on the wear surface, which increased W_{sa}^{-1} up to a temperature T_R at which the recrystallisation of the MM lowered the support of the layers. The MMC with a very hard HSS matrix outperformed the other materials in Ar up to about 600 °C, while the maximum wear resistance of the NiA materials was reached at about 700 °C (Fig. 8). The HP content is most beneficial at 900 °C. In spite of the high Mo content WCI 7 seems to be quite inferior to the MMC. Three NiA materials were also studied in air and - in the temperature range of 600 °C to 800 °C - show a distinctly lower wear resistance compared to runs in Ar. Above 600 °C WC/W_2C started to deteriorate by oxidation while Cr_2C_2 served well up to 900 °C. An assessment of our results leads to the conclusions that MMC offer a new level of wear restistance for parts in crushing, seaving and compacting of hot sinter and dust in the steel industry.

Table 3. Tool life in bolt making

material of die insert	hardness (HV30) die [3]	wire [4]	number of bolts [5]	increase by factor
W6 Mo5Cr4V2C0.9 [1]	640	157	9640	
		187	2800	
HWS+60 vol% HSh[2]	878	157	78000	8.1
		187	17500	6.3

[1]conventionally cast and hot worked high speed steel, [2]HWS = hotwork tool steel Cr5Mo1V1Si1C0.4
[3]hardened and tempered, [4] steel with (mass%) 0.19C and 0.005B annealed to 157 (HV30) or cold drawn to 187 (HV30), [5]12 mm ∅, cold headed from wire by die inserts as given

Table 4. Abrasive wear resistance of in situ MMC compared to reference materials

	hardness [3] HV30	wear resistance $W_{ab}^{-1} \cdot 10^4$ Al_2O_3 [4]	SiC[5]
FDS + 10vol% in situ TiC[1]	730	5.5	3.1
FDS + 10vol% CrB_2[1]	660	4.6	2.2
MSS + 50vol% TiC[2]	790	2.0	1.3

[1] HP size ≈ 80 µm, [2] Martensitic stainless steel with fine TiC (1 to 4 µm), [3] hardened and tempered
[4] 80 mesh, 2060 HV0.05, [5] 80 mesh, 3000 HV0.05

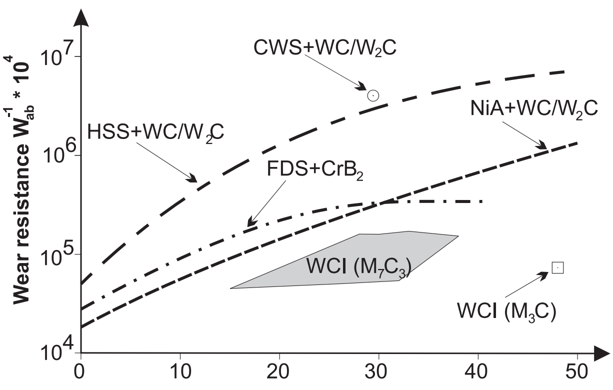


Figure 7. Wear resistance against 80 mesh flint in dependence of the HP content: four MMC are compared to WCI with M_7C_3, respectively M_3C carbides.

Corrosion resistance: The strong chemical bond within the HP enhances their corrosion resistance. The MM has to be protected by passivating elements. Powders of hardenable stainless steels with (mass%) ≥ 13 Cr and ≥ 0.35 C are suitable. It is well known, though, that the corrosion resistance of stainless martensite is considerably improved, if the greater part of C is replaced by N: Gavriljuk, Berns (1999). Therefore a (mass%) Cr15Mo1C0.04 matrix powder was mixed with CrN particles. During HIP the reaction $2CrN \rightarrow Cr_2N + N$ yielded N atoms which were dissolved in the MM and provided a high hardenability and resistance to aqueous corrosion. The passive current density of the MMC Cr15Mo1 + 10 vol% CrN in 1n-H_2SO_4 was even below that of the stainless steel grade Cr17Mo1C0.35, which does not contain coarse HP and wears much faster: Berns, Wang (1993). Even more convincing is the resistance of the MMC to pitting by $^{-}$Cl ions.

Fracture toughness: In section 3 it was shown that an increase of the HP size simultaneously raises the fracture toughness and the abrasive wear resistance . The bending strength is bound to decrease, though (Fig. 3a).

Tools in metal forming live on the crack-free service period and therefore require a high strength. This led to the idea to replace the coarse HP dispersed in the MM by coarse particles with a high content of fine HP dispersed in an MM (Fig. 1e). This double dispersion microstructure may be viewed as clusters of fine HP providing wear resistance, surrounded by MM without HP to enhance the fracture toughness. It was realized by mixing 60 vol% of HSh powder of ≈ 80 µm size containing about 30 vol% fine, dispersed carbides and 40 vol% hot work steel powder (HWS) of 16 µm mean size. According to Fig. 2 and in reality a double dispersion microstructure was obtained after HIP. Die inserts of (mm) 33 ∅ · 12 ∅ · 10 in size were machined, heat treated and used to form bolts by cold extrusion of steel wire: Berns, Nguyen (1996), Berns et al. (1998). In Table 3 the life of double dispersion die inserts is compared to that of conventional high speed steel, the hardness of which had to be kept rather low to avoid cracking. The improvement is quite substantial, because the fracture toughness of HWS + 60 vol% HSh is K_{Ic} = 16 MPa √m. Double dispersion or clustered MMC appear to be a means of raising toughness while a high level of wear resistance is maintained.

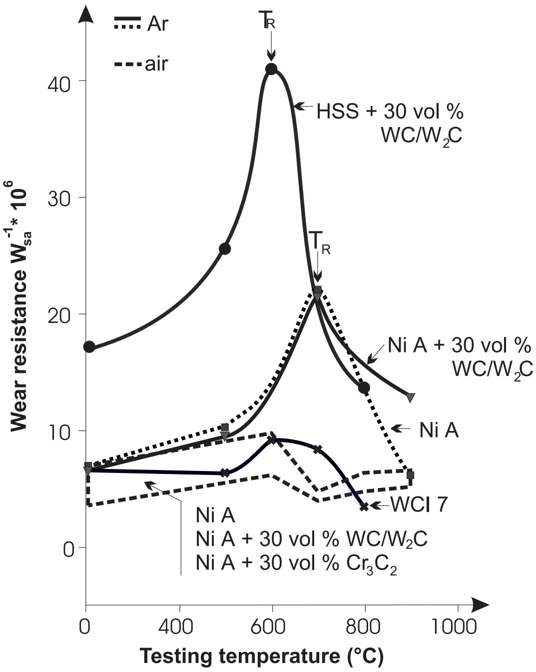


Figure 8. Resistance to sliding abrasion W_{sa}^{-1} of MMC and WCI at elevated temperatures against 80 µm flint.

Low-cost HP: The good properties of WC/W_2C (Fig. 6) are partly canceled out by its high price and density, which make an MMC with 30 vol% HP rather expensive. Therefore the coarse HP powder was replaced by a ferroalloy powder like FeTi with 70 mass% Ti. Upon adding graphite to the MM + FeTi powder mix an ″in situ“ transformation of the FeTi particles to TiC took place during HIP. Within a particle the TiC formed a hard case (2600 to 2750 HV 0.025) and the iron was enriched in the core (900 to 1500 HV 0.05). First results of wear tests carried out with FDS + 10 vol% in situ TiC are presented in Table 4 and compared to reference MMC: Berns, Wewers (2001). The higher hardness of in situ TiC compared to CrB_2 improved W_{ab}^{-1} especially against SiC, the hardest AP used. Compared to pure MM without coarse HP (Table 2) the high content of fine TiC in a commercial MMC actually reduces W_{ab}^{-1}. The reason is seen in the embrittling effect of the dense population of fine HP in front of coarse grooving AP. The material density increases in the order of FeTi, Cr_3C_2, FDS, WC/W_2C. Taking into account the market price for a lot of 250 kg the relative powder cost amounts to 1, 4, 2.3, 50 in the same order. In situ MMC are a promising way to reduce the HP cost.

*

The work of K. Al-Rubaie, C. Broeckmann, M. Buschka, S. Franco, W. Hänsch, S. Koch, O. Lüsebrink, C. v. Nguyen, W. Theisen, W. Trojahn, T. Ümit, B. Wewers carried out at the Ruhr University is gratefully acknowledged.

*

6. References

Berns, H.(ed.), 1998, ″Hard alloys and composites“, Springer Verlag, Berlin (in German)

Berns, H. and Franco, S.D., 1997, ″Effect of coarse hard particles in high temperature sliding abrasion of new metal matrix composites, Wear 203-204, pp. 606-614

Berns, H. and Koch, S. 1999, ″High temperature abrasion of a nickel-base alloy and composite“, Wear 225-229, pp. 154-162 and ″Influence of abrasive particles on wear mechanism and wear resistance in sliding abrasion tests at elevated temperatures“, Wear 233-235, pp. 424-430

Berns, H., Melander, A.,Weichert, D., Asnafi, N., Broeckmann, C., Groß-Wege, A. 1998, ″A new material for cold forging tools“, Computat. Mat. Sci. 11, pp.166-180

Berns, H. and Nguyen, v.C., 1996, ″A new microstructure for PM tooling material“, Met.Phys.Adv.Tech. 16, pp. 693-706

Berns, H. and Wang. G., 1993, ″Stainless martensitic PM-HNS“, Proc. of the High Nitrogen Steels (HNS) 93, Kiev/Ukraine, pp. 415-419

Berns, H. and Wewers, B. 2001, ″Development of an abrasion resistant steel composite with in situ TiC particles“, Conf. Wear of Materials, Vancouver, April 2001 to be published in Wear

Gavriljuk, V.G., Berns, H., 1999, ″High nitrogen steels“ Springer Verlag, Berlin

Zum Gahr, K.H. and Doane, D.V., 1980, ″Optimizing Fracture Toughness and abrasion resistance in white cast irons“, Metall. Trans. 11A, pp. 613-620

REFRIGERATION AND AIR CONDITIONING SYSTEMS FOR THE NEW MILLENNIUM UNDER THE IMPACT OF ENVIRONMENTAL CHALLENGES

Prof. Dr.-Ing. Dr. h.c. H. Kruse
Forschungszentrum für Kältetechnik und Wärmepumpen GmbH, Weidendamm 12-14, 30167 Hannover, Germany,
e-mail, e-mail@fkw-hannover.de

Abstract. The Refrigeration and Air Conditioning Industry has been affected heavily since 1988 by the Montreal Protocol on Substances that deplete the Earth's Ozone Layer in a conditioning phase out process. It will be affected in the future further on by the Kyoto Protocol on Greenhouse Gases.
Therefore, the emission reduction of fluorocarbon refrigerant will be an important task in the new millennium causing a redesign and better maintenance procedures of system concerning tightness or the application of low global warming refrigerants. The different measures for various application of refrigeration and air conditioning systems are discussed and future developments are printed out.

Keywords. *Ozone Depletion, Global warming, Refrigerants Refrigeration and Air-conditioning systems Domestic and commercial refrigeration, mobile and unitary air conditioning, chillers*

1. Introduction

The global environmental issues of ozone depletion and global warming affected the refrigeration, air conditioning and heat pump industry considerably over the last ten years.

The Montreal Protocol has caused the refrigeration industry to introduce newly developed hydrofluorocarbons (HFC) as substitutes for the chlorine containing ozone depleting chlorofluorocarbons (CFC). This process for new systems has been finalised in the industrialized countries and is still ongoing in developing countries. Also the hydrochlorofluorocarbons (HCFC) with much lower Ozone Depleting Potential (ODP) (Figure 1) like R-22 are foreseen to be phased out within the next 20 years worldwide whereas national regulations require a much earlier phase out date especially in European countries. Therefore, the chlorinefree hydrofluorocarbons (HFC) will take over in the next years the market of the chlorine containing refrigerants.

So, HFC-134a as a CFC-12 substitute currently fulfils an important role in almost all refrigeration and air conditioning sectors, from domestic refrigeration to large size chillers, and particularly in mobile air conditioning.

HCFC-22 still is currently the most important refrigerant used in refrigeration and air conditioning worldwide. The Montreal Protocol as well as the EC Regulation control already the production and offer for sale of R-22 and other HCFCs. Both regulations give a phase-out schedule for the stepwise reduction of the permitted quantities of HCFCs, whereby the EC Regulation gives a much shorter phase-out schedule.

In addition, in some European countries exist national regulations. For example, Sweden already had prohibited the use of R-22 and HCFC in new equipment since 1998. In Germany and Denmark, the use of R-22 in new equipment was permitted only until December 31, 1999.

The search for suitable substitutes for R-22 has shown that there is no single replacement which covers its wide application range and which possesses all the technical advantages which this refrigerant offers. Therefore main substitutes for HCFC-22 are nonflammable HFC-mixtures of R-32 or R-143a and R-125 with or without R-134a, as R-404A, R-507, R-407C and R-410A respectively. Further, the hydrocarbon propane (R-290) is discussed as a possible single fluid , but flammable substitute.

The Kyoto-Protocol for the reduction of greenhouse gases among which the hydrofluorocarbons HFC also contribute by their inherent Global Warming Potential (GWP) (Figure 1) to this second environmental effect is a further milestone on the way to alternate refrigeration systems. In the Kyoto-Protocol the HFC are nominated among the mentioned six gases(CO_2, CH_4, N_2O, HFC, PFC, SF_6) as one of the three industrial gases together with perfluorocarbons and sulphurhexafluoride. Therefore, emissions of the new refrigerants into the atmosphere have to be prevented in the future as far as possible.

In this respect Denmark seems to take the most stringent measures announced January 2000 by the Energy and Environmental Ministry to phase out the application of the HFC by the year 2006.

Further on, the United Kingdom (UK) stated in its Climate Change Program, drafted March 2000 by the Department of the Environment, Transport and the Regions, that "HFCs are not a sustainable technology in the long term" and that "HFCs should only be used where other safe, technically feasible, cost effective and more environmentally acceptable alternatives do not exist".

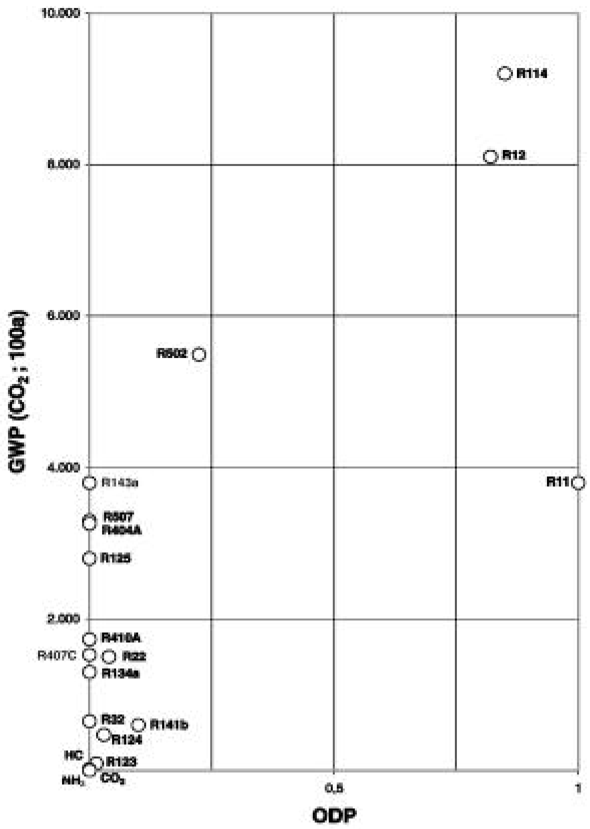


Figure 1. GWP and ODP of several refrigerants

These announcements have led to considerable uncertainty in Europe concerning the long term future of the HFCs as substitutes for the CFCs and HCFCs as refrigerants.

Figure 2 shows the worldwide refrigerant consumption in 1991 of 484.200 t/year used in the different application sectors (Montreal 1991).

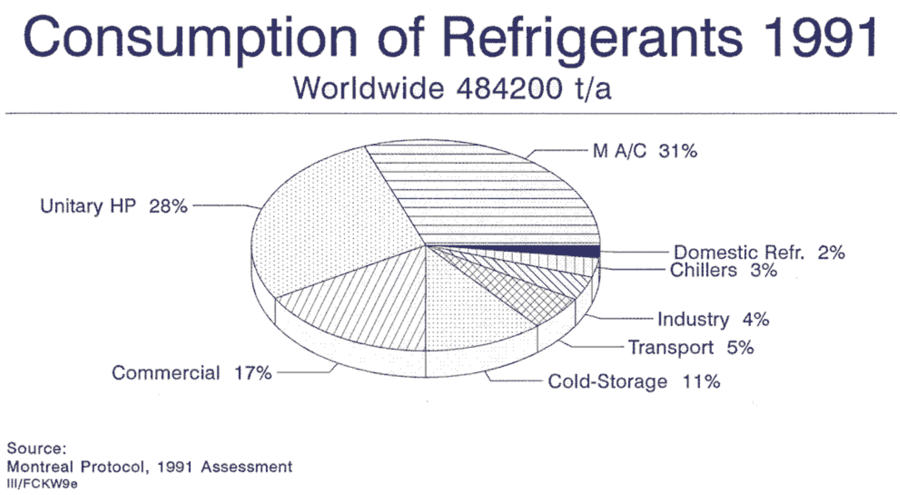


Figure 2. Worldwide Consumption of Refrigerants in 1991

This refrigerant consumption was mainly caused to one third by new systems and two third by after market service purposes, so that leakages may have been an important reason for the refrigerant consumption in the past.

Since the old CFCs and HCFCs and the new HFCs all have a remarkable Global Warming Potential an important task for the future is to reduce leakages and tighten refrigerant systems.

Another solution of the problem could be to use refrigerants with negligible Global Warming Potential (GWP) like the old refrigerants ammonia, hydrocarbons or carbon dioxide (Figure 1). Between these often called natural fluids ammonia has a zero Global Warming Potential due to its short lifetime in the atmosphere whereas hydrocarbons and carbon dioxide have a small but negligible Global Warming Potential.

Since refrigeration systems need energy supply they cause carbon dioxide emissions from fossil fuel power generation and therefore they also contribute indirectly to the greenhouse effect.

The sum of both effects per system, the direct from refrigerant emissions and the indirect caused by the energy to drive the refrigeration system is called Total Equivalent Warming Impact TEWI. The relationship between the direct and the indirect greenhouse effect is different for various applications of refrigeration systems as shown in Figure 3 (DOE 1991).

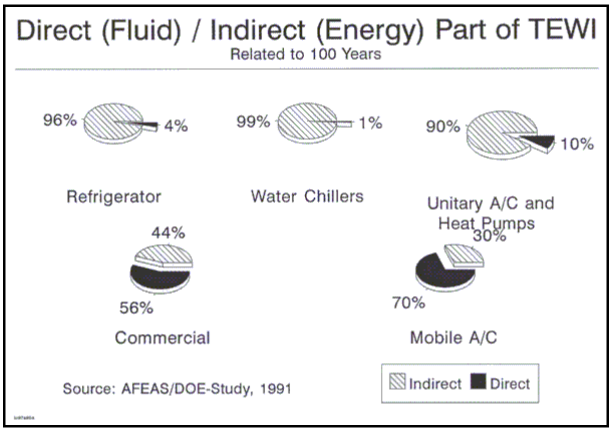


Figure 3. Direct and Indirect Parts of TEWI

2. Future Systems for various Applications

2.1 Domestic Refrigeration

According to Fig. 2 and Fig. 3 domestic refrigerators, because of their low refrigerant charge and their tight hermetic systems, consumed only 2 % of the worldwide refrigerant consumption in 1991 and contributed with their small charge and their foam blowing agent with only 4 % to the direct part of the Total Equivalent Warming Impact per system, whereas the rest of 96 % is caused by the energy consumption. Therefore, here mainly was the challenge for the new future systems and not concerning the refrigerant itself.

In contrary Greenpeace very early took the opportunity of the industrial change during German reunification to introduce 1993 hydrocarbon refrigerants via an East-German manufacturer and caused other West German manufactures to follow with a development from R 134a to Isobutane R 600a as a refrigerant and later to cyclopentane as foam blowing agent. This development was reported about already by the author in Brazil at the CIAR 1995 (Kruse 1993).

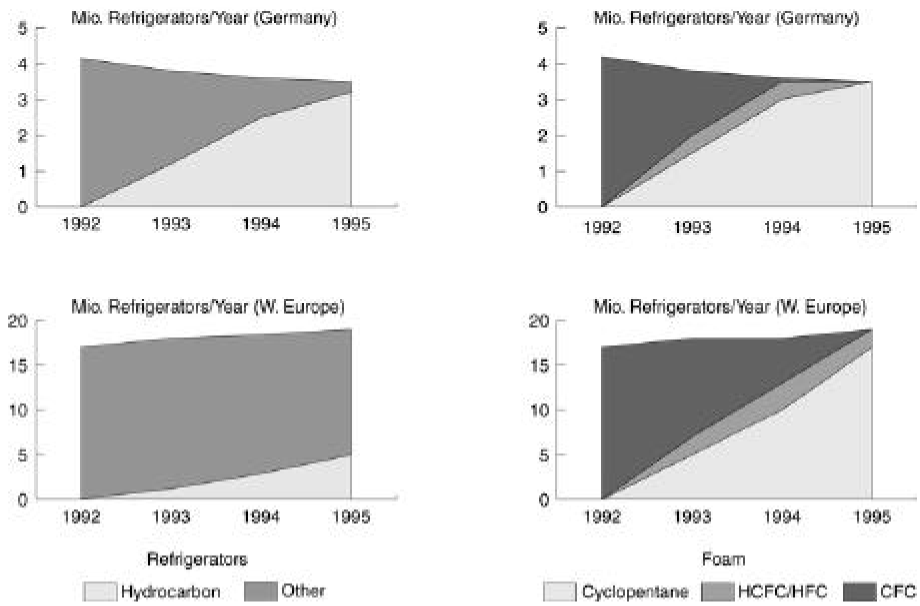


Figure 4. Refrigerator market and market share of hydrocarbon refrigerators and refrigerators with CFC, HCFC/HFC and cyclopentane foam in Germany and Western Europe between 1992 and 1995

Figure 4 reflects the market share of yearly produced refrigerators in Germany and Western Europe using hydrocarbons as refrigerant and foam blowing agent (FKW et al 1996). HC 600a is now being used in about 50 % of the refrigerators being produced in Europe and further production of hydrocarbon refrigerators has been set up in China and India.

More than 80,000,000 domestic refrigerators and freezers are produced annually worldwide.

Both refrigerants, HFC-134a and HC-600a, have demonstrated mass production capability for safe, efficient, reliable and economic use. The application of HC-600a or HFC-134a provides approximately equal efficiency. Other design parameters introduce more efficiency variation opportunities than is presented by the refrigerant choice.

HC-600a and HFC-134a continue to be the dominant alternative refrigerant candidates to replace CFC-12 in domestic refrigeration new equipment. Other alternative candidates have only regional niche appeal, primarily driven by established chemical production capability. Long term new equipment options will likely be limited to these two refrigerants.

The Kyoto Protocol also has increased consideration of energy efficiency initiatives to influence the secondary effects of greenhouse emissions from power generation and distribution in addition to direct product emission prescriptive regulations.

Practice of mature domestic refrigeration technology should provide products that require less than one-half of the electrical energy of the product being replaced. Improvements in compressor efficiency, heat exchange efficiency, insulation efficiency and construction techniques are the main contributors..

2.2 Commercial Refrigeration

From the worldwide refrigerant consumption in 1991 of 484.200 t/year the commercial refrigeration sector was the third largest refrigerant consumer with 17 % (Fig. 2).

In average commercial system contributed to the Total Equivalent Warming Impact TEWI 1991 with 56 % by the direct fluid effect due to the high leakage rate and with 44 % by the indirect energy effect (Fig. 3).

The types of equipment applied in commercial refrigeration are very different, where it concerns their size, the logistics of the specific refrigeration circuit, and the refrigerant charges applied.. Commercial refrigeration consists of four main types of equipment:

Stand-alone equipment covers many different types including vending machines, ice machines, etc. In summary, all kinds of small equipment that is installed in stores or public areas in many developing as well as in the developed countries. 10 to 12 million pieces of this equipment are in use globally. Refrigerant charges vary from 200 g up to 1 kg. The usual refrigerant applied is now HFC-134a, which has largely replaced CFC-12. In some European countries one has started applying hydrocarbons --in principle only HC-600a-- with charges up to 150 g, sometimes up to 800 g, dependent on national regulations. However, HFC-134a is the current dominant option for stand-alone equipment.

Condensing units are typically installed in specialized shops. The refrigerant charge varies between 1 and 5 kg. The number of equipment in use globally is estimated at about 2.5 million. The refrigerant choice is depending on the application temperature level. For medium low temperatures HFC-134a and for low temperatures still HCFC-22, as well as already HFC substitutes for R 22 are the preferred options.. Due to safety concerns, HCs are not a wide spread option for equipment where refrigerant charges are applied in the order of 1 to 5 kg.

Centralised systems can be found in supermarkets, where the estimated number is 120,000 globally; these systems have a wide range of refrigerating capacities. The refrigerant charge varies from 100 kg to about 1,500 kg. The refrigerating system is installed in machinery rooms and the refrigerant circulates from this machinery room to the display cases installed in the sales area.

A relatively new concept, called *"distributed system"*, drastically limits the length of refrigerant pipes applied because the compressors are installed in sound-proof boxes inside or nearby the sales area, which consequently leads to a substantial reduction of the refrigerant charge applied.

For supermarket systems the former CFC-502 has been replaced either intermediately by HCFC-22 or now already by its substitutes.

2.2.1 R-22 and its Substitutes

A thermo dynamical correct performance comparison for refrigerants with different properties in each optimized system is extremely difficult or nearly impossible.

An optimized adjustment of compressors and other equipment components to the different refrigerant properties is necessary. After optimization and adjustments for all possible replacements for R-22 a correct and final comparison of performance and capacity could be done. Of course, this ideal way of comparison is impossible in practice, but with the TEWI concept in mind this should finally be the procedure to select refrigerants environmentally.

In reality the theoretical comparison of fluids is normally the first partial step for an exact comparison, followed by refrigerant compressor performance tests and practical measurements at refrigerating installations.

Theoretical evaluations have been reported and discussed by the author formerly. (Kruse et al 1998)

Experimental comparison of R 22 Substitutes

When comparing the behavior of refrigeration systems there are in principle two possibilities called a) internal behavior and b) external behavior. The comparisons are related either to the temperatures inside the refrigeration cycle i.e. the evaporation temperature, the condensing temperature, the sub cooling temperature and the suction temperature of the compressor or the in- and outlet temperatures of the secondary fluids at the evaporator or the condenser respectively. When making a comparison concerning the internal behavior of the refrigeration system related to the refrigerant temperatures as mentioned then the heat transfer in the heat exchangers, evaporator and condenser, are not taken into account. In contrast, a comparison based on the external fluid temperatures includes the exergetic losses in the heat exchangers given by the temperature differences. So, a first comparison of various refrigerants in a refrigeration system can be done be measuring the internal behavior whereas a total evaluation with heat exchange effects should be done related to the temperatures of the external fluids using complete refrigeration systems.

Comparison according to the Internal Behavior

Experimental evaluations for substitutes of R-22 and R-502 have been performed at the Research Center for Refrigeration and Heat Pumps, Hannover (FKW) by calorimeter tests according to ISO 917 with the substitutes HFCs R-404A, R-407A, R-407B, R-407C, R-507 and R-410A as well as with the hydrocarbon propane R-290 (Arnemann et al 1995). Besides of other experimental results the most important is the energy consumption of the substitutes as compared to R-22. From the upper part of Figure 5 it can be seen that only R-410A and R-290 have an energetic behavior comparable to R-22, whereas the other fluids show a 5 to 15 % higher energy consumption like the old CFC R-502 mainly used in commercial refrigeration formerly.

Especially R-410A shows a very advantageous energetic behavior at low evaporation temperatures, which is even better than R-290. In the higher temperature range R-290 has a small advantage against R-410A. The flammable R-290 has a negligible global warming potential, whereas R-410A is non flammable with a similar global warming potential like R-22.

This is reflected in the direct effect of TEWI in the lower part of Figure 5.

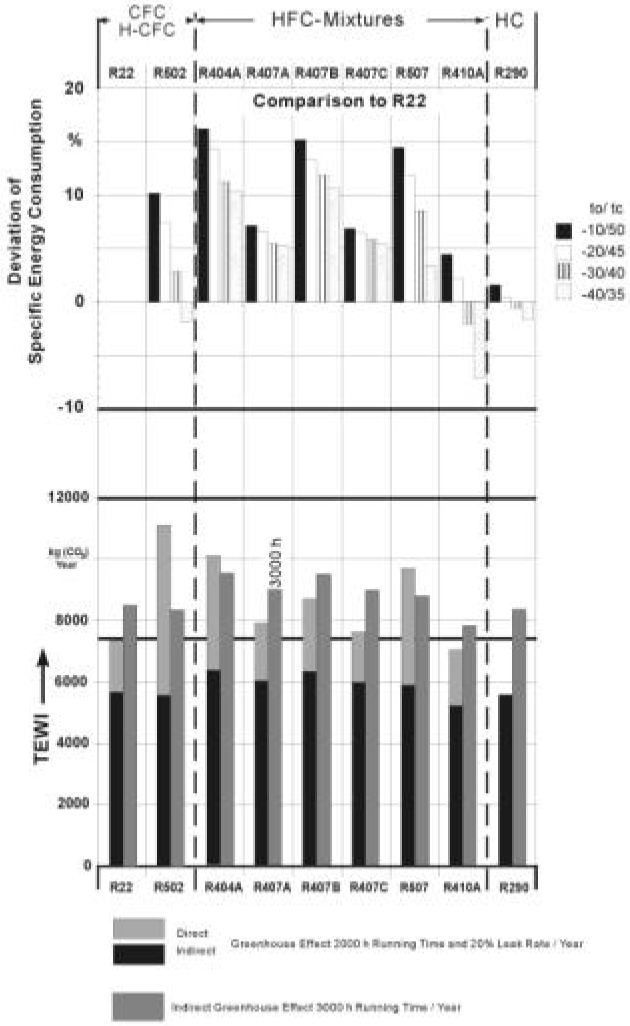


Figure 5. Experimental comparison of energy consumption / Direct and indirect part of the TEWI of R-22 and substitutes

Considering the emissions of CO_2 from the power plant, which contribute to global warming, it can be stated that both, R-410A and propane are slightly better in the indirect part caused by the energy consumption in comparison to R-22. The results shown in Figure 5 are based on a yearly leak rate of 20 % and annual operating hours of 2000 or 3000 hours, respectively.

Taking into account also the direct part caused by refrigerant emissions, then R-410A still has a slight advantage as compared to R-22, whereas R-290 is better under these assumptions, but implies a risk of flammability by the refrigerant emissions. This can partly be overcome by designing indirect systems with secondary fluids for the use of hydrocarbons, but then the energy consumption of such systems would be higher compared to systems with direct evaporation.

In conclusion from the experimental comparison of R-22 substitutes it can be stated that from an energetic and emission point of view R-410A would be the best choice between the non flammable fluids. No large differences more or less exist between the other options, R-407A, R-407B, R-407C, R-404A and R-507 whereby the last two options show the highest TEWI coming from the GWP of their mixtures component R-143a.

Comparison according to the External Behavior

Measurements at a certain refrigeration system concerning the experimental comparison of refrigerant blends as substitutes for R-22 were performed at the Institut für Luft und Kältetechnik Dresden (ILK Dresden) (Ahnefeld et al 1995). The experimental test rig was used to compare in addition to R-22, R-404A, the R-407 - series and R-507 as well as the refrigerant R-410A. When comparing refrigerants according to the external behavior as mentioned before it is necessary to keep the external parameters constant, namely the refrigeration capacity and the temperatures of the secondary fluids. The refrigerating capacity of R-410A is expected to be 50% higher than with R-22 if the compressor would be driven with the same revolutions per minute. Therefore the speed of the compressor was reduced when R-410A was tested to match the refrigerating capacity of R-22. For these tests a reciprocating compressor designed for R-22 was used. In the results for various refrigerants and evaporator inlet air temperatures the COP deviations from R-22 are given in Figure 6 as an example for these type of investigations. The condensing temperature in this example was $t_c = 33°C$.

It can be seen that R-410A would be comparable at low or be slightly advantageous at higher air inlet temperatures, whereas the other substitutes all show a disadvantage of 15 to 8 % respectively as compared to R-22.

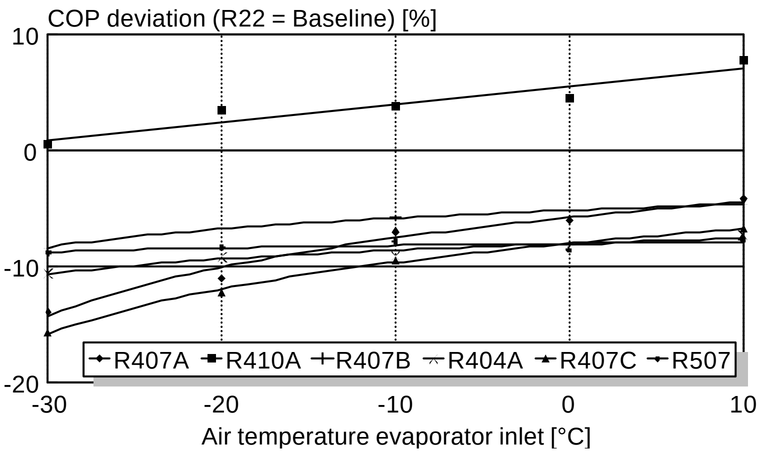


Figure 6. Comparison of R-22, R-410A and other HFC-blends with constant refrigerating capacity and external temperatures for different compressor speeds

The reason for higher COP with R-410 A compared to the theoretical results from section 2 are the high heat transfer of the component R-32 of the near azeotropic blend, the lower pressure drop in tubes and lower losses in the compressor.

For refrigeration and air conditioning systems with remarkable leakages like commercial refrigeration, mobile air conditioning and unitary air conditioning and heat pump systems the refrigerant to be used influences via its GWP and its COP the TEWI number of the respective system. These relations were presented in the first AFEAS/DOE-Study 1991 and refined 1997 in the so called AFEAS/DOE TEWI 3 Project which between others also covered refrigeration systems in Europe (DOE 1991, DOE 1997).

The TEWI 3 study for supermarket systems was performed with different refrigerants for current and future systems like direct evaporation systems as up to now mainly used, indirect cooling systems with a secondary refrigerant loop and distributed systems as favourized in the USA which are small water cooled and acoustically insulated systems distributed in the supermarket area and providing refrigeration via short liquid and suction lines to the cabinets.

The Total Equivalent Warming Impact (TEWI) was calculated for those systems with temperature levels in a supermarket under reasonable assumption for European conditions and also for various parts of the world. The results in Figure 7 show a large Total Equivalent Warming Impact for the conventional direct expansion systems with various refrigerants as compared to secondary loop or distributed systems for low temperature application.

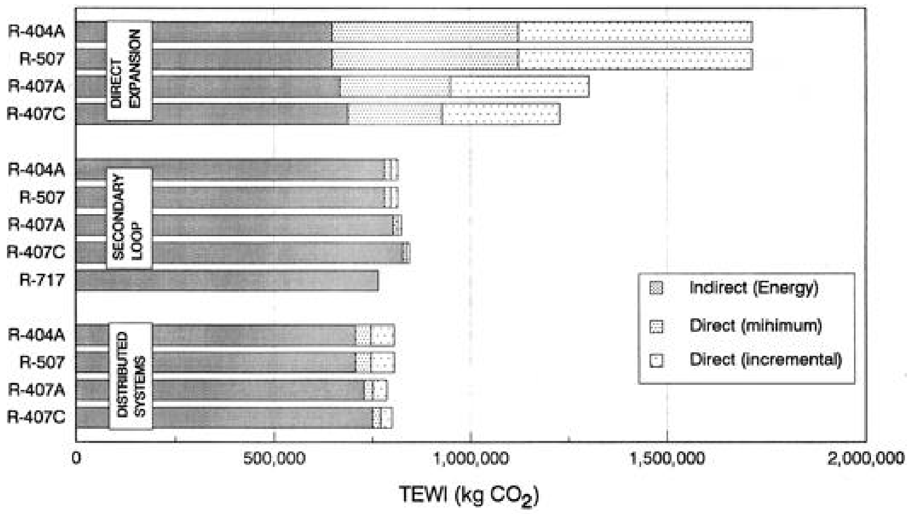


Figure 7. TEWI of low temperature refrigeration in Europe

In conclusion it must be stated that conventional supermarket refrigeration systems with direct expansion and a large HFC refrigerant charge and long refrigerant lines both with a potential to cause reasonable amounts of leakages cannot be the solution for the future if not the leakages can be decreased drastically. Therefore the choice for future HFC refrigeration systems in supermarkets are either indirect systems with secondary coolants or direct distributed water cooled HFC systems.

Another possibility would be the application of refrigerants with very low Global Warming Potential in direct systems. With the exception of carbondioxide the other mentioned natural fluids like ammonia and hydrocarbons show dangerous local behavior because of their flammability or toxicity and therefore are not suited for direct expansion systems inside public areas of supermarkets. So, also for these locally dangerous refrigerants indirect systems have to be provided. For direct systems with natural fluids only carbondioxide could be used which presents other problems, later to be discussed.

2.2.2 Indirect Refrigeration Systems

Theoretical Consideration

As can be seen from the results of the TEWI 3-Project (Figure 7) indirect refrigerant systems even though having a lower Total Equivalent Warming Impact cause higher energy consumption than conventional or distributed direct expansion supermarket systems which have an energy consumption between both.

Since the energy consumption because of the running costs is an important matter for the customers in order not to have higher electricity consumption than with direct expansion systems.

Therefore an important field of future research and development is to find suitable secondary refrigerants with low additional energy consumption as caused by the pumping power requirement for the coolant as well as the additional temperature differences in the primary/secondary refrigerant heat exchangers.

For various conventional secondary coolants as available on the market the pumping power per meter length had been calculated for pipes 35 x 1.5 mm. The heat transfer behavior of the coolants was evaluated by a characteristically number (Kruse et al 1997).

The fluid properties viscosity η, density ρ, specific heat capacity c and heat conductivity λ are important parameters for the pumping power as well as for the heat transfer characteristic. The results for low temperature cooling circuits in supermarkets are shown in Figure 8.

From Figure 8 it is evident that for the application of low temperature cooling the pumping power for hydrous salt solutions is nearly the same and in the order of the old calciumchloride brine whereas for the synthetic fluids much higher pumping powers are necessary. This is valid also for the pumping powers of the higher alcohols especially the Propyleneglycol.

Concerning the heat transfer properties which also influence the system energy consumption because of the resulting lower evaporation temperatures in the primary refrigerant the synthetic fluids and the glycols show also a less favorable behavior than the hydrous salt solutions. In the low temperature application HYCOOL 50 and Tyfoxit F40 are the best fluids followed by calciumchloride.

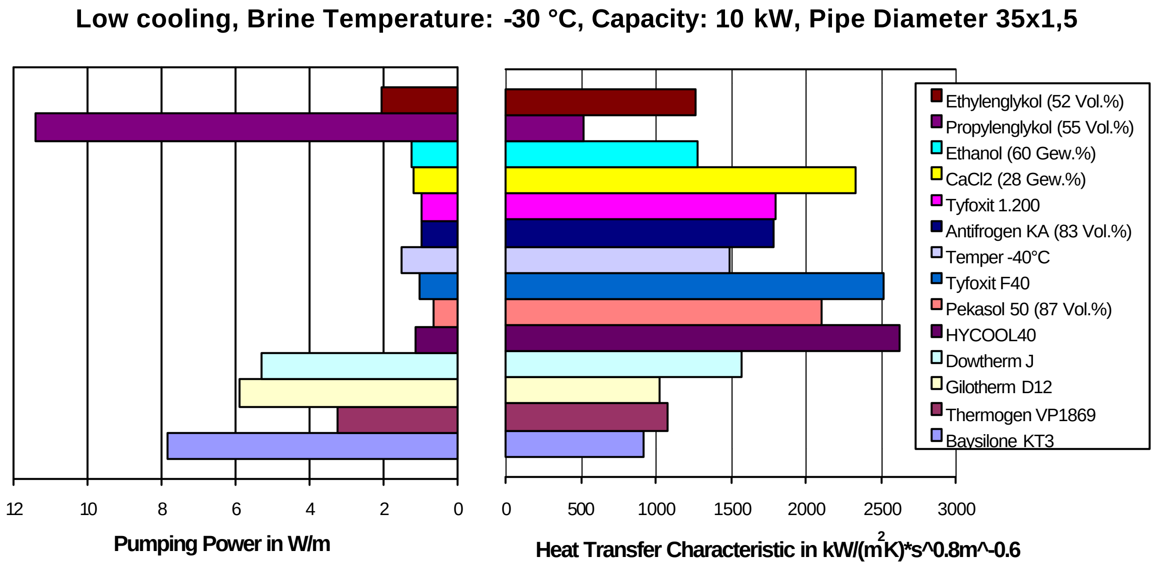


Figure 8. Pumping Power and Heat Transfer Characteristic of Secondary Refrigerants (Low Temperature Cooling)

It can be concluded that among the available secondary refrigerants on the market there is compared to the old calciumchloride brine nearly no one which shows a remarkable better behavior concerning pumping power requirements and heat transfer in order to diminish considerably the energetic disadvantages of indirect cooling systems. The reason is that with these kind of fluids, the sensible heat is used for heat transport meaning that large quantities of liquid have to be circulated with special problems in the low temperature range where high viscosity increases the pumping power.

Experimental Investigations

In order to evaluate the suitability of indirect cooling systems for supermarkets together with new fluids FKW within a research project from the German Federal Government as a subcontractor a the system manufacturer fulfilled the task of measuring the energy consumption and comparing it to that of conventional supermarkets with dry expansion evaporation (Haaf et al 1997). During one year each conventional supermarket has been measured first with the refrigerant R-404A and during the following year with R-407C and the results compared to each other in order to find energetic advantages of one of the two fluids. It was expected that because the second one shows a certain temperature glide this could benefit the energy consumption.

The result was however that during both years of the measurement with the two different refrigerants no remarkable advantage or disadvantage of one of the two fluids could be identified. Therefore assuming the same leakage rate only the different global warming potential of both refrigerants leads to a benefit for R-407C as also can be seen in the results of the TEWI 3-Project for low temperatures (Figure 7).

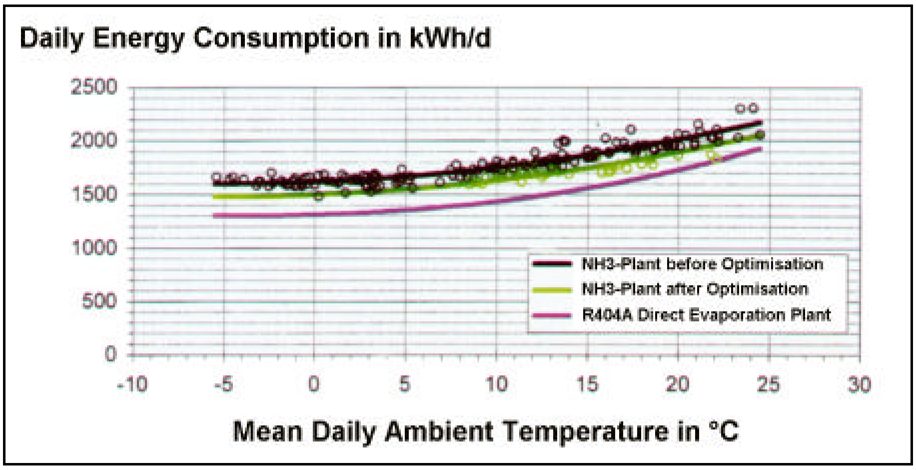


Figure 9. Energy Consumption of a NH_3- and a R-404A-Plant

In order to find the best solution for future supermarkets concerning the greenhouse effect an ammonia supermarket with indirect cooling was built by the same manufacturer and measured by FKW in another research project sponsored also by the German Government and compared as far as possible with the before mentioned conventional supermarket with R-404A and R-407C respectively for the same degree days. The comparison led to an around 15 % higher energy consumption of the indirect system which could be decreased moderately by certain optimization measures to around 10 % (Figure 9) (Haaf 1998) Another supermarket system manufacturer confirmed this statement by measurements showing higher energy consumption of 18 % for Tyfoxit and 11 % for HYCOOL (Arneg 1999).

More or less the same results can be found in the theoretical evaluation of the TEWI 3-Project leading to the conclusion that although a much lower TEWI can be achieved by indirect cooling systems as compared to conventional systems with hydrofluorocarbons the energy consumption is always higher for indirect systems although using the energetic best refrigerant ammonia. This is mainly due to the high viscosity of the secondary coolant in the low temperature part of the supermarket systems where the pumping power depends strongly on the very high viscosity at low cooling temperatures. Therefore the conclusion is that especially for the low temperature cabinets in supermarket systems another secondary coolant has to be found in order to decrease the energy consumption due to the extremely high pumping power there.

2.2.3 Improvement of Indirect Systems

In order to improve such kind of a system the other possibility would be to employ a secondary coolant with phase change material which shows lower pumping power requirements and better heat transfer characteristics especially at low temperatures. This can be achieved by applying fluids which change phase from the liquid to the gaseous state in this region of temperatures which means that refrigerants themselves could be applied as secondary low temperature coolant too. Since the synthetic refrigerants show the well known environmental draw backs concerning Global Warming Potential and the natural refrigerants with exception of carbondioxide show the local safety problems only carbondioxide itself is a possible candidate for such a secondary coolant. It has a much better heat transfer behavior than the conventional single phase fluids and also a very low viscosity which is one to two magnitudes of order lower than that of the already shown conventional coolants available on the market.

The high heat of vaporization leads to a very low mass flow of such a carbondioxide coolant resulting in very small pipe diameters as compared to either direct evaporation systems with R-22 or single phase fluids like Tyfoxit which is one of the bests fluids of that family.

A disadvantage is the relative high pressure of carbondioxide within the secondary circuit, but causing no principal problems.

The application of carbondioxide as secondary refrigerant was first mentioned 1992 by Forbes Pearson and a pilot supermarket system was built in Scotland to investigate this system in reality (Pearson 1993). Also already in 1992 at FKW such a system was built in the laboratory and investigated together with solvents for CO_2 in order to decrease the pressure inside the system. Regrettably all suitable solvents for carbondioxide are flammable fluids and therefore the high pressures have to be accepted (Enkemann et al 1994). In 1996 a Swedish company first built an ammonia system with indirect CO_2 cooling in the low temperature circuit in a commercial supermarket in Lund, Sweden (Rolfsmann 1996).

Furthermore, also HFC-404A- supermarkets with such an indirect cooling system using carbondioxide in the low temperature circuit were built mainly in Scandinavian countries. Nowadays, this seems to be a well proven technology contributing to solve the problem of the high pumping power and low heat transfer of the single phase fluids. Since, the cooling capacity of supermarket systems generally is divided between the normal and low cooling circuits in so far as the latter requires a much lower capacity, i.e. in the ammonia supermarket measured by FKW with around one third of the total capacity, therefore, also at normal temperatures the application of a favorable phase changing fluid as a coolant like for example CO_2 or ice slurry would be desirable.

Direct Cooling

A question is if a carbon dioxide system with direct evaporation in the low temperature region, as shown in Figure 10 is a more competitive system like the before mentioned secondary coolant application of carbon dioxide, as far as costs and energy are concerned.

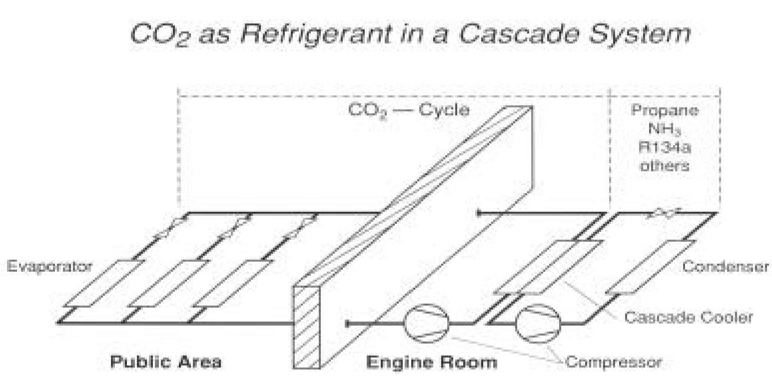


Figure 10. CO_2 as Refrigerant in a Cascade System

Theoretical investigation of both systems were done for comparison (Ferreira et al 1996) coming to the conclusion that both systems are nearly equivalent comparing energy consumption and yearly costs (Figure 11).

One important disadvantage for the cascade system is that the cascade cooler has to have a rather low temperature level in order not to exceed the pressures for conventional compressors in the low temperature cycle. This means that the evaporation temperature for the high temperature cycle must be much too low for the cooling purposes leading to higher heat transfer losses in the high temperature cycle.

Compressors with head pressures of 40 bar as designed i.e. for HFC-410A could presumably be advantageous for this application. Cascade supermarkets with direct expansion carbondioxide (R-744) systems in the low temperature circuits recently have been installed in Denmark (Kauffeld 2001). Another possible solution could be to use especially designed CO_2-compressors for higher pressures in which case it could be possible also to make use of the transcritical carbondioxide cycle as well in the low as also in the high temperature application in supermarkets. Special CO_2 compressors have already been designed by an Italian compressor manufacturer for this purpose (Dorin 1998). Such a system has been theoretically and experimentally investigated and evaluated for distributed systems by Neksa (Neksa et al 1998).

The results show that though such systems with conventional direct expansion of carbon dioxide in the cabinets show no direct emission problems concerning global warming they have much higher energy consumption of the transcritical cycle being one third higher than for a R-22 system.

Certain improvements can be expected if using a two stage CO_2-system, subcritical in the lower and transcritical in the higher temperature stage. Such a system has been evaluated and designed for an Italian supermarket manufacturer (Verondini et al 2001) by FKW.

Figure 12 shows the layout of the CO_2-System with intermediate pressure vessel for providing liquid CO_2 for the high temperature pumping system and recondensation of the low temperature direct injection system. A two stage high temperature compressor with intercooler is needed to reach energy consumption of a comparable R-22-system.

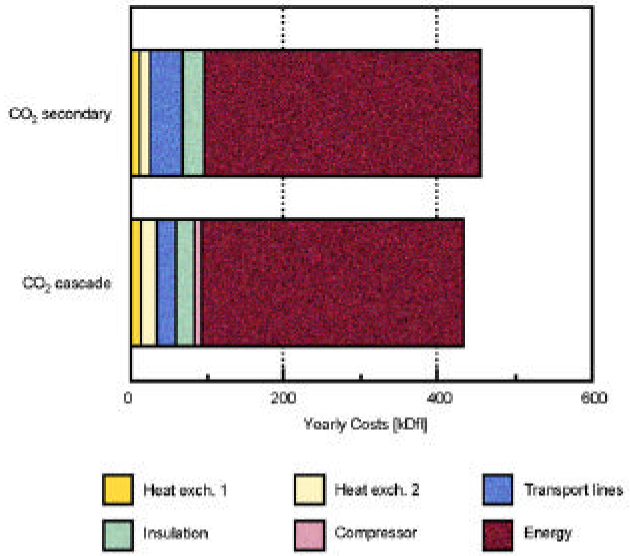


Figure 11. Annual Costs of CO_2-Systems

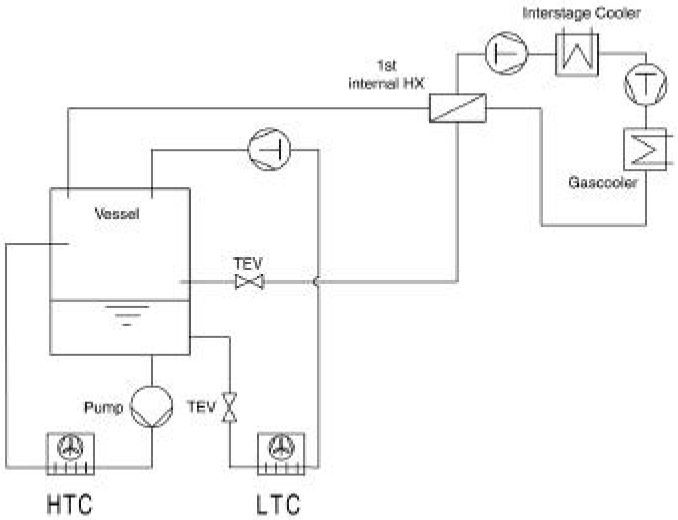


Figure 12. Two temperature level CO_2 supermarket system with 2-stage compressor in the high temperature circuit.

2.3 Mobile Air-Conditioning and Refrigeration

The mobile air conditioning sector (Figure 2 and Figure 3) showed the largest sector of refrigerant consumption and of the emission per system. Therefore, first attention to the use of CO_2 as a refrigerant was drawn to this application necessarily employing the transcritical cycle because of higher outside air and heat rejection temperatures when running mobile air-conditioning systems (Lorentzen 1990).

According to Figure 13 CO_2 (R-744) could lead to systems with smaller TEWI numbers as compared to R-134a direct and hydrocarbon indirect systems as well in Japan, Europe and North America. The reason is the still considerable leakage of mobile R-134a direct systems and the higher energy consumption of hydrocarbon indirect systems.

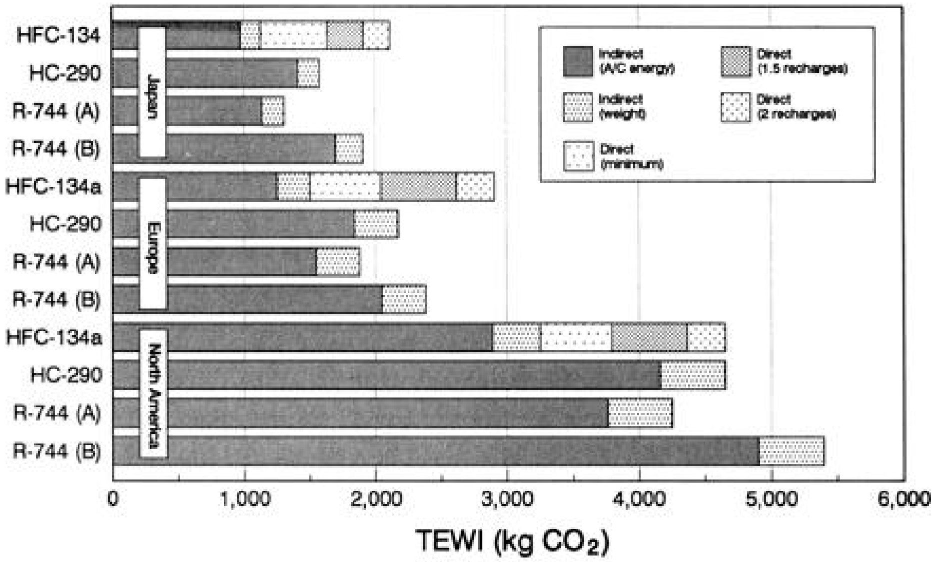


Figure 13. TEWI for R-134a, R-600a and CO_2 air conditioners in selected design climates

Leakage problems are also valid for all vehicle air conditioning and refrigeration systems such as those used in trains., trams and trucks and containers. Because of engine and vehicle vibrations and unforeseen accidents, the risk of refrigerant emissions in these applications is much higher than in standard factory manufactured stationary air-conditioning and refrigerating systems. Thus, the main application for CO_2 as a refrigerant can be the entire transport sector.

HFC-134a replaced CFC-12 in virtually all vehicle air conditioners produced in the developed countries after 1994. It is predicted that in the period 2000-2010, 70-80% of all new vehicles produced globally will have HFC-134a air conditioners. It is technically and economically feasible to significantly reduce emissions of HFC-134a refrigerants. This includes recovery and recycling, the use of high quality components with low leakage rates, and by minimizing the refrigerant charge, as well as the use of different systems using alternative refrigerants. Efficiency improvements and smaller AC units can further reduce the energy related carbon dioxide emissions.

Manufacturers are working to increase the energy efficiency, and reduce the emissions of HFC-134a systems. Refrigerant charges are also being reduced to levels below 0.8 kg per vehicle, with lowest charges currently applied in Japan and in Europe. This will show demonstrable progress during the period 2001-2003, which implies that improved HFC-134a systems can be introduced faster and at lower incremental cost than alternative systems. This is also related to the introduction of hybrid vehicles where an electricity driven hermetic refrigeration cycle using HFC-134a would virtually phase out emissions during the useful life of the equipment.

However, substantial activities are underway to develop alternatives to HFC-134a air conditioning systems for vehicles, in particular the trans-critical carbon dioxide cycle and the hydrocarbon (secondary loop) system. The carbon dioxide systems are assumed to have energy efficiencies comparable to HFC-134a. However, their high operating pressures require substantial new engineering, component reliability testing, technician training etc. It is estimated that the first CO_2 systems could be commercialized in 4-7 years.

Where it concerns secondary loop systems with a flammable refrigerant in the refrigeration cycle, it is also estimated that the energy efficiency of these systems can be brought to a level comparable of that of HFC-134a. Where it concerns new engineering, reliability testing etc., the introduction of secondary loop systems would require fewer technical innovation than would be the case for carbon dioxide systems, which implies that these systems could be implemented in 3-5 years.

Where it concerns road transport and its continuing use of HCFC-22, the market share of systems with R-404A is growing. Moreover, for a smaller amount of refrigerated road transport systems R-410A has been selected. Where it concerns the air-conditioning systems in railcars, R-134a is the leading candidate for future systems; however, the application of air-cycle based systems in high speed trains has been started in the year 2000, particularly in Germany.

Here, the air-cycle with less energy efficiency but zero direct TEWI effect (GWP=0) seems to be an attractive solution for the future, however for other reasons like simplicity and maintenance than for substituting ozone depleting fluids. In all transport sub-sectors flammable hydrocarbons have not gained any importance as refrigerants and this applies to railway road and sea transport.

2.4 Unitary Air-Conditioning and Heat Pump Systems

The unitary air conditioning and heat pump systems show the second largest sector in refrigerant consumption and the third one in emissions per system. Although not contributing per unit as much to refrigerant emissions as the commercial sector, the number of units worldwide places them second in terms of total refrigerant consumption. The TEWI 3 study for unitary air conditioning and heat pump systems shows for European conditions no large differences in TEWI (Figure 14) mainly influenced by their respective energy consumption in these systems.

Here, the question is whether energetic drawbacks of the high-pressure transcritical CO_2 cycle can be environmentally compensated for by the rather small savings in greenhouse gas refrigerant emissions if using CO_2 in hermetic systems, or if the HFC system can be further improved in terms of energy saving and leakage in order to meet the standards of other stationary systems such as domestic refrigerators or water chillers (Figure 3).

In the unitary air conditioning sector there has been continued progress in the development of alternative technologies needed to replace ODSs in unitary air conditioning and heat pump systems. During the past three years, manufacturers in the developed countries have continued to commercialize non-ODP technologies, mostly HFC based products. The dominant HFC refrigerants being employed are the blends R-407C and R-410A.

Japan has made the greatest progress in the conversion with significant portions of its residential and commercial markets being converted to R-410A and R-407C. In Europe, R-407C has been the dominant replacement for HCFC-22. Europe has also seen some penetration of R-290 and R-410A into commercialized products.

In the US a number of manufacturers have introduced non-ODP unitary systems however, currently less than 5% of the unitary market is using non-ODP refrigerants. It is anticipated that there will be a significant shift toward the non-ODP technologies in 2006, which will coincide with the implementation of new minimum efficiency standards.

The search for other alternative refrigerants has continued to move forward. The development of non-HFC technologies also received a boost from the Coca-Cola and McDonald's companies when they announced plans to phase out the use of HCFCs and HFCs by 2004.

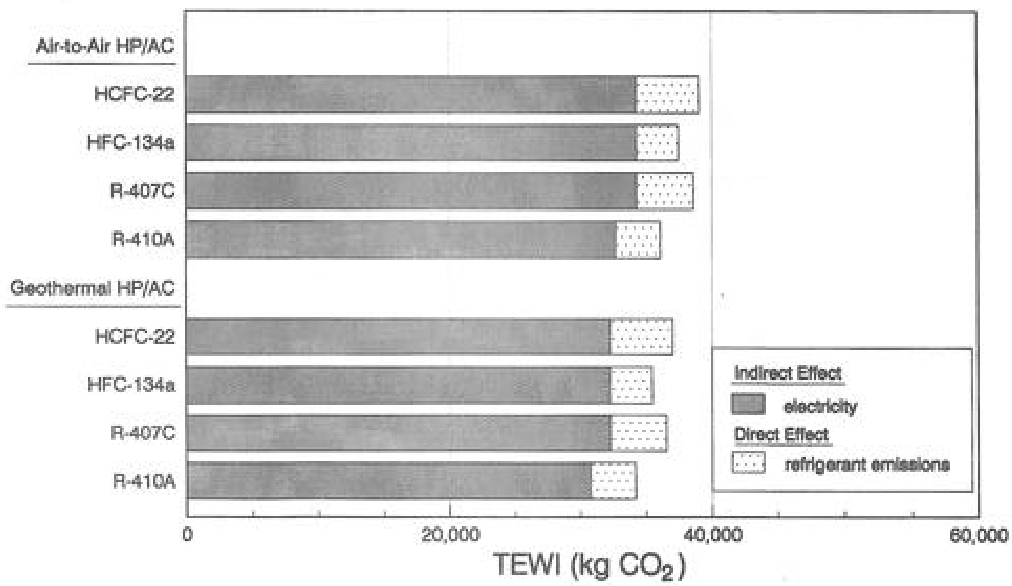


Figure 14. TEWI for residential cooling-only options in southern Europe (Greece)

2.5 Chillers

According to Fig. 2 and 3 1991 the refrigerant consumption for chillers was only 3 % caused obviously by the tight systems manufactured at the producers site, leading also to the low direct TEWI effect of 1 %.

HCFC-22 still remains the most commonly used refrigerant in positive displacement chillers (7.0 kW up to over 700 kW) employing reciprocating, screw, or scroll compressors. This refrigerant is scheduled for phase-out in new products by 2010 in most countries. A number of national regulations, particularly in the European Union member states, mandate the phase-out of HCFC-22 in new systems even earlier. In response, manufacturers have introduced new equipment employing HFCs including R-134a (particularly for water-cooled chillers), R-407C (particularly for air-cooled chillers), and less commonly R-404A, R-717 (ammonia), and hydrocarbons.

Following the CFC phase-out, the principal refrigerants used in centrifugal water chillers have been HCFC-123 and HFC-134a. New products continue to be offered for both of these refrigerants. There is a clear trend to improve the energy efficiency of these chillers, particularly in the United States where new ASHRAE energy efficiency standards have been introduced. Although subject to phase-out after 2020 under the Montreal Protocol, HCFC-123 remains the most efficient refrigerant for water chillers.

New chillers of both types – positive displacement and centrifugal – are being designed to have low refrigerant emissions during their operating life. Studies have shown that refrigerant releases throughout the life-cycle of chillers can be held to less than 0.5 % per year. Regulations are being put in place around the world to require service personnel to minimize refrigerant emissions during their activities and to require refrigerant to be reclaimed or destroyed when units are taken out of service. These measures, together with increased energy efficiency, are substantially reducing the environmental impact of new water chilling equipment.

Another option for the future is that chillers which are already serving indirect systems could also run with ammonia or hydrocarbons if they are installed for safety reason in special machinery rooms or outside of buildings.

3. Conclusion

Summarizing the considerations regarding future refrigerant and air conditioning systems under the impact of environmental regulations it can be concluded that a phase out of ozone depleting substances is well under way and earlier than the Montreal Protocol prescribes as well in industrialized as also developing countries, mainly by applying hydrofluorocarbons as substitutes for the CFCs and HCFCs .

The considered emission controls of the HFCs according to the Kyoto Protocol caused additional concern regarding the application of these fluids. Necessary measures have to be pursued against leakages of these greenhouse gases into the atmosphere. This can be done by changing direct expansion systems with high leakage rates to indirect systems but then the consumption increases leading to more carbon dioxide emissions during the power generation for driving the refrigeration air conditioning systems.

Indirect refrigeration systems would allow to reduce leakages remarkably and are necessary for locally dangerous refrigerants like hydrocarbons or ammonia as for instance in supermarkets but show a higher energy consumption which can be reduced when applying carbondioxide as secondary coolant.

Since liquid chillers for indirect systems can be made at the manufacturers site with great accuracy and tightness control the choice of the refrigerant between the environmentally or locally dangerous types of the synthetical or natural fluids respectively is then not important from the aspects of the environmental problems of refrigeration systems.

This choice then can be made between synthetic or natural fluids from the point of costs which in the one case are more the costs of the fluids in the other the costs of the protection equipment against local danger. Therefore it might be advantageous to apply a local safe natural fluid like carbondioxide in direct expansion systems but the supercritical heat rejection for CO_2 leads to lower energetic cycle performance which partly may be compensated by better component efficiencies, and cycle configurations, but presumably at higher systems costs.

4. References

Ahnefeld, G., Wobst, E., Performance comparison of a zeotropic refrigerant blend with HCFC-22; Ki Luft- und Kältetechnik, Vol. 31, No. 3, 1995, P. 134-138.

Arneg SPA, 1999, Future Supermarket Systems under the Aspect of Environmental Issues, Marsango, Italy.

Arnemann, M., Gebhardt, D., Kruse, H.: Experimentelle Bewertung neuer Kältemittelgemische als Ersatz für R-22 und R502; DIE KÄLTE und Klimatechnik 2/1995, p. 66.

DOE/AFEAS, December 1991, GW Project: Energy and Global Warming Impacts of CFC Alternative Technologies, Executive Summary, Oak Ridge National Laboratory.

DOE/AFEAS, March 1997, TEWI-III Report: Energy and Global Warming Impacts of HFC Refrigerants and Emerging Technologies: TEWI Phase 3, Oak Ridge National Laboratory.

Dorin, October 1998, 19. International Trade Fair for Refrigeration and Air Conditioning, IKK, Nürnberg, Germany.

Enkemann, T. and Arnemann, M., Hannover, 1994, Investigation of CO2 as a secondary refrigerant. Proc. Int. Conf. IIR, New applications of natural working fluids in refrigeration and air conditioning, p. 721.

Enkemann, T., December 1996, Kälteträger für Anlagen mit indirekter Kühlung , Übersicht, FKW-Seminar XVIII: Aktuelle Entwicklungen in der Kältetechnik in Supermärkten/Gewerbekälte, Hannover,

Ferreira Infatec, Danmark, 1996, C.A. and Boukens, R.A., Carbon dioxide – secondary coolant or refrigerant for cascade systems?, Proc. Int. Conf. IIR, Applications for natural refrigerants, Aarhus, p. 185.

FKW, 1996, Natural Fluid Based Refrigeration, Trends in natural fluid based Technologies for Domestic and other Refrigeration Applications

Haaf, S., Enkemann, T., Ammoniak-Kälteanlagen für Supermärkte – ein Beitrag zur TEWI-Reduzierung, DKV-Tagungsbericht Hamburg, 1997, (24), No. II.1, p. 48, Deutscher Kälte- und Klimatechnischer Verein, DKV e.V.

Haaf, S., Ammoniak-Kälteanlagen für Supermärkte, KK-Die Kälte & Klimatechnik 8/1998 (51), p. 520

Kauffeld, M., 2001, FKW Seminar XXI, Unter-und Überkritische CO2 Anwendungen in Dänemark.

Kruse, H., 1995, CIAR , Sao Paulo, European developments in substituting CFC and HCFC refrigerants.

Kruse, H., Enkemann, T., Ki Luft- und Kältetechnik 10/1997, p. 472, Minderung des Energieverbrauchs indirekter Systeme durch Auswahl von Kälteträgern.

Kruse, H., König, H., October, 1998, System Comparison of R-22 Replacement Refrigerants: R407C and R410A, Conference Proceedings: The Earth Forum, Washington, DC.

Lorentzen, G., 1990, Trans-critical vapor compression cycle device. Patent WO 90/07683.

Montreal Protocol, December, 1991, Assessment, Report of the Refrigeration, Air conditioning and Heat Pumps Technical Options Committee, UNEP.

Neksa, P., Girotto, S. and Schiefloe, Oslo, Norway, 1998, P.A., Commercial refrigeration using CO2 as refrigerant – system design and experimental results. Proc. Int. Conf. IIR, Natural working fluids 1998, p. 227.

Pearson, F., London, 1993, Development of Improved Secondary Refrigerants, Paper presented at The Institute of Refrigeration,

Rolfsman, L., Aarhus, Danmark, 1996, CO2 and NH3 in the supermarket ICA-focus. Proc. Int. Conf. IIR, Applications for natural refrigerants, p. 219.

SOME SELECTED PROBLEMS IN SUPERSONIC COMBUSTION

Vladimir A. Sabel'nikov
Office National d'Études et de Recherches Aérospatiales,
Département Énergétique Fondamentale et Appliquée,
BP 72 – 29, avenue de la Division Leclerc,
92322 CHATILLON Cedex, France
e-mail: vladimir.sabelnikov@onera.fr

Abstract. *Three problems of turbulent combustion in supersonic flow are addressed. The first one is the flamelet modeling of nonpremixed combustion in supersonic turbulent flow. The second problem involves the combustion stabilization of the liquid kerosene in supersonic flow using a free recirculation bubble which is obtained by the interaction of a concentrated vortex with a normal shock wave. The third problem is the potential resources of the supersonic combustion enhancement using aerated (by hydrogen or air) liquid kerosene jets (effervescent sprays) injected through elliptic nozzles from tube-injectors and fin-injectors.*

Keywords*. supersonic flow, combustion, flamelet, stabilization, jet*

1. Flamelet model for nonpemixed combustion in supersonic turbulent flow

This section presents briefly the revisited supersonic flamelet modeling (Sabel'nikov et al., 1997). The results of the detailed chemistry flamelet calculation and comparison with the experimental results of Evans et al. (1978) may be found elsewhere (Sabel'nikov et al., 1997; Morgenthaler et al., 1997; Morgenthaler et al., 2001).

1.1. Introduction

Combustion of nonpremixed gases in supersonic turbulent flows has a large interest both from practical and fundamental points of view (see, e. g. Bray et al., 1994; Curran et al., 1996). Turbulent reactive flow is a complex phenomena and the prediction of supersonic turbulent combustion requires modeling of the various physical and chemical processes : thermal expansion and high speed compressibility, shock waves, mixing and finite rate chemistry. Much attention has been paid to turbulent flame structure and combustion mechanisms, and in particular to understanding the interaction between turbulence and combustion for low speed flows (see, e.g. Bilger, 1976, 1989; Peters, 1984, 1986; Bray and Peters, 1994; Kuznetsov and Sabel'nikov, 1990; Chen and Kollman, 1994). Different regimes in nonpremixed turbulent combustion were identified : 1) distributed reaction zones, where chemical times are long in comparison with the largest time scales of the turbulent flow; 2) flamelet regime, where the longest chemical time scale τ_c is shorter than the smallest time scale of the turbulence (Kolmogorov time scale τ_η) and rms mixture fraction Z' fluctuation is larger than the flame thickness in mixture fraction space $(\Delta Z)_F$; 3) connected reaction zones, where $\tau_\eta >> \tau_c$ and $Z' < (\Delta Z)_F$ (see, e.g. Bray and Peters, 1994).

In the last three decades an enormous effort has been devoted to the development of mathematical models for the prediction of turbulent diffusion flames (Pope, 1990). These models consist usually of two parts : turbulence and combustion models. The turbulence models are commonly based on such moment closures as k-ε models or second-order moment closures (e.g. Chen and Kollmann, 1994; Bray and Peters, 1994). The chemistry models are currently in use are fast or equilibrium chemistry, stretched laminar flamelet models, conditional moment closure (Klimenko, 1990; Bilger, 1993) and eddy break-up models. Thermochemical statistics is obtained from the probability density function (pdf) of mixture fraction, Z, for the case of fast or equilibrium chemistry and from the joint pdf of Z and its scalar dissipation, $N=D(\nabla Z)^2$ for the flamelet approach (see, e.g. Borghi, 1988; Kuznetsov and Sabel'nikov, 1990; Bray and Peters, 1994; Lentini, 1994). The shape of pdf of Z is usually assumed, so that the pdf is uniquely determined by the mean and the variance of the Z. The joint pdf of Z and N is factorized using the assumption about statistical independence of Z and N and finally, for the pdf of N log-normal distribution is used (e.g., Kuznetsov and Sabel'nikov, 1990; Lentini, 1994; Buriko et al., 1994). The assumed pdf method, though successful for nonpremixed flames with simple chemistry, cannot be readily extended to more general case. The composition joint pdf transport equation provide a more accurate closure. The large dimensionality of the joint pdf is the key obstacle to solve directly the pdf transport equation. This difficulty can be partly overcomed through the use of Monte Carlo methods (Pope, 1985, 1990; Kollmann, 1990).

We now turn to supersonic combustion. Raman measurements of mixing and finite - rate chemistry in supersonic hydrogen-air nonpremixed flame done by Cheng et al. (1994) show that this type of combustion is quite different from combustion in low speed flows. Indeed, they concluded that: 1) finite chemical reaction rate effects become particularly important in supersonic flows; 2) higher fluctuations of temperature and species concentrations are observed in this supersonic flame compared to a subsonic flame; this effect is due to the interaction of velocity and temperature in supersonic compressible flow. Balakrishnan and Williams (1993) estimated the possible turbulent combustion regimes in hydrogen-air systems for applications in supersonic combustor of hypersonic aircraft such as the National Aerospace

Plane (NASP) over the range of flight Mach numbers up to 25 and at altitudes from 30 to 70 km. Their results indicate that laminar flamelets are likely to exist in supersonic combustor under above conditions.

Recent review of the fundamentals of supersonic combustion and extensive bibliography has been given by Bray et al. (1994). It is useful, similar to the case of low speed turbulent combustion, to try do separate the problems of developing models of turbulent supersonic combustion into those linked with gasdynamics and chemical reactions (e.g. Baev et al., 1979; Zimont et al., 1983; Meshcheryakov and Sabel'nikov, 1988; Farshchi, 1990; Villasenor et al., 1992; Zheng, 1993; Zheng and Bray, 1994, 1997; Bray et al., 1994; Jouve, 1996). The starting point of such an analysis is the treatment of the influence of compressibility effects on turbulence without chemical reactions, a subject which has been studied for many years. The current state of our knowledge of compressible turbulence is reviewed by Zeman (1990), Speziale and Sarkar (1991), Lele (1994) and Smits and Dussauge (1996). An important parameter which characterizes the effect of compressibility on turbulence is a turbulence Mach number $M_t = (2k/a^2)^{1/2}$, where k and a are characteristic values of the turbulent kinetic energy and speed of sound, respectively. Compressibility begins to have a noticeable effect on turbulence if $M_t > 0.3$ (e.g. Zeman, 1990). The dilatational dissipation and the pressure-dilatation covariance (which are clearly zero in constant density turbulence) led to a reduction in turbulent kinetic energy and hence in turbulent mixing at $M_t > 0.3$. The observed reduction in the spreading rate of a supersonic shear layer is explained by these effects.

Different combustion models for supersonic flows have been reviewed by Bray et al. (1994). Among these, Eifler and Kollmann (1993) derived and modeled the transport equation for the joint pdf of velocity, density, internal energy, mixture fraction and velocity dilatation. This set of variables provides a complete descriptions of the local mechanical and thermodynamic state of the gas for the equilibrium chemistry. Eifler and Kollmann (1993) solved their pdf equation by Monte Carlo method (Pope, 1985) and compared their results with experimental data of Evans et al. (1978). Details of the closure model, the numerical procedure and the results of calculation can be found in Kollmann (1990) and Eifler and Kollmann (1993). Although joint pdf approach of Eifler and Kollmann (1993) is most comprehensive treatment of supersonic combustion, it can hardly to be used at the present time for the application aims.

Zheng and Bray (1992, 1994, 1997) extended the flamelet model to supersonic combustion. The most restrictive assumption of the extension is the neglect the unsteady pressure term in the equation for the stagnation enthalpy. This basic assumption has an uncertain validity in high speed flows. The main difficulties that arise when the supersonic flamelet modeling is developed are due to the role played by dynamic compressibility, i.e. the interaction of velocity and the temperature in flamelet. The kinetic to thermal energy conversion can become comparable to the heat released by combustion. A particular attention is paid to recently revisited flamelet modeling by Sabel'nikov et al. (1997).

1.2. Subsonic flamelet. Basic equations

It is convenient for future discussion of high-speed flamelet model to recall the basic assumptions and governing equations of the classical subsonic flamelet model. Flamelet models for combustion in turbulent flows consider the instantaneous turbulent flame as ensemble of one-dimensional (1-D) laminar stretched flames (see, for instance Bray and Peters, 1994; Kuznetsov and Sabel'nikov, 1990). In such a modeling the Kolmogorov micro-scale of turbulence η is assumed to be larger that the instantaneous flame thickness. In order to write the simplest equations for the elementary flamelet for subsonic flows, the structure of the 1-D flamelet is often supposed to satisfy the classical diffusion flame theory assumptions, namely that: 1) the unsteady pressure term can be neglected compared to the chemical heat release in the energy equation (even though the pressure gradient is kept in the instantaneous momentum equations); 2) the coefficients of heat and mass diffusion are identical for all chemical species; 3) heat losses due to radiation are neglected; 4) the unsteadiness of the flow field inside of laminar flamelet is not taken into consideration, i.e. a quasi-steady simplification is accepted. Under above assumptions, the equations governing the structure of the flamelet in a reference system which is attached to the stoichiometric iso-scalar surface Z_{st} can be written as (Bray and Peters, 1994; Kuznetsov and Sabel'nikov, 1990; Buriko et al., 1994)

$$N \frac{d^2 Y_k}{dZ^2} = \dot{\omega}_k W_k, \qquad k = 1,..., K \,, \tag{1.1}$$

$$h = \sum_k h_k(T) Y_k = A_1 Z + B_1 \,, \tag{1.2}$$

$$p = \rho R T \sum_k Y_k / W_k \,. \tag{1.3}$$

In these equations, $h_k(T) = h_k^0 + \int_{T_0}^{T} c_{pk} dT$ is the static enthalpy per unit mass of species k, and N is the scalar dissipation rate, $N = D(\nabla Z)^2$; Z is the mixture fraction; R is the universal gas constant. For each of the K species, Y_k is the mass fraction, W_k is the molecular weight and $\dot{\omega}_k$ is the molar production rate. The pressure, temperature and density are noted by p, T and ρ, respectively. The coefficients A_1 and B_1 are given by $A_1 = h_j - h_e$ and $B_1 = h_e$, where the subscripts j and e represent the conditions prevailing at either side of the flamelet. In the above equations only diffusion normal to the iso-scalar surfaces and the chemical source term are conserved. Non-steady, convective and tangential diffusion terms have been neglected. The transient behavior of strained laminar nonpremixed flames in low speed flows has been the subject of intensive research in the last few years (see, e.g. Haworth et al., 1988; Darabiha, 1992; Mell et al., 1994; Thévenin and Candel, 1994; Im et al., 1995). It is recognized that transient effects in some cases may play the considerable role in the internal structure of flamelets. The degree of differential diffusion in (non-reacting) turbulent jet flows was experimentally examined by Smith et al., 1995. One should notice that Eq. (1.2) is valid as long as the equation for the mixture fraction Z and the static enthalpy h are identical. Moreover, the assumptions relative to the role of the unsteady pressure term and to the equality of all molecular transport coefficients must also be applicable to the flow external to the flamelet.

Equations (1.1), (1.2) have to be solved with boundary conditions which, in general case, are dependent on the interaction of the flamelet with the turbulent flow. This problem is not solved at present time (cf. Bish and Dahm, 1995; Everest et. al., 1995). Boundary conditions are usually chosen simply assuming that within the outer flow of oxidizer no combustible remains, and vice-versa, and that Eq. (1.1) applies in the whole range (0,1) of Z, i.e.

$$Z = 0, \text{ pure oxidizer,} \tag{1.4}$$
$$Z = 1, \text{ pure fuel.} \tag{1.5}$$

Using these boundary conditions, Eqs. (1.1) – (1.3) can be solved for each given value of N. Thus, the instantaneous values of the temperature and the mass fractions are expressed as

$$T = T(Z, N), \tag{1.6}$$
$$Y_k = Y_k(Z, N). \tag{1.7}$$

Equations (1.6), (1.7) constitute the flamelet library from which the expectation of the arbitrary function of the temperature and composition $F(T, Y_k)$ can be recovered if the joint pdf of Z and N is known

$$\langle F \rangle = \iint F(T(Z, N), Y_k(Z, N))P(Z, N)dZdN. \tag{1.8}$$

The joint pdf $P(Z, N)$ can be factorized using the assumption of statistical independence of Z and N (e.g. Kuznetsov and Sabel'nikov, 1990), i.e.

$$P(Z, N) = P(Z)P(N). \tag{1.9}$$

The individual pdf's $P(Z)$ and $P(N)$ are approximated using methods described, e.g. by Kuznetsov and Sabel'nikov (1990), Bray and Peters (1994), Lentini (1994), Buriko et al. (1994).

1.3. Supersonic flamelet

The extension of such a formulation to the case of supersonic flows present three major difficulties. First, the above formulation does not hold when shock waves are present (Eifler and Kollmann, 1993). This problem will not be addressed here since only the case of adapted supersonic fuel jets (i.e. jets which expand ideally into co-flow oxidizer stream) is considered. The second is related to the decrease of the Kolmogorov micro-scale η in supersonic flows. Balakrishnan and Williams (1993) have shown that the flamelet theory can still be applied in a range of Reynolds and Damköler numbers representative of supersonic combustion applications. The third is associated to the increase of kinetic energy in supersonic flows, which becomes of the same order of the chemical heat release. Taking into account the influence of the kinetic energy requires the equation for static enthalpy to be written as an equation for total enthalpy H

$$H = \sum_k h_k(T)Y_k + \frac{u^2}{2} = A_2 Z + B_2, \qquad A_2 = H_j - H_e,\ B_2 = H_e. \tag{1.10}$$

Recall that Eq. (1.10) is valid only in the case when the unsteady pressure term is neglected in the equation for the stagnation enthalpy. It must be noticed that, even though the form of the equation for H remains analogous to Eq. (1.2), a new parameter is introduced to the flamelet library u, which is the modulus of the flow velocity. As a consequence, the functional dependence of temperature and composition now includes a third random parameter u

$$T = T(Z, N, u)\ , \tag{1.11}$$
$$Y_k = Y_k(Z, N, u)\ . \tag{1.12}$$

This supplementary functional dependence appears implicitly in Eq. (1.1) through the dependence of $\dot{\omega}_k$ on temperature, $\dot{\omega}_k = \dot{\omega}_k(Y_k, Z, u)$. It is clear that such revised flamelet model, which includes the effect of kinetic energy, needs the knowledge of the joint pdf of *Z, N* and *u*, *P*(*Z, N, u*), in order to recover the mean flow properties. Recalling that *N* is characteristic of the small scales of turbulence, while *Z* and *u* are dependent on the large scale characteristics of the turbulent flow, one can apply the classical Kolmogorov's hypothesis of statistical independence (e.g., Kuznetsov and Sabel'nikov, 1990), i.e.

$$P(Z, N, u) = P(Z, u)P(N). \tag{1.13}$$

Moreover, the joint pdf of *Z* and *u* can be rewritten using the Bayes theorem

$$P(Z, u) = P(Z)\, P(u \mid Z), \tag{1.14}$$

where $P(u \mid Z)$ is the conditional pdf of velocity at a given value of *Z*. This pdf must represent correctly the coupling that exists in supersonic flows between the flow velocity and enthalpy.

1.4. Conditional moment closure of the supersonic flamelet

In the present case of a compressible 3-parameter flamelet library described by Eqs. (1.11), (1.12), Eq. (1.8) can be rewritten using Eqs. (1.13), (1.14) as

$$\langle F \rangle = \iiint F(T(Z,N,u), Y_k(Z,N,u))P(Z)P(N)P(u \mid Z)\,dZdNdu. \tag{1.15}$$

This equation can be modified by the use of the conditional moment closure technique similar to the one developed by Bilger (1993) and using the hypothesis that the fluctuations around the conditionally averaged velocity $< u \mid Z >$ are small, i.e. the conditional pdf is approximated by Dirac δ-function

$$P(u \mid Z) = \delta(u - < u \mid Z >), \qquad \langle F \rangle = \iint F(T(Z,N,< u \mid Z >), Y_k(Z,N,< u \mid Z >))P(Z)P(N)\,dZdN. \tag{1.16}$$

This hypothesis leads to the following conservation equations in the flamelet library

$$N\frac{d^2\tilde{Y}_k}{dZ^2} = \dot{\omega}_k(Z, N, \langle u \mid Z \rangle)W_k, \qquad k = 1,\ldots,K\ , \tag{1.17}$$

$$H = \sum_k h_k(\tilde{T})\tilde{Y}_k + \frac{\langle u \mid Z \rangle^2}{2} = A_2 Z + B_2, \tag{1.18}$$

where the temperature $\tilde{T}$ and the mass fractions $\tilde{Y}_k$ are the conditionally averaged values $\tilde{T} = T(Z,N,< u \mid Z >)$ and $\tilde{Y}_k = Y_k(Z,N,< u \mid Z >)$. Since the exact functional form of $< u \mid Z >$ is presently unknown, it can be assumed as a first approximation that a linear relationship exists between $< u \mid Z >$ and *Z*, i. e.

$$< u \mid Z > = a_1 + b_1 Z. \tag{1.19}$$

The coefficients a_1 and b_1 can be considered to be constant for a given velocity difference $\Delta u = u_F - u_O$, fixed by the flow of fuel and oxidizer external to the flamelet, $a_1 = u_O$ and $b_1 = \Delta u$. This linear dependence assumption reduces the three-parameter library to a two-parameter library for the given velocity difference Δu. This dramatically simplifies the calculations and the calculation of the moments of the pdf at each point. The functional dependence of the conditionally averaged values of temperature and composition reduces to: $\tilde{T} = \tilde{T}(Z, N)$ and $\tilde{Y}_k = \tilde{Y}_k(Z, N)$. Notice that obtaining this reduction in the number of parameters for the flamelet library requires that $< u \mid Z >$ be a function of *Z* only.

1.5. Results and discussion

It is useful first to perform a simplified analysis which gives the order of magnitude of the modifications induced by taking into account the kinetic energy effect using the formulation presented above. Comparing Eqs. (1.2), (1.18), recalling Eq. (1.19) for $< u|Z >$ and supposing that the specific heat at constant pressure c_p is the same for all species, the local temperature increase within the flamelet $\Delta T|_{\Delta u}$ due to kinetic to thermal energy conversion may be expressed as

$$\Delta T\Big|_{\Delta u} = \frac{(\Delta u)^2}{2c_p} Z(1-Z). \tag{1.20}$$

Within the flamelet, as it follows from Eq. (1.20), the maximum temperature increase corresponds to Z = 0.5. Recalling that the flamelet is positioned in the vicinity of the value of Z_{st} which corresponds to stoichiometric composition and that in cases of practical interest, when undiluted fuel streams react with air, $Z_{st}(1\text{-}Z_{st}) \ll 1$, it can be readily seen that the increase in the maximum temperature will be much smaller that the one corresponding to Z = 0.5. The analysis (Sabel'nikov et al., 1997) shows that accounting for kinetic to thermal energy conversion increases also the critical value of the scalar dissipation rate N at which flamelet extinction occurs. The results of the detailed chemistry flamelet calculation and comparison with the experimental results of Evans et al. (1978) may be found elsewhere (Sabel'nikov et al., 1997; Morgenthaler et al.; 1997, Morgenthaler et al., 2001).

2. Combustion stabilization in supersonic flow using free recirculating bubble

This section presents the experimental results (Sabel'nikov et al., 1998; Sabel'nikov and Penzin, 2000) of the investigation of the self-ignition and combustion stabilization of aerated by air liquid kerosene in supersonic flow using the free recirculating bubble which was generated by the interaction of a concentrated vortex with a normal shock wave. Tests were conducted in the hypersonic facility T - 131B of the Central Aero-Hydrodynamic Institute of Moscow (TsAGI) in a nearly matched to atmospheric pressure supersonic jet at the exit of the two-dimensional channel over Mach number range M = 2.0 - 2.8 and total temperature range T_t = 1200 - 1400 K. Initial thickness of jet was equal 67 mm. Self-ignition and stabilization were obtained at T_t = 1400 K. Self-ignition was not achieved at T_t = 1200 K and 1300 K. Thickness of free recirculating bubble was about 25 - 30 mm.

2.1. Introduction

Experimental investigations (Zatoloka et al., 1975, 1978; Ivanyushkin et al., 1989; Delery et al., 1984; Mettwally et al., 1989; Glotov, 1989; Michael et al., 1995; Kalkhoran et al., 1996) revealed that the interaction of a concentrated vortex with a shock wave at excess of critical conditions results in the breakup of the vortex and generation of a free recirculating zone (bubble). Vortices in the works of Zatoloka et al. (1975, 1978) and Ivanyushkin et al. (1989) were generated using different types of generators, e.g. semispan wing having a diamond shape airfoil section which was installed at the angle of attack to the incoming flow. This kind of the vortex-generator allowed quite easily to vary the vortex intensity by the variation of the angle of attack. Shock waves in the works of Zatoloka et al. (1975, 1978) and Ivanyushkin et al. (1989) were produced by: 1) supersonic two-dimensional inlets, 2) blunt bodies, 3) wedges, 4) axisymmetric diffusers (which were similar to supersonic nozzle). In the works of Mettwally et al. (1989) and Glotov (1989) phenomenon of free recirculating bubble generation was studied with the interaction of the vortex with a central shock wave (Mach stem) in overexpanded jet. Already by Ivanyushkin et al. (1989) it was concluded that breakup of vortex and generation of free recirculating bubble is determined mainly by intensities of vortex and shock wave and does not depend on type of vortex and shock generators.

It is very attractive idea to use a free recirculating bubble for combustion stabilization in supersonic flow. High efficiency of such a kind of combustion stabilization in supersonic flow was demonstrated by Winterfeld (1968) at studying H_2 – air combustion (with a forced ignition). Free recirculating bubble by Winterfeld (1968) was generated by interaction of shock waves with base wake behind of cylindrical model of diameter 20 mm in supersonic flow with M = 2.1 and T_t = 375 K. Sabel'nikov (1997), Figueira da Silva et al. (1997) demonstrated, using the numerical simulation, the self-ignition and combustion stabilization of methane in free recirculating bubble (with lateral dimension of the latter of about 5 cm) in supersonic flow with M = 3.0 and T_t = 1400 K.

The objective of the present experimental investigation was to study the self-ignition and combustion stabilization of aerated by air liquid kerosene (sometime called in the literature as the barbotaged kerosene) in supersonic flow using a free recirculating bubble which was generated by the interaction of concentrated vortex with normal shock wave. The investigation was conducted in the hypersonic facility T-131B of TsAGI. Tests were performed in nearly matched supersonic jet at the vicinity of the exit of two-dimensional channel at M = 2.0 - 2.8 and T_t = 1200 - 1400 K.

2.2. Minimum dimension of recirculating bubble needed for self-ignition and combustion stabilization

For the evaluation of the minimal size of a free recirculating bubble which is required for the self-ignition and combustion stabilization we will use the criterion of combustion stabilization behind a bluff body in subsonic flow of homogeneous fuel mixture (see, e.g. Winterfeld, 1968; Shchetinkov, 1965; Zakkay, 1970; Baev et al., 1984)

$$\tau_{res} / \tau_{ind} \approx 1, \tag{2.1}$$

where τ_{res} and τ_{ind} are the residence and induction times, respectively. In the experiments of Winterfeld (1968) it was shown that this criterion is applicable both for a supersonic flow and a nonpremixed combustion. Residence time can be roughly estimated by the equation (Winterfeld, 1968; Shchetinkov, 1965; Zakkay, 1970; Baev et al., 1984)

$$\tau_{res} = k \cdot h / u_e , \tag{2.2}$$

where k = 30 - 50, h is maximum thickness of the recirculating bubble, u_e is velocity at the boundary of the recirculating bubble (i.e. behind the conical shock). Estimation for T_t = 1200 - 1400 K and M = 2.5 gives $u_e \approx$1000 m/s. The value k = 50 was obtained by Zakkay (1970), where residence time was measured in a base flow region of an axisymmetric body at M = 4.1. For the estimation of the induction time we assume the static temperature inside the recirculating bubble is nearly total temperature. Another assumption is that static pressure inside the recirculating bubble is equal to atmospheric pressure. For T_t = 1200 K and 1400 K and atmospheric pressure induction times for stoichiometric kerosene-air mixture are (e.g. Shchetinkov, 1965; Westbrook and Dryer, 1984) $\tau_{ind} \approx 5 \cdot 10^{-3} s$ and $3 \cdot 10^{-4} s$, respectively (for the hydrogen-air mixture $\tau_{ind} \approx 5 \cdot 10^{-5} s$ and $1 \cdot 10^{-5} s$, respectively). Hence, as one can obtain from Eqs. (2.1), (2.2), for the stabilization of combustion of kerosene-air mixture in the recirculating bubble it is necessary that the thickness of recirculating bubble were greater of some critical value h_{cr}

$$h > h_{cr} \approx u_e \cdot \tau_{ind} / k, \tag{2.3}$$

where $h_{cr} \approx 10 - 16\ cm$ at T_t = 1200 K and $h_{cr} \approx 0.6 - 1\ cm$ at T_t = 1400 K.

In our experiment the entrance diameter of the diffuser (shock-generator) was d = 25 mm. It is known that the thickness of recirculating bubble h is almost the same as d (Zatoloka et al., 1975, 1978; Ivanyushkin et al., 1989). Thus we can expect the self-ignition and stabilization of combustion in our tests will occur if $T_t \geq$ 1400 K.

2.3. Scheme of the experiment. Experimental model and tests conditions

Experiments were conducted in the hypersonic facility T - 131B of TsAGI which is used for scramjet combustor connected pipe tests. "Vitiated" air parameters in the kerosene-air pre-heater were T_t = 1200 - 1400 K, $P_t \leq$ 4.0 MPa. Scheme of the experiment is presented in Fig. (2.1).

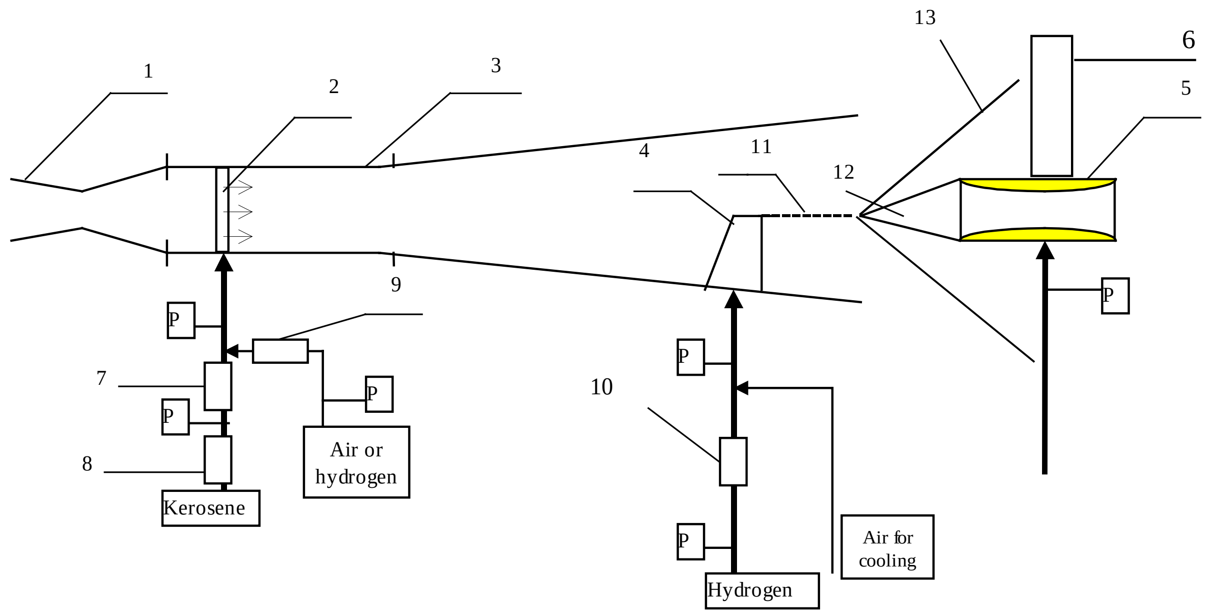


Figure 2.1. Scheme of experiment.

Supersonic nearly matched jet was exhausted into still atmosphere from the exit of a rectangular divergent channel (3) which is connected to a rectangular nozzle (1) with the exit cross section 30x100 mm^2 and designed for M = 2.5. Free recirculating bubble was generated due to the interaction between vortex (11) generated by vortex generator (4) mounted on the channel wall and bow shock wave in front of the diffuser (5), which is placed downstream the flow. Experimental model includes channel (3), injectors (2), pylon-vortex generator (4), diffuser (shock-generator) (5), transversing equipment (6), Fig. (2.2).

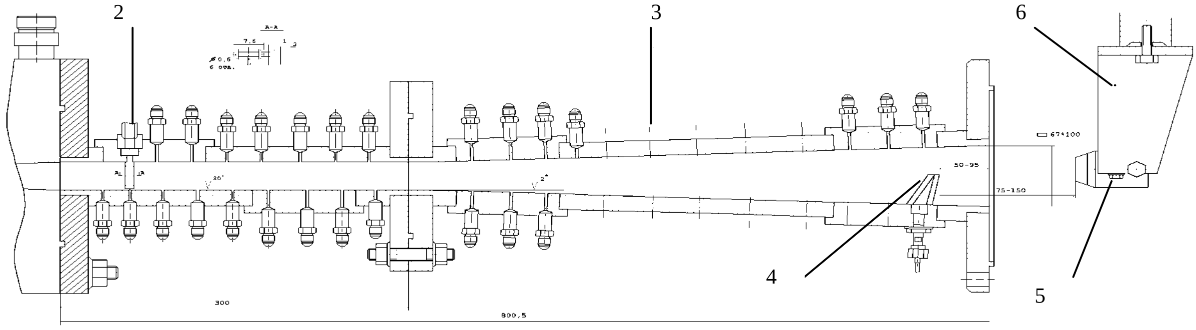
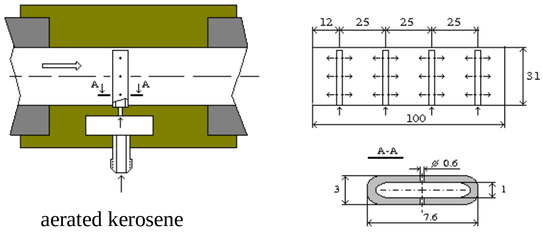


Figure 2.2. Scheme of the experimental model (all dimensions in mm).

The channel (3) consists of two sections. The front section has length of 300 mm and entrance cross-section 30x100 mm^2, it diverges along bottom wall with an angle of half degree. Second section has length of 500 mm, it diverges along bottom and top walls with an angle of 2 degrees. Exit cross-section of the channel is 67x100 mm^2. Aerated by air liquid kerosene was injected perpendicular to the main flow through the holes drilled in the walls of four tube-injectors (2), which were installed at 70 mm distance from the channel entrance. Injectors were mounted on the lid of a hatch across the channel. Each of the four injectors had 3 holes (of the diameter 0.6 mm) on each side, Fig. (2.3). The distance between injectors and channel exit was about 1 m, thus the kerosene-air mixture was nearly homogeneous at the channel exit. The aeration of the liquid kerosene was carried out with the aid of the aeration device (Avrashkov et al., 1990a, 1990b), which provides a kerosene-air mixture with mass rate ≤ 0.25 kg/s and gas mass fraction in the mixture less than 5 percents, Fig. (2.4). Aeration procedure allowed to provide better mixing of liquid kerosene jets with a supersonic air flow, as was shown by Avrashkov et al. (1990a, 1990b).



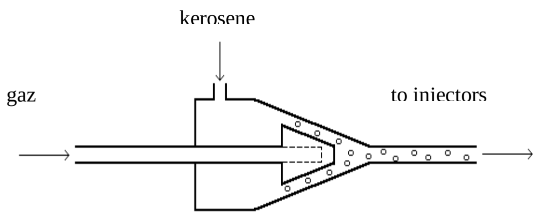
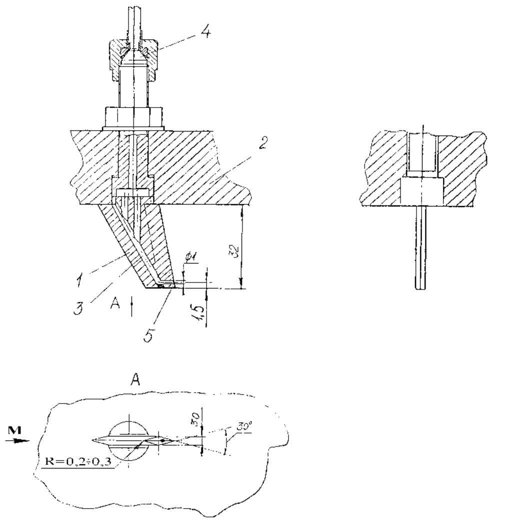


Figure 2.3. (to the left) Geometry and location of tube-injectors (all dimensions in mm).
Figure 2.4. (to the right) Scheme of the aeration device.

The vortex-generator (1) in Fig. (2.1) ((4) in Fig. 2.2) was semispan wing having a diamond shape airfoil section with a chord length of 15 mm, a span of 32 mm, a half-angle of 15 degrees, and angle of attack in the range ±180 degrees, Fig. (2.5). The vortex-generator was mounted on the wall of the channel at the distance of 50 mm upstream of the exit. The construction of vortex-generator had a cooling duct (3) to cool the generator during start-up of the test facility. This duct was also used to supply hydrogen (or another fuel) into the vortex and in this way into recirculating bubble. To provide the durability of vortex-generator under high temperatures and pressures it was manufactured from a strong alloy steel.



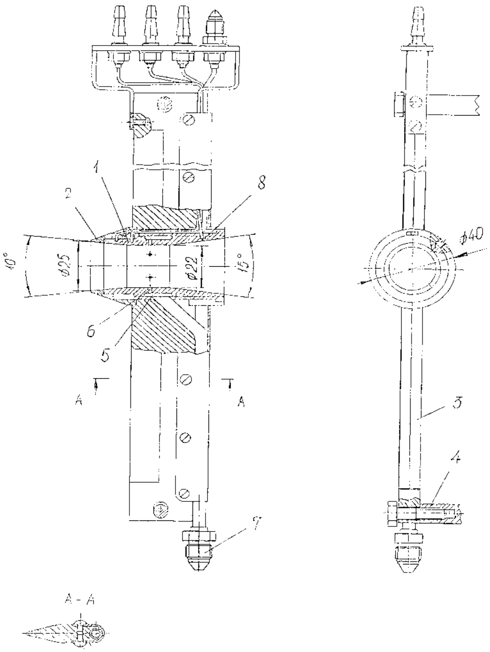


Figure 2.5. (to the left) Scheme of the vortex-generator (all dimensions in mm).
Figure 2.6. (to the right) Scheme of the shock-generator (all dimensions in mm).

Schematic plot of the diffuser (shock-generator) is given in Fig. (2.6). As mentioned above the dimensions of the diffuser were chosen to get the free recirculating bubble, emerging as the result of interaction between vortex and bow shock in front of the diffuser, sufficient to sustain self-ignition and combustion stabilization. Diffuser was fixed in a special holder (3) which is mounted on the stand of the transversing equipment (4). Transversing equipment allowed to move the holder with the diffuser in and out of the desired region of the flow during a test run. Diffuser had a sharp edge lip. Relation between throat area entrance and diffuser area was chosen according to the recommendations of the work of Ivanyushkin and Korotkov (1995). The fulfillment of the condition found in this work assured the start of the diffuser in supersonic flow with $M > 2.1$ including the cases when the concentrated vortex hits the diffuser. Throttling hollow (5) was connected to the diffuser throat by six holes (6). The holes were spread evenly across the circle of the channel's cross-section and were 1.5 mm in diameter. The throttling of the diffuser was implemented with the aid of air jets which were injected into the throttling hollow through the connecting pipe (7). This gas-dynamic throttling allowed to vary effective throat area of diffuser, and to realize in continuous way the change of working regimes of diffuser from start to unstart. Static pressure along the diffuser wall was measured in three points during the test runs. To ensure the proper operation under the high temperatures and loads the holder and constructions of diffuser were made of endurable steel, and its lips were fabricated from heat-resistant steel.

2.4. Tests methodology and measurements

At the first stage, the conditions of the generation of the free recirculating bubble were found in runs No.1-No.4. These runs were carried out for T_t in the range T_t = 1200 K - 1400 K for both with injectors (run No.1) and without injectors (run No.2-No.4). In the first case the fuel was not supplied into the injectors. The following parameters were varied during tests: the angle of attack of the vortex generator, the distance between the vortex-generator and the diffuser, the throttling intensity. Besides, the influence of air or hydrogen injection into the vortex center (throughout the vortex generator) on the generation of the free recirculating bubble was investigated. The second stage - investigation of self-ignition and combustion stabilization was conducted afterwards in runs No.5-No.11 when the conditions of the appearance of the free recirculating bubble were established. These test runs were also carried out over range T_t = 1200 K - 1400 K. During the test runs the following measurements were done: 1) axial static pressure distributions on the top and bottom walls of the channel, 2) axial static pressure distribution on the wall of diffuser, 3) pre-heater pressure P_t, 4) air, hydrogen and oxygen pressures , and temperatures upstream of flow-meter nozzles, 5) pressure and mass rate of kerosene upstream of the aeration device, 6) fuel mixture pressure upstream of the injectors, 7) pressure of hydrogen (or air) supply into the vortex generator, 8) pressure of the throttling air in the diffuser, 9) videotaping of the flow in the vicinity of the exit of the rectangular channel made with the aid of the schlieren system.

2.5. Experimental results

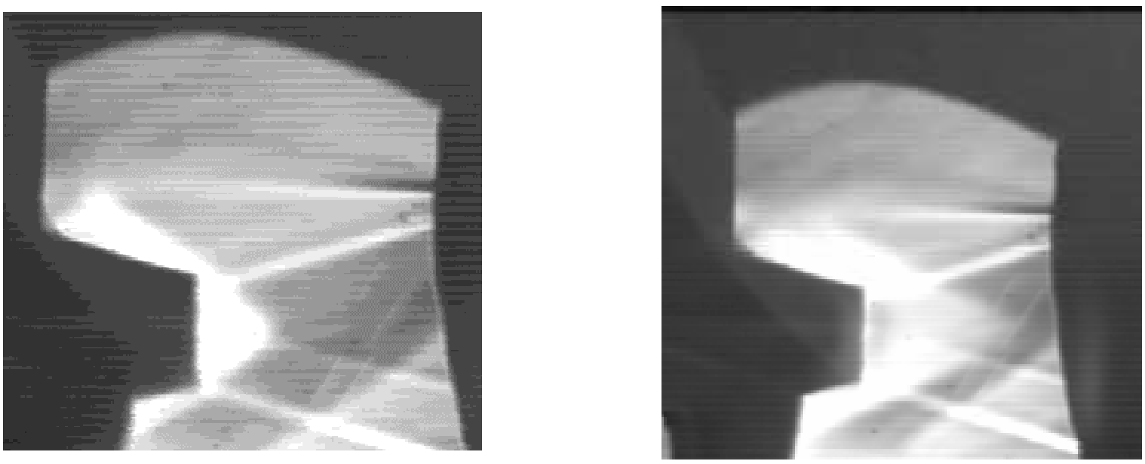
The list of the test runs and the conditions at which they were conducted are given in the Tab. (2.1). Stagnation pressure fields were measured at the vicinity of the channel exit in the two vertical sections located at the distances 18 mm and 50 mm downstream of the channel exit both with and without injectors, in runs No.1 and No.2, respectively.

Table 2.1. Tests conditions and tests results.

Run	P_t, MPa	T_t, K	M_{ent}	Injectors	Diffuser	Vortex	ER	Objectives and results
	3.85	1418	2.5	yes	no	no	0	Stagnation pressure fields at supersonic jet at distance x = 18mm and x = 50mm from the channel exit were measured.
No.1	3.81	1408	2.5	yes	no	no	0	
	3.79	1403	2.5	yes	no	no	0	
	3.81	1402	2.5	yes	no	no	0	
	3.83	1407	2.5	yes	no	no	0	
	3.87	1433	2.5	no	no	no	0	
	3.90	1441	2.5	no	no	no	0	
No.2	3.87	1425	2.5	no	no	no	0	
	3.89	1429	2.5	no	no	no	0	
	3.87	1424	2.5	no	no	no	0	
No.3	3.95	1197	2.5	no	yes	no	0	Bow shock in front of the shock wave-generator was obtained.
	3.93	1194	2.5	no	yes	no	0	
	3.91	1187	2.5	no	yes	yes	0	Free bubble in front of the shock wave-generator as a result of vortex and bow shock interaction was obtained.
No.4	3.91	1192	2.5	no	yes	yes	0	
	3.88	1185	2.5	no	yes	yes	0	
No.5	2.75	1220	2.5	yes	yes	yes	0	No self-ignition and combustion stabilization of aerated kerosene in free recirculation bubble.
	2.76	1210	2.5	yes	yes	yes	1.23	
No.6	2.75	1280	2.5	yes	yes	yes	0	
	2.72	1275	2.5	yes	yes	yes	0.77	
	3.80	1400	2.5	yes	yes	yes	0	Self-ignition and combustion stabilization of aerated kerosene in bubble was obtained. Outward propagation of combustion took place.
No.7	3.82	1407	2.5	yes	yes	yes	0	
	3.80	1401	2.5	yes	yes	yes	0	
	3.79	1396	2.5	yes	yes	yes	0.7	
	3.93	1433	2.5	yes	yes	yes	0	Self-ignition and combustion stabilization of pure (nonaerated) kerosene in free recirculation bubble was obtained. No outward propagation of combustion.
No.8	3.94	1436	2.5	yes	yes	yes	0	
	3.92	1423	2.5	yes	yes	yes	1.46	
	3.97	1461	2.5	yes	yes	yes	0	
No.9	3.96	1453	2.5	yes	yes	yes	0	
	3.97	1460	2.5	yes	yes	yes	1.56	
No.10	2.53	1190	2.5	yes	yes	no	0	Role played alone by bow shock in combustion stabilization was studied. It was shown that self-ignition was absent.
	2.54	1190	2.5	yes	yes	no	0.51	
No.11	2.66	1384	2.5	yes	yes	no	0	
	2.69	1406	2.5	yes	yes	no	0.9	

Stagnation pressures were measured by scanning horizontal ten-point rake, which was vertically moved by transversing equipment. Conclusions inferred from the measurements of the static pressure distributions along the channel walls and stagnation pressure field for run No.2 were the following (further details in Sabel'nikov et al., 1998; Sabel'nikov and Penzin, 2000): 1) static pressures on the walls close to the channel exit were within range 0.8 - 1.1 bar, i.e. the supersonic jet was nearly matched to atmospheric pressure; 2) stagnation pressure variation inside of the internal part of jet in vertical direction was negligible, and so the practically uniform supersonic flow core thickness was about 30 mm which is larger than the entrance diameter of the diffuser d = 25 mm. The Mach numbers, calculated on the base of measured stagnation pressures and static pressure equal 0.1 MPa were within range M = 2.5 - 2.8 and M = 2.57 - 2.85 at distances 18 mm and 50 mm downstream of the channel exit, respectively. Installation of injectors in run No. 1 resulted in 1.5 time increase of static pressure at the exit of channel in comparison with run No.2. For jet matched condition to be obtained at the channel exit in this case it was needed to decrease the pressure in pre-heater to P_t = 2.7 MPa. In run No.3 the methodology of bow shock wave generation was adjusted. Run was performed without injectors and vortex generator. Diffuser was located at the distance 50 mm downstream of the channel exit. Axial pressure distributions on channel and diffuser walls were measured for two cases: 1) without diffuser throttling, and 2) with diffuser throttling using injection of the air jets in the diffuser throat. These distributions showed (further details in Sabel'nikov et al., 1998; Sabel'nikov and Penzin, 2000) that the flow was supersonic inside of diffuser for the case without throttling. Bow shock was generated in front of the diffuser for the case with throttling. Pressure rise in the bow shock roughly corresponded to M = 2.5. Subsonic flow behind of bow shock wave accelerated and reached sonic velocity at the throttling section.

Generation of free recirculating bubble was performed in run No.4 at T_t = 1200 K. Test was carried out without injectors, with vortex-generator and diffuser with throttling of the latter. Angle of attack of vortex generator was 15 degrees. It was installed at distance 50 mm upstream of the channel exit. Diffuser was installed at the distance 50 mm downstream of channel exit. Two cases were considered: 1) with air supply into the vortex (through the vortex-generator); and 2) without air supply into vortex. Schliren pictures are presented in Figs. (2.7a), (2.7b), respectively. In first case air mass rate was 1.5 g/s (pressure supply was 1.0 MPa). It is seen from Figs. (2.7a,b) that in both cases free recirculating bubble arose in front of diffuser and that recirculating bubble is larger in the second case. Thus, this test shows that injection of high pressure gas into the vortex results in the decrease of vortex intensity, the decrease of dimension of recirculating bubble was a consequence of that. It can be assumed that some critical value (dependent, of course, on shock intensity and vortex structure) of air mass rate exists at excess of which breakup of vortex does not take place. In this case interaction of vortex with shock is referred as weak (see, e.g. Zatoloka et al., 1975) and only bending of shock is realized during such interference. Unfortunately, this interesting problem was outside of our investigation.



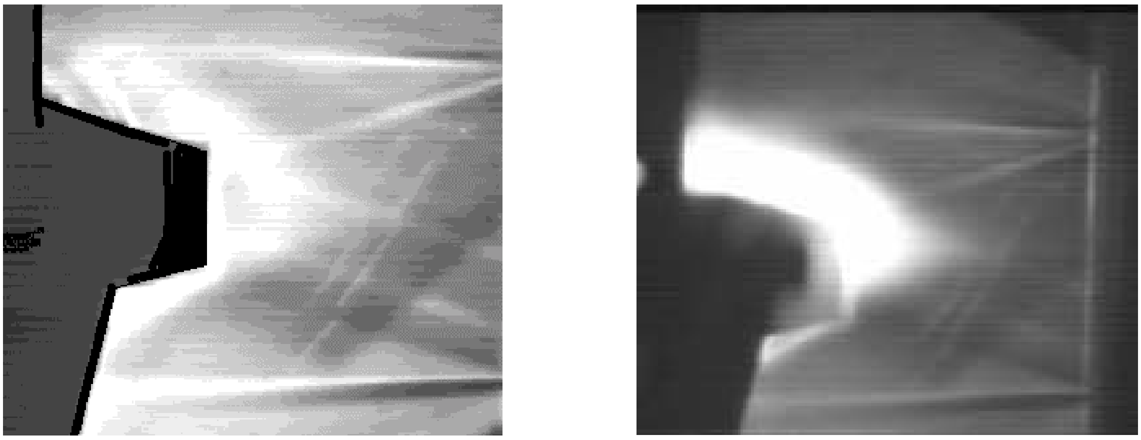
(a) (b)

Figure 2.7. Schlieren picture of flowfield in run No.4. a) air supply through vortex, b) no air supply through vortex

In the runs No.5-No.11 the aerated kerosene was injected into the channel through the four tube-injectors. Stagnation pressure fields at the vicinity of the channel exit for runs No.5 - No.11 were not measured, but rough estimation based on the results of the run No.1 showed that injectors installation resulted in the decrease of Mach number down to M = 2.0 at the distance 18 mm downstream of the channel exit.

Attempts to obtain the self-ignition and combustion stabilization of aerated kerosene with the aid of the free bubble were undertaken in runs No.5 and No.6 at T_t = 1200 K, and in run No.7 at T_t = 1400 K in the following sequence. At first, free recirculating bubble was generated in supersonic flow, Fig. (2.8a). Afterwards supply of the aerated kerosene throughout was performed and hydrogen was injected throughout vortex generator. For the case T_t = 1200 K the attempt was not successful. But run No.6 with T_t = 1400 K (ER = 0.7, mass rate of H_2 was 0.5 g/s) was succsefull. During this run, suddenly, in some region of free recirculating bubble self-ignition arose followed by the propagation of flame throughout the free recirculating bubble and at the final stage outward of zone to supersonic flow of kerosene-air mixture, Fig. (2.8b). Upstream propagation of flame was not observed. Approximately 3s after beginning of kerosene supply diffuser and holder were thermally damaged due to the intensive heat release in front of and in the vicinity of the diffuser.

In runs No.8 and No.9 the attempts were undertaken to obtain the self-ignition of the pure liquid kerosene (i. e. without aeration) at the same temperature $T_t = 1400$ K as in run No.7. Fuel equivalence ratio was ER $\approx$ 1.5. Diffuser was installed at the distance 25 mm downstream of the channel exit. In run No.8 air was supplied throughout the vortex generator. In run No.9 after some time delay the air supply was switched off, and H_2 was injected throughout the vortex-generator. The radiation of the light was observed mainly from the recirculating bubble, i.e. the burning in the external supersonic flow of air-kerosene mixture was practically absent. Possible reason of such a decreasing of the intensity of combustion, as compared to the run No.7, was the worsening of mixing between kerosene and air without aeration of the former.



(a) (b)

Figure 2.8. Schlieren picture of flowfield in run No.7. a) just before self-ignition, b) just after self-ignition.

The last two runs No.10 and No.11 were directed to learn the role of shock waves in the combustion stabilization. To this end these runs were conducted without the vortex-generator and, consequently without the free recirculating bubble but only with the bow shock in front of the diffuser. Tests were performed at two stagnation temperatures $T_t = 1200$ K (run No.10) and $T_t = 1400$ K (run No.11). Tests showed the absence of self-ignition of aerated kerosene in the vicinity of the diffuser for start and unstart regimes of the diffuser work. At $T_t = 1400$ K the self-ignition was observed in the far wake region downstream of the diffuser. Thus, it can be concluded that in the present experiment the free recirculating bubble in front of the diffuser played the decisive role in the self-ignition of the kerosene-air mixture.

2.6. Conclusions

Experiments were carried out to study the self-ignition and combustion stabilization of aerated by gas liquid kerosene in supersonic flow using a free recirculating bubble which was generated by the interaction of concentrated vortex with a normal shock wave. The following results were obtained:

1. Method of the combustion stabilization in supersonic flow using free recirculating bubble is proposed.
2. Aerodynamical model for experimental confirmation of combustion stabilization method for aerated kerosene in supersonic high-entalphy flow in the hypersonic facility T - 131B of TsAGI was designed and manufactured.
3. Dimensions of the aerodynamical model to realize the self-ignition and combustion stabilization were chosen using the criterion of stabilization of flame in the wake of a bluff body. It was found that the dimensions of the facility T - 131B allow to obtain the self-ignition and combustion stabilization of aerated liquid kerosene at the stagnation temperature $T_t \geq 1400$K.
4. Experiments confirmed the principles on which the design of aerodynamical model was based. At $T_t = 1400$ K the self-ignition and combustion stabilization of aerated kerosene was obtained in the nearly matched supersonic jet with the aid of the free recirculating bubble generated by the interaction of concentrated vortex and bow shock. Outward combustion propagation into supersonic flow was obtained (without upstream propagation). Self-ignition of pure liquid kerosene took place only inside of free recirculating bubble without outward propagation into supersonic flow.
5. Self-ignition and combustion stabilization of aerated kerosene in a free bubble at $T_t = 1200$ K - 1300 K was not achieved.
6. It was shown that the injection of the high pressure gas into the vortex center results in weakening of concentrated vortex and, as a consequence, the decrease of dimensions of free recirculating bubble. Hypothesis was put forward that some critical value of injection mass rate exists at excess of which free recirculating bubble will disappear.

3. The supersonic combustion enhancement using effervescent sprays and injectors with noncircular nozzles

This last section presents the results of the ducted combustion test program on the supersonic mixing and combustion enhancement in two-dimensional diverging-area supersonic combustor using aerated (by hydrogen or air) liquid kerosene jets (effervescent sprays) injected through elliptic nozzles from tube-injectors and fin-injectors (Sabel’nikov et al, 1998; Sabel’nikov and Penzin, 2000). Aerated kerosene jets were injected in two ways: 1) through the nozzles drilled in the tube-injectors at an angle of 45 degrees relative to the mainstream air flow direction, and 2) through the nozzles drilled at the base of the fin-injectors. In the latter case a co-flow injection was used. Tests were conducted at an entrance Mach number of M = 2.5 and a total temperature in the range T_t = 1650 - 1800 K. The experimental results for the elliptic nozzles were compared to the baseline results obtained with the circular nozzles. The axial static pressure distributions on the combustor walls as well as the combustion induced pressure-area integrals showed that the effervescent kerosene sprays performed better for elliptic nozzles than for round nozzles when injected from tube-injectors. Injection from fin-injectors did not show noticeable difference in combustion efficiency for elliptic and round nozzles.

3.1. Introduction

Combustion in supersonic combustor is considerably dependent (along with the kinetics) on the intensity of turbulent mixing. The factors that give to the methods of supersonic mixing enhancement a special significance are as follows: 1) the decrease of mixing intensity in supersonic flows; 2) the small residence time due to the length of the combustor that does not exceed 3 m and a flow speed which is in the range 1-2 km/s. Among enhancement techniques we can mention (see also the review paper of Gutmark et al.,1995, the papers of Haimovitch et al., 1994, 1997; Kopchenov and Lomkov, 1992): 1) the interaction between fuel jets, shock and expansion waves; 2) the use of injectors which geometry favors the generation of intense longitudinal vortices (e.g., NASA swept wedges); 3) the use of a noncircular nozzle geometry (e.g., elliptic nozzles) for the fuel supply. Up to now, listed above techniques of mixing intensification were used basically to accelerate the gaseous fuel jet mixing. The main mechanism of mixing intensification in the gaseous jet is the vortex-induced one and it is related to the excitation of large-scale modes of instability. Opportunities of such mechanism are apparently limited for jets of liquid fuel (e.g., kerosene, the promising fuel for small-dimension hypersonic vehicles). Hence, the idea of the aeration of liquid fuel jets by the gas (effervescent sprays) is considered as appealing. Investigations conducted at the Moscow Aviation Institute (MAI) by Avrashkov et al. (1990a, 1990b) and at TsAGI showed that effervescent sprays by their expansion angle are close to gaseous jets.

The main objective of the present investigation was to study the potential possibilities of supersonic mixing combustion enhancement using aerated (by hydrogen or air) liquid kerosene and injection of the last through the noncircular nozzles. Fuel was injected through elliptic nozzles from injectors of two geometry’s: 1) tube-injectors and 2) fin-injectors. Flow parameters at the combustor entrance were M = 2.5 and T_t = 1650 - 1800 K. Ignition delay lenghts, wall pressure distributions, combustion efficiencies and combustion induced pressure-area integrals are presented.

3.2. Experimental model. Test methodology

Tests were conducted using the supersonic combustor and hypersonic facility of MAI equipped with air-kerosene pre-heater (vitiated air). Oxygen mass fraction in the vitiated air Y^0_{O2} was slightly lower than in the atmospheric air. With an oxygen mass fraction in the atmospheric air of 0.232 the kerosene equivalence ratio (ER) in vitiated air in the combustor can be determined from the following equation

$$ER = (0.232/Y^0_{O2})\, L_0\, G_{ker}/G_1\,, \qquad (3.1)$$

where L_0 = 14.7 is the stoichiometric coefficient for kerosene combustion in atmospheric air, G_{ker} and G_1 are the mass rates of kerosene and vitiated air, respectively. Figure (3.1a) depicts the schematic view of the combustor. The combustor has four sections: 1) a 150 mm length section with a constant cross-area 52x104 mm^2, of the height h = 52 mm and of the width w = 2h; 2) a 150 mm length section with a divergence angle of 6.85 degree along upper wall leading to an exit cross-area of 70x104 mm^2; 3) a 300 mm length section with divergence angle of 1.9 degree along upper wall leading to an exit cross-area of 80x104 mm^2; 4) a 570 mm length section with a constant cross-area. Thus, the total length of the combustor is 1050 mm with an area-expansion ratio of 1.7. The flow from the combustor was exhausted into a still atmosphere. Axial pressure distributions were measured by taps placed on the upper and lower combustor walls.

Kerosene was aerated by hydrogen or air. Aeration device was similar to the one used for the combustion stabilisation investigation presented in the section 2. Scheme of the aeration device was shown earlier in Fig. (2.4). Mass fraction of gas used for aeration was small enough: indeed, while kerosene mass rate was 60 -130 g/s, hydrogen mass rate was about 1 g/s and air mass rate was about 10 g/s. Mixture pressure in the aeration device was in the range 1.5 - 2.5 MPa. The volume fractions of the kerosene and the gas at the nozzles exit of the injectors were of the same order of magnitude. Injection of effervescent kerosene sprays into the flow with a much lower static pressure level causes the explosion of the jet that promotes the mixing and vaporization of liquid kerosene (Avrashkov et al., 1990a, 1990b). Fuel was injected into the combustor in two ways: 1) at the angle of 45 degrees relative to the mainstream air flow direction throughout the tube-injectors; 2) in the co-flow direction with mainstream flow throughout the fin-injectors. The injectors were mounted in rows of four pieces on the upper and lower combustor walls. The distance between the combustor entrance and the injectors location was 105 mm. Fuel injection for both injector types was performed through either round nozzles of diameter 1.2 mm or elliptic nozzles with dimensions of principal axis 0.6 mm and 1.9 mm. Injector geometry's are given in Fig. (3.1b) - tube-injectors, and in Fig. (3.1c) - fin-injectors; their placement scheme is given in Fig. (3.1d). Total flow parameters and the other parameters characterizing the facility and the combustor operation regimes are given in Tab. (3.1).

Table 3.1. Tests parameters.

RUN	Nozzles	P_t, MPa	T_t, K	P_m, MPa	G_{air}, kg/s	G_{O2} kg/s	$G_{h,ker}$, kg/s	G_{ker}, kg/s	G_{bg}/G_{ker}	Y^0_{O2}	ER
Tube-injectors injectors, aeration by hydrogen											
1	Round	1.44	1690	*	2.125	0.522	0.103	0	*	0.16	0
1	Round	1.44	1650	1.5	2.125	0.522	0.103	0.132	~0.015	0.16	1.06
2	Elliptic	1.46	1790	*	2.185	0.257	0.101	0	*	0.1548	0
2	Elliptic	1.45	1780	2.18	2.185	0.257	0.101	0.08	~0.01	0.1548	0.69
3	Elliptic	1.41	1780	*	2.097	0.257	0.106	0	*	0.1539	0
3	Elliptic	1.42	1750	1.6	2.097	0.257	0.106	0.07	~0.01	0.1547	0.62
tube-injectors injectors, aeration by air											
9	Elliptic	1.46	1754	*	2.175	0.257	0.1189	0	*	0.1354	0
9	Elliptic	1.45	1793	2.29	2.125	0.2514	0.1182	0.094	0.16	0.1347	0.945
9	Elliptic	1.44	1756	1.96	2.115	0.2514	0.1187	0.075	0.18	0.1335	0.763
9	Elliptic	1.44	1756	1.86	2.116	0.2514	0.1184	0.07	0.19	0.1341	0.709
10	Round	1.43	1771	*	2.08	0.257	0.1254	0	*	0.1224	0
10	Round	1.45	1736	2.39	2.155	0.2514	0.1220	0.094	0.16	0.1304	0.962
10	Round	1.46	1775	2.16	2.137	0.2514	0.1212	0.075	0.18	0.1308	0.771
10	Round	1.44	1727	2.15	2.144	0.2514	0.1217	0.07	0.19	0.1304	0.720
fin-injectors injectors, aeration by air											
11	Round	1.43	1765	*	2.109	0.257	0.1226	0	*	0.1278	0
11	Round	1.45	1745	2.45	2.045	0.2514	0.1231	0.1011	0.16	0.1272	1.1
11	Round	1.42	1788	2.08	2.064	0.2514	0.1226	0.0787	0.18	0.1259	0.87
11	Round	1.43	1807	2.05	2.064	0.2514	0.1223	0.0782	0.19	0.1263	0.86
12	Elliptic	1.41	1732	*	2.077	0.257	0.1216	0	*	0.1286	0
12	Elliptic	1.42	1755	2.19	2.087	0.2536	0.1218	0.0984	0.16	0.1288	1.05
12	Elliptic	1.42	1764	1.95	2.082	0.2536	0.1216	0.0817	0.18	0.1288	0.87
12	Elliptic	1.42	1762	1.91	2.080	0.2536	0.1217	0.0795	0.19	0.1286	0.85

* - air supply through injectors for cooling the latter (performed for the cases of no fuel supply); G_{air} - air mass rate through pre-heater, kg/s; G_{O2} - oxygen mass rate through pre-heater, kg/s; $G_{h,ker}$ - kerosene mass rate through pre-heater, kg/s; G_{ker} - kerosene mass rate through supersonic combustor, kg/s; P_m - total pressure of kerosene-gas mixture in injectors; G_{bg}/G_{ker} – aeration gas mass rate to kerosene mass rate ratio.

Test runs No.1-No.3 were carried out at fixed fuel equivalence ratios. Kerosene was aerated by hydrogen. For other four test runs No.9-No.12 with hydrogen used for aeration of the kerosene the value of ER was changing continuously during the tests in the following sequence: after reaching the desired combustor entrance conditions the fuel was injected during 20 - 30 s. During this time interval the magnitude of ER was gradually decreased and changed from the values nearly stoichiometric to the values at which the combustion blowout took place.

To ignite the combustor, the high pressure throttling air jets were injected during 0.5 - 1.0 s in the section located at the distance 780 mm from the combustor entrance, Fig. (3.1a). After ignition the air throttling jets were switched off. During the tests, the axial pressure distributions on the upper and lower walls of combustor were measured. In test run No. 2 the total pressure field was measured in the combustor exit plane. Measurements were carried out by 10-point transversing rake. In the other test runs, the total and static pressures were measured in a single point at the combustor exit plane.

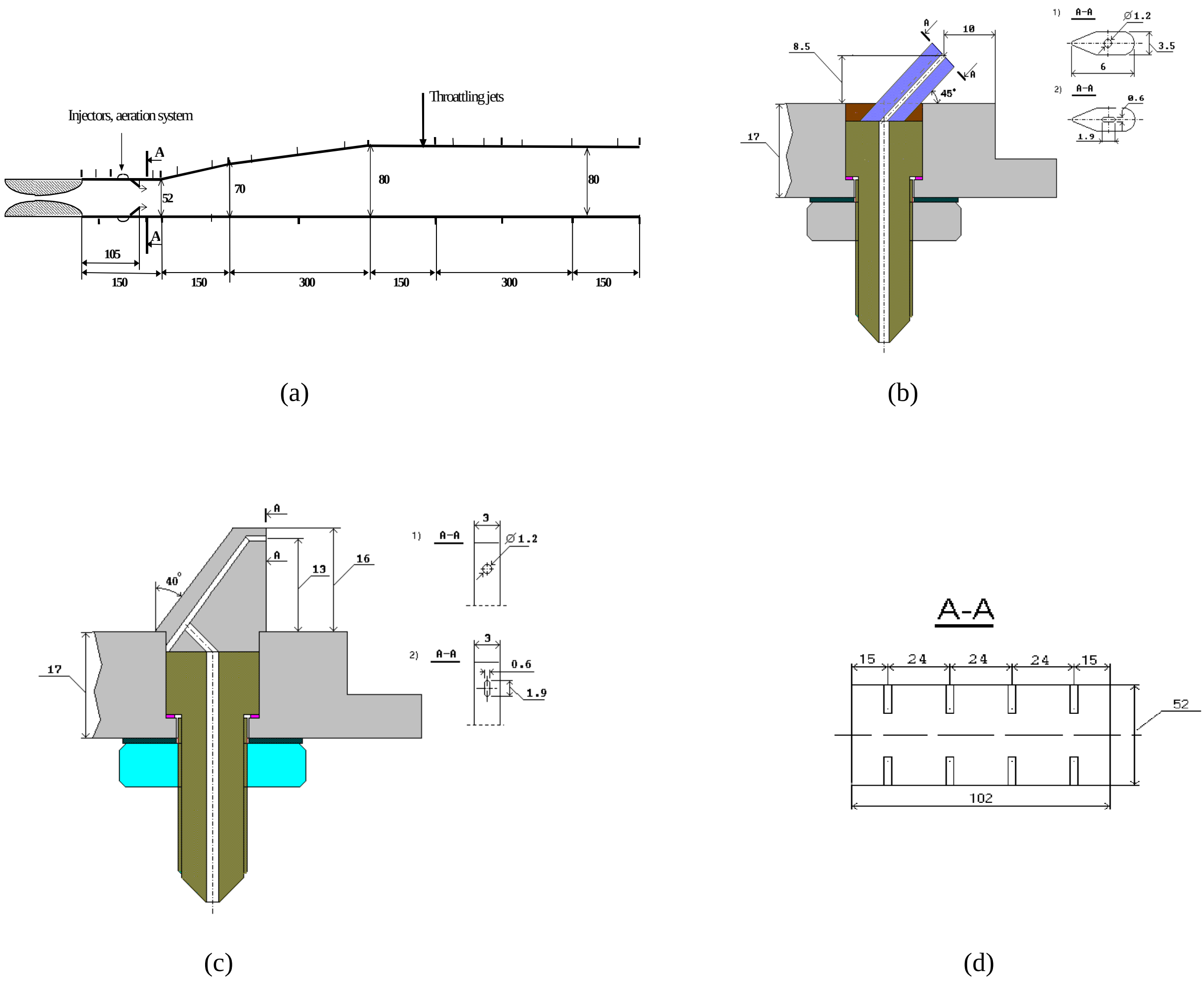


Figure 3.1. (a) Scheme of supersonic combustor, (b) Geometry of tube-injector, (c) Geometry of fin-injector, (d) Scheme of injectors location (all dimensions in mm).

Experimental data were analyzed using 1-D method. This method is based on the solution of the conservation equations of the energy, mass and impulse at known (from experiment) axial pressure distributions on the walls of the combustor (pressure is assumed constant over cross-sections of the combustor). The 1-D calculation results for the section located at the distance 900 mm from combustor entrance are given in Tab. (3.2).

3.3. Tests results

3.3.1. Tube-injectors

Tube-injectors were studied in the runs No.1-No.3 and No.9-No.10. Kerosene was aerated by hydrogen in the runs No.1-No.3 and by air in the runs No.9-No.10 (see Tab. (3.1)). Figure (3.2) compares the axial normalized static pressure (static pressure P divided by the pressure P_t in the pre-heater) distributions on the combustor walls for the runs No.9 (elliptic nozzles) and No.10 (round nozzles) with practically the same values of ER for elliptic and round nozzles. Recall that the fuel was injected at the angle of 45 degrees relative to the mainstream air flow direction throughout tube-injectors. As 1-D calculations showed, the flow in the combustor remained supersonic in the test runs No.9 and No.10 over range of ER given in Tab. (3.2). It can be seen from Fig. (3.2) that for the tests with the combustion the values of the pressure along the length of the combustor are almost everywhere higher in the case of the elliptic nozzles, i. e. the enhancement of the supersonic combustion took place when aerated kerosene jets were injected through elliptic nozzles. Figure (3.2) shows that at the aft of the combustor, a flow separation occurred for the case without combustion, i. e. at ER = 0 (due to the over-expansion of the flow). It is also seen from Fig. (3.2) that after some ignition delay (for the combustion cases) pressure increased monotonously along the combustor (with the exception of the aft of the combustor). Figure (3.3) shows the dependence of ignition delay lengths on ER. It can be concluded that the ignition delay lengths were shorter for the case of injection of aerated kerosene through elliptic nozzles.

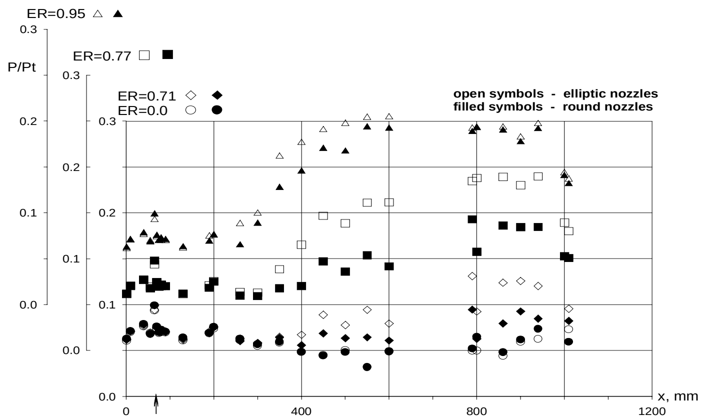
Table 3.2. Flow parameters at distance of 900 mm from the combustor entrance.

Run	Nozzles	ER	M	combustion efficiency	total pressure recovery coefficient
tube-injectors injectors, aeration by hydrogen					
1	Round	1.06	0.98	1	0.368
2	Elliptic	0.69	1.02	1	0.354
3	Elliptic	0.62	1.08	1	0.345
tube-injectors injectors, aeration by air					
9	elliptic	0.945	1.06	1	0.344
9	elliptic	0.763	1.05	1	0.334
9	elliptic	0.709	1.5	0.7	0.352
10	round	0.962	1.1	0.98	0.338
10	round	0.771	1.33	0.8	0.342
10	round	0.720	1.9	0.4	0.322
fin-injector injectors, aeration by air					
11	round	1.1	1.1	1	0.336
11	round	0.87	1.1	0.93	0.33
11	round	0.86	1.2	0.79	0.34
12	elliptic	1.05	1.1	1	0.337
12	elliptic	0.87	1.1	0.95	0.333
12	elliptic	0.85	1.2	0.89	0.331

The most simply, the efficiency of the supersonic combustion enhancement can be analyzed using a local characteristic – the ratio of the difference of pressure rises due to combustion for elliptic and round nozzles, respectively, to the difference of the pressure rise due to combustion for the baseline configuration, i.e. for the round nozzles

$$\Delta\overline{P} = (P_{ell}-P_{round})/(P_{round}-P_{no\ combustion}). \tag{3.2}$$

Here $P_{no\ combustion}$ is the wall static pressure distribution without combustion, P_{ell} and P_{round} are the combustion induced pressure distributions for the elliptic and round nozzles, respectively. The results of calculation of $\Delta\overline{P}$ using experimental data given in Fig. (3.2) are presented in Fig. (3.4). It can be seen that $\Delta\overline{P} > 0$ along the length of the combustor, i. e. the elliptic nozzles provide better combustion performance than round nozzles.



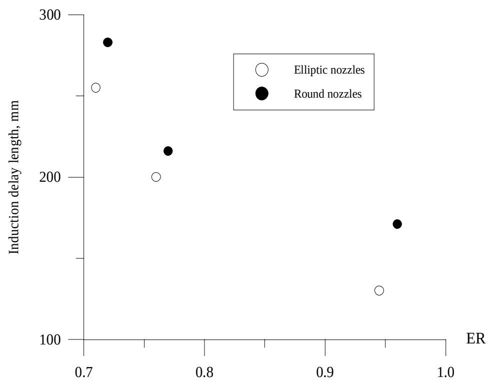
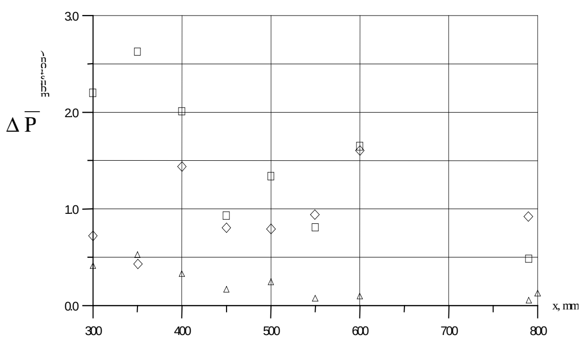


Figure 3.2. (to the left) Wall pressure distributions for tube-injectors, arrow shows the place of fuel supply.
Figure 3.3. (to the right) Ignition delay lengths for tube-injectors.

But the better quantitative indicator of supersonic mixing and combustion enhancement is obtained through the analysis of the impact of the fuel supply mode on the particular integral characteristic of the diverging-area supersonic combustor – pressure-area integral (see, e.g. Kay et al., 1990; Stouffer et al., 1994). The combustion induced pressure-area integrals were calculated from the measured axial wall pressure distributions using the following relationship (see, e. g. Stouffer et al., 1994)

$$\Delta F = w \int (P_{combustion} - P_{no\ combustion}) tg\theta dx, \quad (3.3)$$

where θ is the local wall angle with respect to flow and x is the axial coordinate. Figure (3.5) shows the comparison of the normalized combustion induced pressure-area integrals $\Delta\,\overline{F} = \Delta F / I_1$ for elliptic and round nozzles, here $I_1 = (P+\rho u^2)_1 hw$ is the axial impulse function at the combustor entrance. It is seen that the magnitudes of $\Delta\,\overline{F}$ increase with increasing of ER (as it should be) for both type of nozzles, and that $\Delta\,\overline{F}$ are higher for elliptic nozzles with respect to the round nozzles.



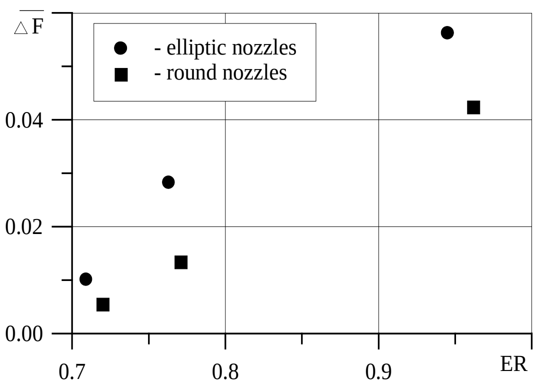
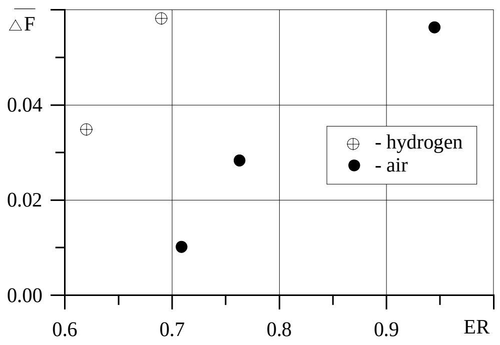


Figure 3.4. (to the left) Normalized difference of combustion induced pressure rises for elliptic and round nozzles (for caption see Fig. 3.2).
Figure 3.5. (to the right) Comparison of normalized combustion induced pressure-area integrals for tube-injectors with elliptic and round nozzles, air-kerosene effervescent sprays.



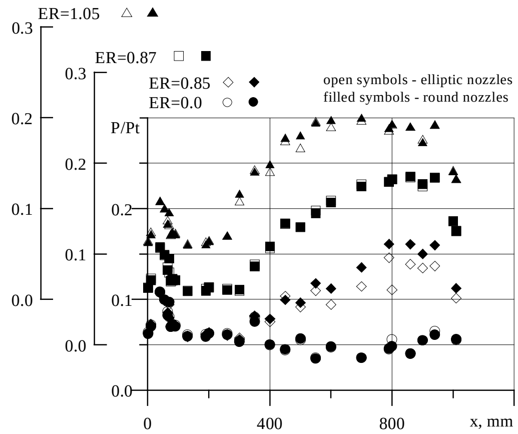


Figure 3.6. (to the left) Impact of type of the gas for kerosene aeration on normalized combustion induced pressure-area integral for tube-injectors with elliptic nozzles.

Figure 3.7. (to the right) Axial static pressure distributions on combustor wall for fin-injectors; open and filled symbols for elliptic and round nozzles, respectively.

To study the influence of the type of gas used for the aeration of the kerosene on the combustion and mixing enhancement hydrogen was used for the aeration of kerosene in test runs No.1-No.3 (see Table 3.1). The comparison of the values of $\Delta \overline{F}$ obtained with hydrogen and air as aerated gas is illustrated in Fig.3.6. One can conclude that aeration of kerosene by hydrogen provided better mixing and combustion enhancement than aeration by air. The possible reasons of greater hydrogen aeration efficiency are the following: 1) greater specific work capacity of hydrogen during expansion compared to that of the air; 2) favorable influence of hydrogen on combustion kinetics of kerosene. The last factor is hardly possible since hydrogen fraction in the mixture is quite low (about 1 percent).

3.3.2. Fin-injectors

The fin-injectors were studied in the runs No.11 and No.12. Air was used for the aeration of the kerosene. Effervescent kerosene sprays were injected through round (test run No.11) and elliptic nozzles (test run No.12) located at the base of fin-injectors in the co-flow direction to the mainstream flow. Figure (3.7) compares the axial pressure distributions (with practically the same values of ER for elliptic and round nozzles) for two tests. One can conclude from Fig. (3.7) that the combustion induced pressure rises for elliptic and round nozzles are nearly the same, i.e. the mixing and combustion efficiencies practically coincide for both types of nozzles. This conclusion is confirmed by the calculation of the combustion induced pressure-area integrals for both types of nozzles, Fig. (3.8).

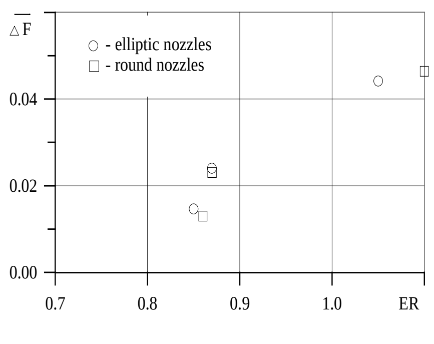


Figure 3.8. Comparison of normalized combustion induced pressure-area integral for fin-injectors with elliptic and round nozzles, air-kerosene effervescent sprays.

3.3. Conclusions

An experimental study was carried out to study the supersonic mixing and combustion enhancement in scramjet combustor using aerated by gas liquid kerosene jets (effervescent sprays) injected through elliptic nozzles from tube-injectors and fin-injectors. The following results were obtained:

1. Elliptic nozzles provided greater mixing and combustion efficiencies in comparison with round nozzles for the cases when aerated kerosene was injected from tube-injectors at the angle of 45 degrees relative to the mainstream air flow direction.

2. Aeration of kerosene by hydrogen provided higher mixing and combustion enhancement compared to aeration by air at injection from tube-injectors.
3. Co-flow injection of aerated kerosene throughout the fin-injectors did not show noticeable difference in the mixing and combustion efficiencies for round and elliptic nozzles.
4. Injection from tube-injectors at an angle of 45 degrees relative to the mainstream air flow direction provided greater mixing and combustion efficiencies in comparison to co-flow injection from fin-injectors.

The investigation showed that the use of the effervescent sprays and elliptic nozzles for the injection of the fuel enable to realize the efficient supersonic combustion of the liquid hydrocarbon fuels. It would be interesting in future work to study the possibilities of the supersonic combustion enhancement using the elliptic nozzles drilled at the base of NASA swept wedges.

4. Acknowledgments

The content of this paper represents the sustained interest of the author in the problems associated with supersonic turbulent combustion, as well as the collaboration with several other investigators. The work presented in the section 1 was accomplished while the author was an Associate Professor at ENSMA (École Nationale Supérieure de Mécanique et d'Aérotechnique) and LCD (Laboratoire de Combustion et de Détonique), France. The work presented in the section 2 was supported by USA Air Force Office of Scientific Research (AFMC), EOARD, contract No. SPC-96-4043, with Dr. W.L. Bain as the technical monitor. The work presented in the section 3 was supported by NAVY USA, Office of Naval Research contract No. N00014-96-1-0869, with Dr. G. Roy as the technical monitor. The author is appreciative of the support he has received from the Russian Fund of Basic Research (RFBR, grant No. 00-03-32066).

5. References

Avrashkov, V.N., Baranovsky, S.I., and Davidenko, D.M., 1990a, "Penetration height of a liquid jet saturated by gas bubbles", Izvestiya Vuzov, Aviatsionnaya Tekhnika, No.4, pp. 96-98, (In Russian).

Avrashkov, V.N., Baranovsky, S.I. and Levin, V.M., 1990b, "Gasdynamic features of supersonic kerosene combustion in model combustion chamber", AIAA Paper 90-5268, 7 p.

Baev, V.K., Golovichev, V.I., Tret'yakov, P.K., Garanin, A.F., Konstatinovsky, V.A. and Yasakov, V.A. "Combustion in Supersonic Flow", 1984, Nauka, Moskow, Novosibirsk, (In Russian).

Balakrishnan, G. & Williams, F.A., 1993, "Turbulent combustion regimes for hypersonic propulsion employing hydrogen-air diffusion flames", Journal of Propulsion and Power, Vol. 10, No. 3, pp. 434-436.

Bilger, R.W., 1976, "The structure of diffusion flames", Combust. Sci. Technol., Vol.13, pp. 155-170.

Bilger, R.W., 1989, "Turbulent diffusion flames", Ann. Rev. Fluid. Mech., Vol. 21, pp. 101-135.

Bilger, R.W., 1993, "Conditional moment closure for turbulent reacting flow", Phys. Fluids A, Vol. 5, pp. 436-444.

Bish, E.S. and Dahm, W.J.A., 1995, "Strained dissipation and reaction layer analysis of nonequilibrium chemistry in turbulent reacting flows", Combustion and Flame, Vol. 100, pp.457-464.

Borghi, R., 1988. "Turbulent combustion modeling", Prog.Energy Combust. Sci., Vol. 14, pp. 245-292.

Bray, K.N.C., Libby, P.A. and Williams, F.A., 1994, "High speed turbulent combustion", In: P.A.Libby and F.A.Williams (Eds.), "Turbulent Reacting Flows", Academic Press Inc., San Diego, 609-638.

Bray, K.N.C. and Peters, N. 1994. "Laminar flamelets in turbulent flames", In: P.A.Libby and F.A.Williams (Eds.), "Turbulent Reacting Flows", Academic Press Inc., San Diego, pp. 63-113.

Buriko, Y.Y., Kuznetsov, V.R., Volkov, D.V., Zaitsev, S.A. and Uryvsky, A.F., 1994, "A test of flamelet model for turbulent nonpremixed combustion", Combustion and Flame, Vol. 96, pp. 104-120.

Chen, J.-Y. and Kollmann, W., 1994, "Comparison of prediction and measurements in nonpremixed turbulent flames", In: P.A.Libby and F.A.Williams (Eds.), Turbulent Reacting Flows, Academic Press Inc., San Diego, pp. 211-300.

Cheng, T.S., Wehrmeyer, J.A., Pitz, R.W., Jarrett, O.Jr. and Northam, G.B., 1994, "Raman measurements of mixing and finite-rate chemistry in a supersonic hydrogen-air diffusion flame", Combustion and Flame, Vol. 99, pp. 157-173.

Curran, E.T., Heiser, W.H. and Pratt, D.T., 1996, "Fluid phenomena in scramjet combustion systems", Ann. Rev. Fluid Mech., Vol. 28, pp. 323-360.

Darabiha, N., 1992, "Transient behavior of laminar counter-flow hydrogen-air diffusion flames with complex chemistry", Combust. Sci. and Technol., Vol. 86, pp. 163-181.

Delery, J., Horovitz, E., Leuchter, O. and Solingac, J., 1984, "Etudes fondamentales sur les écoulements tourbillonnaires", La Recherche Aérospatiale, No.2, pp. 81-104.

Eifler, P. and Kollmann, W., 1993, "Pdf predictions of supersonic hydrogen flames", AIAA Paper 93-0448, 31st Aerospace Sciences Meeting, Reno, NV.

Eklund, D.R., Drummond, J.Ph. and Hassan, H.A., 1989, "Numerical modeling of turbulent supersonic reacting coaxial jets", AIAA-89-0660.

Evans, J.S., Schexnayder, C.J. and Beach, H.L., 1978, "Application of a two-dimensional parabolic computer program to prediction of turbulent reacting flames", NASA TP 1169.

Everest, D.A., Driscoll, J.F., Dahm, W.J.A. and Feikema, D.A., 1995, "Images of the temperature field and temperature gradients to quantify mixing rates within a nonpremixed turbulent jet flame", Combustion and Flame, Vol. 101, pp. 58-68.

Farshchi, M., 1990, "Chemically reacting supersonic flow calculation using an assumed pdf model", AIAA-90-0731.

Figueira da Silva, L.F., Sabel'nikov, V.A., and Deshaies, B., 1997, "The stabilization of supersonic combustion by a free recirculating bubble: a numerical study.", AIAA Journal, Vol. 35, No.11, pp.1782-1784.

Gutmark, E.J, Schadow, K.S., and Yu, K.H., 1995, "Mixing enhancement in supersonic free shear flows", Ann. Rev. Fluid Mech., Vol. 27, pp. 375-417.

Glotov, G.Ph., 1989, "Interference of concentrated vortex with shock waves in free flow and nondesigned jets", Uchenye Zapiski TsAGI, Vol. 10, No. 5, pp. 21-32 (In Russian).

Haimovitch, Y., Gartenberg, E., and Roberts, A.S. Jr.,1994, "Investigation of rump injector for supersonic mixing enhancement", NASA-CR-4634,1994.

Haimovitch, Y., Gartenberg, E, Roberts, A.S. Jr., and Northam, G.B.,1997, "Effects of internal nozzle geometry on compression-ramp mixing in supersonic flow", AIAA Journal, Vol. 35, No.4, pp. 663-670.

Haworth, D.C., Drake., M.C., Pope, S.B. and R.J. Blint., 1988, "The importance of time-dependent flame structures in stretched laminar flamelet models for turbulent jet diffusion flames", 22nd Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp. 589-597.

Im., H.G., Bechtold, J.K. and Law, C.K., 1995, "Counterflow diffusion flames with unsteady strain rates", Combust. Sci. and Technol., Vol. 106, pp. 345-361.

Ivanyushkin, A.K. and Korotkov, Yu.V., 1995, "Influence of flow vorticity on diffuser start at supersonic and hypersonic speeds", TsAGI Workshop - School Fluid Mechanics : "Research in Hypersonic Flows and Hypersonic Technologies", TsAGI, pp. 27-28.

Ivanyushkin, A.K., Korotkov, Yu.V. and Nikolaev, A.V., 1989, "Some peculiarities of interference of shock waves with aerodynamic wake behind body", Uchenye Zapiski TsAGI, Vol. 10, No. 5, pp. 33-42, (In Russian).

Jouve, P.-A., 1996, "Modélisation de la combustion supersonique turbulente dans des configurations de zones de mélange bidimesionnelles planes", Ph.D. Thesis, Ecole Centrale de Paris.

Kalkhoran, I.M., Smart, K.M. and Betti, A., 1996, "Interaction of supersonic wing-tip vortices with a normal shock", AIAA Journal, Vol. 34, No.9, pp.1855-1861.

Kay, I.W., Peschke, W.T., and Guile, R.N., 1990, "Hydrocarbone-fuelled scramjet combustor investigation", AIAA Paper 90-2337, 8p.

Klimenko, A.Y., 1990, "Multicomponent diffusion of various admixtures in turbulent flow", Fluid Dyn., Vol. 25, pp. 327-334.

Kollmann, W., 1990, "The pdf approach to turbulent flow", Theor. Comput. Fluid Dyn., Vol. 1, 249-285.

Kopchenov, V.I., and Lomkov, K., 1992, "The enhancement of the mixing and combustion processes applied to scramjet-engine", AIAA Paper 92-3428, 8 p.

Kuznetsov, V.R. and Sabel'nikov, V.A., 1990, "Turbulence and Combustion", New York, Hemisphere.

Lele, S.K., 1994, "Compressibility effects on turbulence", Ann. Rev. Fluid Mech., Vol. 26, pp. 211-254.

Lentini, D., 1994, "Assessment of the stretched laminar flamelet approach for nonpremixed turbulent combustion", Combust. Sci. and Technol., Vol. 100, pp. 95-122.

Mell, W.E., Nilsen, V., Kosaly, G. and Riley, J.J., 1994, "Investigation of closure models for nonpremixed turbulent reacting flows", Physics of Fluids, Vol. 6, No. 3, pp. 1331-1356.

Meshcheryakov, E.A. and Sabel'nikov, V.A., 1988, "Reduced heat production due to mixing and kinetic factors in kinetics factors in supersonic combustion of nonpremixed gases in an expanding channel", Fiz. Goreniya i Vzryva, Vol. 24, No. 5, pp. 23-32.

Mettwally, O., Settles, G., Horsman, C., 1989, "An Experimental Study of Shock Wave/Vortex Interaction.", AIAA Paper 89-0082, 12 p.

Michael, K., Smart, K.M. and Kalkhoran, I. M., 1995, "Effect of Shock Strength on Oblique Shock-Wave/Vortex Interaction", AIAA Journal, Vol. 33, No.11, pp. 2137-2143.

Morgenthaler, V., Figueira da Silva, L.F., Deshaies, B. and Sabel'nikov, V.A., 1997, "Extended flamelet approach for supersonic combustion of nonpremixed gases", In: "Advanced Computation and Analysis of Combustion", International Colloquium, Moscow, ENAS Publishers, Moscow, Russia, pp. 267-283.

Morgenthaler, V., Figueira da Silva, L.F., Sabel'nikov, V.A. and Deshaies, B., 2001, "A numerical study of supersonic turbulent hydrogen-air jets using non-premixed laminar flamelet modeling. Influence of the kinetic to thermal energy conversion", Combustion and Flame (submitted).

Peters, N., 1984, "Laminar diffusion flamelets models in nonpremixed turbulent combustion", Progr. Energy Combust. Sci., Vol. 10, pp. 319-339.

Peters, N., 1986, “Laminar flamelets concepts in turbulent combustion”, 21st Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp. 1231-1250.
Pope, S.B., 1985, “Pdf methods for turbulent reacting flows”, Progr. Energy Combust. Sci., Vol. 11, pp. 119-192.
Pope, S.B., 1990, “Computations of turbulent combustion: progress and challenges”, 23rd Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp. 591-612.
Sabel’nikov, V.A., 1997, “Supersonic turbulent combustion of nonpremixed gases - status and perspectives”, In: “Advanced Computation and Analysis of Combustion”, International Colloquium, Moscow, ENAS Publishers, Moscow, pp. 208-237.
Sabel’nikov, V.A. and Penzin., V.I., 2000, “Scramjet Research and Development in Russia”, In : “Scramjet Propulsion”, E.T. Curran and S.N.B. Murthy (Editors), AIAA Progress Series, Chapter 5, pp. 223-367.
Sabel’nikov, V.A., Figueira da Silva, L.F. and Dehaies B., 1997, “Revisited flamelet model for nonpremixed combustion in supersonic turbulent flows”, Combustion and Flame, Vol. 114, No.3/4, pp.577-584.
Sabel’nikov, V.A., Korontsvit, Yu. Ph., Ivanyushkin, A.K. and Ivanov, V.V., 1998a., “Experimental investigation of combustion stabilization in supersonic flow using free recirculation zone”, AIAA Paper 98-1515, Norfolk, USA, 8th International Spaces and Hypersonic Systems and Technologies Conference, April 1998. A collection of technical papers, pp. 93-103.
Sabel’nikov, V.A., Korontsvit, Yu. Ph., Schadow, K., Ivanov, V.V. Voloschenko, O.V. and Zosimov, S.A., 1998b, “Investigation of supersonic combustion enhancement using barbotage and injections with noncircular nozzles”, AIAA Paper 98-1516, Norfolk, USA, 8th International Spaces and Hypersonic Systems and Technologies Conference, April 1998. A collection of technical papers, pp. 104-113.
Shchetinkov, E.S., 1965, “ Physics of Gas Combustion”, Nauka, Moscow, (In Russian).
Smith, L.L., Dibble, R.W., Talbot, L., Barlow, R.S. and Carter C.D.,1995, “Laser Raman scattering of differential molecular diffusion in nonreacting turbulent jets of H_2 / CO_2 mixing with air”, Physics of Fluids, Vol. 7, No. 2, pp. 1455-1466.
Smits, A.J. and Dussauge, J.-P., 1996, “Turbulent Shear Layers in Supersonic Flow”, AIP Press, Woodbury, New York.
Speziale, C.G. and Sarkar, S., 1991, “Second-order closure model for supersonic turbulent flows”, ICASE Report 91-9.
Stouffer, S.D., Vandsburger, U., and Northam, G.B., 1994, “Comparison of wall mixing concepts for scramjet combustors”, AIAA Paper 94-0587, 13p.,1994.
Thévenin, D. and Candel, S., 1994, “Diffusion and premixed flame ignition dynamics in a field of variable strain rate”, 24th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp.1547-1554.
Villasenor, R., Chen, J.-Y. and Pitz, R.W., 1992, “Interaction between chemical reaction and turbulence in supersonic nonpremixed H_2 -air combustion”, AIAA Journal, Vol. 30, No. 10, pp. 2552-2554.
Westbrook,C.K. and Dryer, F.L., 1984, “Chemical kinetic modeling of hydrocarbon combustion”, Progress in Energy and Combustion Science”, Vol. 10, pp. 1-57.
Winterfeld, G., 1968, “On the burning limits of flame-holder-stabilized flames in supersonic flow”, AGARD, IX, Vol. 2, No. 34, p. 12.
Zakkay,V. and Sinha, R., 1970, “Residence time within a wake recirculation region in an axisymmetric supersonic flow”, AIAA Paper 70-111, 11 p.
Zatoloka, V.V., Ivanyushkin, A.K. and Nikolaev, A.V.,1975, “Interference of concentrated vortices with shock wave in inlet. Breakup of vortices”, Uchenye Zapiski TsAGI, Vol. 6, No. 2, pp. 134-138, (In Russian).
Zatoloka, V.V., Ivanyushkin, A.K. and Nikolaev, A.V., 1978, “Interference of vortexes with shocks in airscoops. Dissipation of vortexes”, Fluid Mechanics - Soviet Research, Vol. 7, No. 4, July - August. pp. 153-158.
Zeman, O., 1990, “Dilatation dissipation: the concept and application in modeling compressible mixing layers”, Physics of Fluids A, Vol. 2, No. 1, pp. 170-188.
Zheng, L.L. and Bray, K.N.C., 1992, “Effects of dilatation dissipation on turbulent shear layer combustion in high speed flow”, 24th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, pp. 405-411.
Zheng, L.L. and Bray, K.N.C., 1994, “The application of new combustion and turbulence models to H_2 -air nonpremixed supersonic combustion”, Combustion and Flame, Vol. 99, pp. 440-448.
Zheng, L.L. and Bray, K.N.C., 1997, “Effects of laminar flamelet structures on supersonic turbulent combustion”, In: M.Champion and B.Deshaies (Eds.), IUTAM Symposium on Combustion in Supersonic Flows, Kluwer Academic Piblishers, Inc., Doordrecht, pp. 111-117.
Zimont, V.L., Levin, V.M., Meshcheryakov, E.A. and Sabel’nikov, V.A., 1983, “Characteristics of supersonic combustion of nonpremixed gases in channels”, Combustion, Explosion and Shock Waves, Vol. 19, No. 4, pp. 441-443.

THE ROLE OF TECHNOLOGICAL DEVELOPMENT ON THE BRAZILIAN COMPANY EMBRAER

Hugo Borelli Resende
EMBRAER
hresende@embraer.com.br
Brazil

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

LATTICE-GAS MODELS FOR FLUID FLOW

P.C. Philippi
Porous Media and Thermophysical Properties Laboratory (LMPT)
Mechanical Engineering Department
Federal University of Santa Catarina
88040-900 Florianópolis SC Brazil
e-mail: philippi@lmpt.ufsc.br

L.O. E. Santos, S.L. Bertoli*, P.C. Facin• F. Wolf
* Permanent address: Chemical Engineering Department, Regional University of Blumenau, 89010-971, Blumenau, SC, Brazil
• Permanent address: Department of Physics, State University of Ponta Grossa, Ponta Grossa, Pr, Brazil

Abstract: *Lattice-gas models are to be considered as lower-level models with respect to macroscopic scale: fluids are supposed to be represented by large systems of particles, moving along the directions of a regular lattice and colliding at lattice vertices. In this paper, two kinds of lattice-gas models are presented: the Boolean microscale models and Bolltzmann mesoscale models. In their simplest form, Boolean models consist of regular lattices populated with particles that hop from site to site in discrete time steps in a process, often, called propagation. After propagation, particles in each site interact with each other in a process called collision, in which the number of particles and momentum are preserved. An exclusion principle is imposed in order to achieve better computational efficiency and a Boolean variable* n_i *(r, t) is assigned to each direction in a site to indicate the presence (*$n_i=1$*) or absence (*$n_i=0$*) of a particle in that direction. Based on Bhatnagar, Gross and Krook (BGK) relaxation term, Boltzmann models have the same principles but work with real variables related to the ensemble average values of* $N_i = \langle n_i (\mathbf{r}, t)\rangle$*. A four-bit Boolean lattice gas model and a Boltzmann model based on field mediators are presented for simulating the flow of immiscible fluids. Present field mediators simulate long-range action, enabling the use of local rules in separation step and, avoiding optimization step, reduce computer processing-time with respect to previous models. In addition, field strength and interaction distance is modeled by introducing distinct emission,* P_e*, and extinction,* P_a*, probabilities, for field mediators, enabling to control interfacial tension and transition thickness. Model's microdynamics is fully described.. Simulation results are presented for several sample case studies.*

Keywords*: lattice-gas models, lattice Boltzmann, miscible fluids, immiscible fluids,*

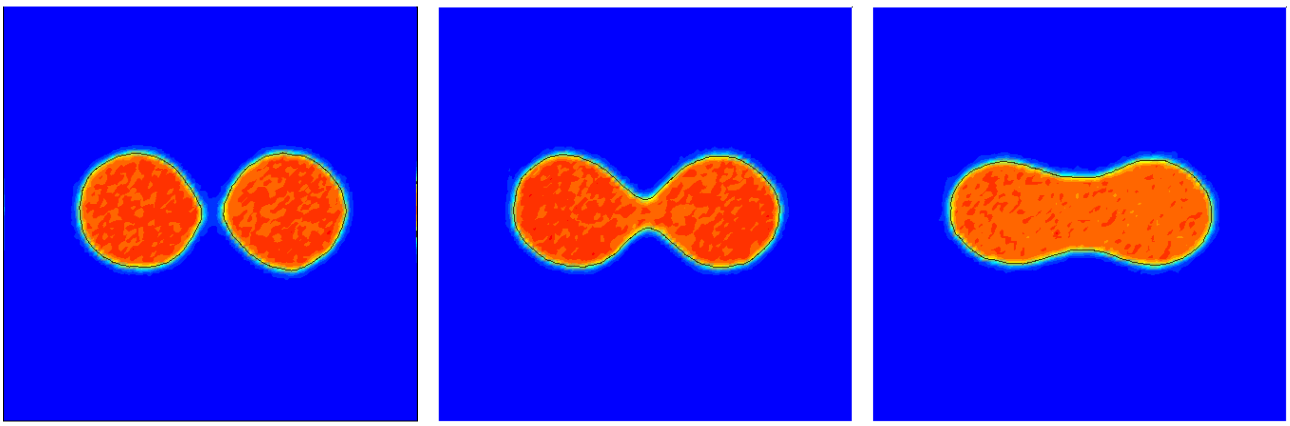
1. Introduction

Considering the macroscopic description of fluid flow, lattice gas models are to be considered as *lower-level* models based on the statistical behavior of a large set of particles moving along the discrete directions of a regular lattice and colliding in lattice vertices.

In fact, there are important physical phenomena related to fluid flow that are difficult to, or cannot, be described by considering the only information given by the macroscopic equations, at the macroscopic level. In these cases *downscaling* to a lower scale is, often, necessary for the correct understanding and mathematical description of the particular physical phenomena of interest, that needs a more refined treatment.

Some of these phenomena, related to interfacial dynamics, are presented bellow

When two droplets of a fluid r are put very closely, Figure 1, long-range fields arising from each one of the droplets attract molecules belonging to the second droplet, giving rise to *coalescence*. Coalescence is a very difficult interfacial phenomenon, which can, only, be fully described, in the molecular scale, related to interaction length of long-range forces.



(a) (b) (c)

Figure 1. Coalescence between two droplets.

Although very interesting from a physical point of view, droplet formation from a dropper is a very difficult problem, when we consider classical discrete methods of fluid mechanics. Droplet formation is pictured in Figure 2 (Adamson, 1990) showing a sequence of drawings based on high speed photographs.

From a macroscopic point of view droplet's shape time evolution is linked to the competition it is subjected between gravity action, viscosity of the droplet fluid and interfacial tension. In this way, interfacial forces hold the droplet until break-off, as droplet weight increases. Break-off starts with the development of a throat, which becomes thinner in time and from where droplet fluid is pulled downward and redistributed horizontally by viscous forces, giving an almost ellipsoidal shape to the falling droplet, with a major axis oriented along horizontal direction.

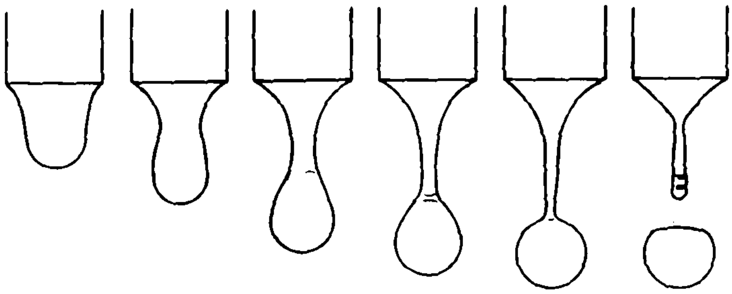


Figure 2 Droplet formation (Adamson, 1990)

The use of a large set of particles in random motion for representing the macroscopic behavior of fluids is not a new idea, dating from Bernoulli (1738), who tried to explain *elasticity* of gases considering them as a set of particles in random motion founding the kinetic theory of gases. Nevertheless, kinetic theory main development occurred in the second half of XIX century by Maxwell and Boltzmann. This was achieved by introducing *probability theory* in the study of N-body problem in classical Lagrangian mechanics.

In fact, no *general* solution exists for the N-body problem when N is larger than 2.

Considering a gas as a set of a *very large* number N of material points, with translational degrees of freedom, it is possible to use probability laws when considering

$$f(\mathbf{r}, \mathbf{c}, t)$$

as a probability density function for the number of particles with velocities between $\mathbf{c}$ and $\mathbf{c} + d\mathbf{c}$ found, at time t, inside an elementary volume $d\mathbf{r}$ of the physical space.

Considered as a continuous function, in the absence of external forces, the velocity distribution function $f(\mathbf{r}, \mathbf{c}, t)$ is modified by the streaming of particles and by collisions in **r, c** space. Its evolution is given by Boltzmann's equation:

$$\partial_t f + c_\alpha \partial_\alpha f = (\partial_t f)_{coll}$$

where ∂_t is a time derivative and ∂_α means a spatial derivative.

Boltzmann's equation has an H-theorem and an equilibrium solution, explaining irreversibility of macroscopic behavior as due to inter-particle collisions. In this way, collisions are considered to be the main mechanism responsible for dissipation phenomena in fluids.

Boltzmann's equation is a mesoscale level description. Macroscopic equations relating the main macroscopic properties can be obtained by integrating this equation over the velocities space.

In the early XX century, Chapman and Enskog, simultaneously, formally retrieved hydrodynamic transport equations from Boltzmann's equation, when Knudsen number $Kn \rightarrow 0$, by considering the first statistical moments of the velocities distribution function (Chapman and Cowling, 1970):

$$\partial_t(\rho) + \partial_\beta(\rho v_\beta) = 0$$
$$\rho\partial_t(v_\alpha) + \rho v_\beta \partial_\beta[v_\alpha] = \rho g_\alpha - \partial_\alpha(p) + \partial_\beta\{\mu[\partial_\beta(v_\alpha) + \partial_\alpha(v_\beta)]\} + \partial_\alpha\{\kappa[\partial_\beta(v_\beta)]\}$$

where $\mathbf{v}$ designates fluid velocity.

In the above equations: i) pressure p is directly related to mass density ρ by ideal gas law, ii) first, μ, and second, κ, viscosity coefficients are given in terms of the collision term in the Boltzmann's equation and related to ideal gas behavior.

These two above remarks are very important in the context of lattice gas development. In fact, this means that a) a large set of particles follow ideal gas law, when *long-range* interaction is not considered and b) hydrodynamic equations are *insensible* to the details of collision processes, which appear related, only, to the transport coefficients themselves, μ and κ.

Last remark, also, means that in despite of strong simplifications with respect to molecular behavior, microscopic models are capable of retrieving correct macroscopic behavior, when they are able to catch the main features at molecular scale.

Lattice gas models were, firstly, developed for single-phase flows based, mainly, on *cellular automata* and on *kinetic theory*. Two classes of models have been developed: Boolean and lattice Boltzmann models.

Boolean lattice gas automata models (LGA) are microscopic models based on particles, which dynamics try to mimic the main overall dynamics of a large set of molecules, preserving mass, momentum and, more recently, energy.

In Boolean models, a Boolean variable $n_i(X,T)$ is attributed to direction i of each site X of a discrete lattice, at time step T, indicating the presence (n_i=1) or absence (n_i=0) of a fluid particle, following an exclusion principle. For each time step, the dynamic evolution of the model is given in two steps. In the first step, designated as *collision step*, the state of site X is changed following collision rules conceived so as to preserve total mass and momentum of the site. In the second step, called *propagation step*, particles are propagated to the neighbor sites, in accordance with their direction at site X after collision step. The use of such models to study and simulate fluid dynamics was firstly introduced by Hardy, de Pazzis and Pomeau (1973), but it was only after 1986 that these models grew in increased importance due to the work of Frisch *et al.* (1986, 1987) These authors formally demonstrated that the dynamics of such models under certain conditions was described by the Navier-Stokes equations for incompressible flows, and could be used to simulate such flows. In fact, based on a square two-dimensional lattice, HPP model of Hardy *et al.* (1973): i) does not have isotropic fourth-order tensors, such as viscosity and ii) preserves spurious quantities, giving a non-physical behavior to the model.

Main contribution of Frish *et al.* (1986), was to demonstrate that hexagonal lattices have the necessary number of degrees of freedom to give isotropic fourth-order viscosity tensor and the elimination of spurious invariants. In three-dimensions, isotropy has been investigated by d'Humières et al. (1986), who introduced the three-dimensional projection of a four-dimensional lattice, the face-centered hypercubic lattice (FCHC) as the simplest lattice giving isotropy of fourth-order tensors. FCHC lattices have 24 degrees of freedom.

Although very suitable from a computational point of view, regarding parallelism and numerical stability, the use of Boolean variables has very serious overcomes that may be summarized by considering: i) non-physical terms in macroscopic equations due to exclusion principle, ii) high noise/signal level, produced by excessively drastic transitions and requiring spatial averages and iii) high transport coefficients.

For this reason, LGA models have been, gradually, replaced by Boltzmann, mesoscale models, in practical applications, although some work is, very recently, being undertaken, for reducing, or eliminating, these overcomes.

At mesoscale, lattice Boltzmann equation (LBE) resulted from the statistical averaging $N_i = \langle n_i \rangle$ of a large set of realizations of Boolean evolution equation, considering molecular-chaos hypothesis (McNamara and Zanetti, 1988). Fluctuations were drastically reduced in simulating N_i (**X**, T), but the remaining overcomes of Boolean models were preserved. Higuera and Jimenez (1989) proposed the use of a linearized collision term, further simplified by Qian et al. (1992) to give a single relaxation time model. In its present form, LBE is to be considered as a mesoscale relaxation equation, which main *collision term* does not follow Boolean transitions, but is written following some main fundamental principles, such as mass and momentum conservation and considering lattice symmetries.

In addition of enabling the description of physical process that require downscaling, lattice Boltzmann models present some computational features that can be forwarded to persuade CFD practitioner to adopt these models in simulating fluid flow: simplicity of the algorithm, easy of dealing with complicated geometric boundaries and high level of parallelism in the implementation.

This presentation is organized in the following manner. In Section 2, Boolean lattice gas models are discussed. Section 3 is devoted to Boltzmann mesoscale models, based on LBE relaxation equation. Section 4 present a Boolean model based on field mediators for studying the flow of miscible and immiscible fluids. Boltzmann's counterparts are presented in Sections 5 and 6.

2. Boolean Models for Monophasic Flows

2.1 Microscopic Dynamics

Consider a regular lattice, Figure 3, where each site X has b_m neighbors. A Boolean variable n_i (**X**, T) is assigned to site **X** to indicate the presence (n_i=1) or absence (n_i=0) of a particle in direction i at time T. Vector $\mathbf{c}_i$ indicates the unitary velocity vector pointing in direction i. A finite, at most b_r, number of, undistinguishable, rest particles is allowed to populate site **X**. Let $b = b_m + b_r$. Let S be the set of all possible states of a given lattice site.

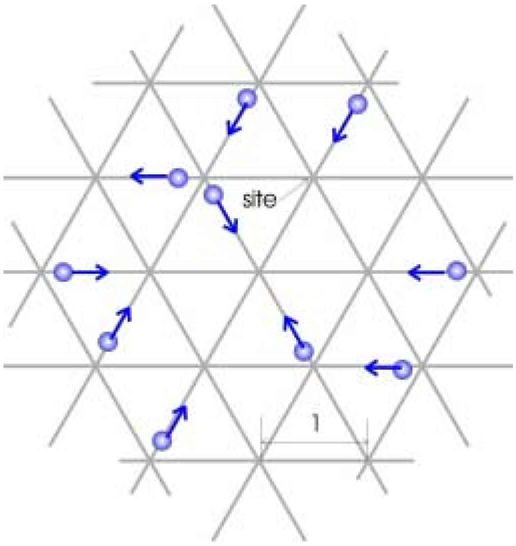


Figure 3 A two-dimensional hexagonal lattice.

A given state s of S can be represented by the array:

$$s=(s_{o1},..,s_{obr}, s_1, ...,s_{bm}) , \tag{2.1}$$

where the first b_r bits indicate rest particles and the following b_m bits indicate moving particles, distributed along the b_m lattice directions.

Microscopic evolution is described by the following equation:

$$n_i(\mathbf{X}+\mathbf{c}_i, T+1)=n_i(\mathbf{X},T) + \omega_i(n_{o1},.. n_{obr}, n_1, ..,n_{bm}) , \tag{2.2}$$

where ω_i: $(n_{o1},.. n_{obr}, n_1, ..,n_{bm})\rightarrow\{-1,0,1\}$ represents the collision operator which can take the values –1, 1 or 0, depending on the state $(n_{o1},.. n_{obr}, n_1, ..,n_{bm})$ of site **X**, before the collision.

Considering $\xi(\mathbf{X},T)$ to be a random number attributed to site **X** at time T and $\alpha_{\xi(\mathbf{X},T)}(s,s')$ to be the transition matrix changing state s to one of all possible post-collisional states s', in accordance with $\xi(\mathbf{X},T)$, collision term can be written as

$$\omega_i(n_{o1},..n_{obr},n_1,..,n_{bm}) = \sum_s \left[\sum_{s'} \alpha_{\xi(\mathbf{X},T)}(s,s')(s'_i - s_i) \prod_{j=1}^{b} \delta(n_{ji}, s_j) \right] . \tag{2.3}$$

2.2 Ensemble averages. Macroscopic behavior of LGA model.

Lattice-gas models have three description levels. In the more detailed level, $n_i(\mathbf{X},T)$ are described for every **X** and T. In general, this is too refined in the description of macroscopic phenomena. A less detailed description is given by furnishing the expected values $N_i=\langle n_i(\mathbf{X},T)\rangle$, obtained as *ensemble* averages over a large number of realizations. In the third level, only the first moments of N_i are furnished for each **X** and T. In fact, in the continuum limit, when Knudsen number is very small, it can be show that the first moments of N_i are related between themselves through a closed system of equations, i.e., the hydrodynamic equations.

Classically, in the framework of *hydrodynamics*, we try to solve this closed system of equations and obtain numerical values for pressure, density and velocity fields. In LGA conception, expected values N_i (**X**, T) result from several realizations of a given Boolean model. Macroscopic equations are, then, obtained from the first moments of N_i. It may be show that, under certain restrictions, these moments satisfy classical hydrodynamic equations (Frish *et al.*, 1986).

Distribution $N_i(\mathbf{X},T)$ is defined as the expected value of $n_i(\mathbf{X},T)$, over an *ensemble* of realizations, run using randomly chosen initial conditions and satisfies.

$$N_i(\mathbf{X}+\mathbf{c}_i, T+1)=N_i(\mathbf{X},T) + \Omega_i(N_o, N_1, ..,N_{bm}) , \tag{2.4}$$

which is Boltzmann equation for the lattice, in discrete form. Since rest particles are undistinguishable, $N_o(\mathbf{X},T)$ means the probability of finding rest particles on site **X** at time T, in several realizations. Taking molecular chaos hypothesis into account, the collision term can be written as:

$$\Omega_i(N_o, N_1, .., N_{bm}) =$$

$$< \omega_i(n_{o1},..n_{obr}, n_1,..,n_{bm}) >= \sum_s \left[\sum_{s'} A(s,s')(s'_i - s_i) \prod_{j=1}^{b} N_j^{s_j} (1-N_j)^{1-s_j} \right]. \tag{2.5}$$

In the above equation,

$$A(s, s') = <\alpha(s, s')>. \tag{2.6}$$

Using the semi-detailed balance condition,

$$\sum_s A(s,s') = 1 , \qquad \forall s' \quad , \tag{2.7}$$

it may be show that Eq. (2.4) has an H-Theorem and an equilibrium solution ,

$$N_i^o = \frac{1}{1+\exp(\alpha + \mathbf{q}.\mathbf{c}_i)} ., \tag{2.8}$$

which is a Fermi-Dirac distribution, as a consequence of *exclusion principle.*

Due to the discrete nature of the model, a linear low velocity approximation is used, written in terms of the density

$$\rho = \sum_{i=1}^{b_m} N_i + N_o b_r , \tag{2.9}$$

and the mean velocity

$$\mathbf{u} = \frac{1}{\rho} \sum_{i=1}^{b_m} N_i \mathbf{c}_i , \tag{2.10}$$

which can be used for finding constants α and q.

This equilibrium solution can be written as

$$N_i^o = f\left[1 + \frac{Db}{c^2 b_m} c_{i\alpha} u_\alpha + \frac{D^2}{2c^4}\frac{b^2}{b_m^2}\left(\frac{1-2f}{1-f}\right)\left(c_{i\alpha}c_{i\beta} - \frac{c^2}{D}(1-b_r)\delta_{\alpha\beta}\right)u_\alpha u_\beta\right] + O(u^3) , \tag{2.11}$$

for moving particles whereas for rest particles,

$$N_o^o = f\left[1 - \frac{D}{2c^2}\frac{b}{b_m}\left(\frac{1-2f}{1-f}\right)u^2\right] + O(u^4) . \tag{2.12}$$

In the above equations D is the Euclidean dimension of the lattice and f=ρ/b.

2.3 Lattice Gas Hydrodynamic Equations

Use of Chapman-Enskog method on the N_i evolution equation, Eq.(2.4), leads to lattice gas hydrodynamic equations, in the limit of low Knudsen number, Kn and low Mach number, M:

$$\partial_t(\rho) + \partial_\beta(\rho u_\beta) = 0 \tag{2.13}$$

$$\partial_t(\rho u_\alpha) + \partial_\beta\left[g(\rho)\rho u_\alpha u_\beta\right] = -\partial_\alpha(p(\rho,u^2)) + \nu\partial_\beta\left[\partial_\beta(\rho u_\alpha) + \partial_\alpha(\rho u_\beta)\right] + \eta\partial_\alpha\left[\partial_\beta(\rho u_\beta)\right] \quad (2.14)$$

where

$$g(\rho) = \frac{bD}{b_m(D+2)}\left(\frac{1-2f}{1-f}\right) \quad (2.15)$$

$$p(\rho,u^2) = c_s{}^2\rho - M^2\rho g(\rho)\left(1 + \frac{D}{2} - \frac{bD}{2b_m}\right) \quad (2.16)$$

and $c_s{}^2 = \dfrac{b_m c^2}{bD}$ is the square of LGA sound speed. The first and the second viscosity coefficients, respectively ν and η, are related to the eigenvalues of collision operator Ω. Equations (2.13) and (2.14) differ from Navier-Stokes hydrodynamic equations: i) by the inclusion of a $g(\rho)$ dependence in the inertial term, breaking Galilean invariance, ii) by a $O(M^2)$ additional term in the pressure equation, Eq. (2.16), considering U as a characteristic macroscopic speed, in lattice units and taking M as the Mach number, $M = U/c_s$, and iii) by the inclusion of density ρ *inside* the spatial derivatives in the viscous terms. In low Mach numbers limit (Rothman and Zaleski, 1997), the incompressibility condition

$$\nabla . \mathbf{u} = 0 \quad (2.17)$$

is recovered and the following momentum equations are found, in this limit :

$$\rho\partial_t(u_\alpha) + \rho g(\rho) u_\beta \partial_\alpha u_\beta = \partial_\alpha(p) + \rho\nu\partial_\beta\partial_\beta u_\alpha \, , \quad (2.18)$$

which are the correct Navier-Stokes equation, for incompressible flows, excepting by the inclusion of a $g(\rho)$ factor in the inertial term. Considering Eq. (2.15), it can be seen that $g(\rho)\to 1$ when the following two conditions are, simultaneously, satisfied.

a) Factor

b)

$$(1-2f)/(1-f) \to 1 \, , \quad (2.19)$$

related to the fact that lattice effects due to *exclusion principle* are reduced, when reducing lattice density.

b) Factor

$$\frac{bD}{b_m(D+2)} \to 1 \quad (2.20)$$

meaning that lattice effects due to the use of a *finite number of directions* are reduced, when increasing the number of lattice degrees of freedom, increasing $b_r = b - b_m$ for rest particles, for compensating factor $D/(D+2) < 1$ and/or working with three-dimensional lattice projections of higher D hyper-spaces.

The first alternative is limited by *noise increase* in the simulations, requiring the use of larger spatial averages. The second one is limited to computer resident-memory capacity, as memory requirements increase with $(b_r + 1)\, 2^{bm}$.

Lattice effects can be, also avoided by rescaling variables **u** and p (Rothman and Zaleski, 1997), which is a, presently, frequently used simulation *strategy*.

2.4 Energy Conservation in Boolean Models

In lattice-gas models, translation speed $c = |\mathbf{c}_i|$ is a constant and different planes with different energy levels must be introduced to take energy relaxation into account. A simple model that is, presently, under development* uses a $b_r + 2b_m$ bits variable to describe the state of a site

$$s = (s_{o1}, .. , s_{obr}, s_1, ..., s_{bm}, e_1, ..., e_{bm}) \quad (2.21)$$

* Facin, P.C., Doctoral Thesis, Mechanical Eng. Department, UFSC, Floranópolis (2001)

where s_i designate particles, distributed in the b_m lattice directions, s_{oi} designate rest particles and e_i designate the energy attribute $e_i = \{0,1\}$ of a lattice particle in direction i. Energy attribute e_i may be considered as an *internal energy* attribute in addition to translational energy.

In this way, mass, momentum and energy conservation require:

$$\sum_{i=1}^{b_m} n_i + \sum_{i=1}^{b_{br}} n_{oi} = \sum_{i=1}^{b_m} n'_i + \sum_{i=1}^{b_{br}} n'_{oi} \tag{2.22}$$

$$\sum_{i=1}^{b_m} n_i \mathbf{c}_i = \sum_{i=1}^{b_m} n'_i \mathbf{c}_i \tag{2.23}$$

$$\frac{1}{2}\sum_{i=1}^{b_m} n_i c_i^2 + \sum_{i=1}^{b_m} n_i e_i c_i^2 = \frac{1}{2}\sum_{i=1}^{b_m} n'_i c_i^2 + \sum_{i=1}^{b_m} n'_i e'_i c_i^2 \tag{2.24}$$

Attributing to internal modes the double of the corresponding translational energy is a convenient strategy since more simplified collision tables can be used, when performing the transition. In this way, when a moving particle loss its internal energy, it can put *two* rest particles into opposite lattice directions, without breaking momentum conservation.

Macroscopic variables can be written as

$$\rho = \sum_{i=1}^{b_m} n_i + \sum_{i=1}^{b_{br}} n_{oi} \tag{2.25}$$

$$\rho \mathbf{u} = \sum_{i=1}^{b_m} n_i \mathbf{c}_i \tag{2.26}$$

$$\rho e_{tr} = \frac{1}{2}\sum_{i=1}^{b_m} n_i (\mathbf{c}_i - \mathbf{u})^2 + \frac{1}{2}\sum_{i=1}^{b_r} n_{oi} u^2 \tag{2.27}$$

related to translational energy fluctuation and giving the translational temperature T_{tr}

$$\rho e_{int} = \sum_{i=1}^{bm} n_i e_i c_i^2 \tag{2.28}$$

related to internal energy. Remark that total energy e_T is given by

$$\rho e_T = \left(\frac{1}{2}\sum_{i=1}^{b_m} n_i (\mathbf{c}_i - \mathbf{u})^2 + \frac{1}{2}\sum_{i=1}^{b_r} n_{oi} u^2 \right) + \frac{1}{2}\rho u^2 + \sum_{i=1}^{b_m} n_i e_i c_i^2 \tag{2.29}$$

and that model enables energy relaxation between different energy modes.

Boundary conditions of first order are simulated by imposing a given configuration on boundary sites: fluid particles are reflected back with a probability

$$P = \frac{\sum_{i=1}^{b_m} e_{i,b}}{\sum_{i=1}^{b_m} n_{i,b}} \tag{2.30}$$

of changing their energy-level to 1. Reflected hot particles increase total energy of a given site in fluid domain, which is redistributed between its different energy modes in collisions. For every 1-P realizations hot particles that reach the boundary will loose their internal energy to boundary sites.

2.5 Sample application: prediction of intrinsic permeability of porous rocks (Santos *et al*., 2001.a)

At authors' knowledge, lattice-gas hydrodynamic models for flow through two and three-dimensional artificially constructed porous microstructures were described by Chen *et al.* (1991.b). Kohring (1991a-b), Kohring (1992), McCarthy (1994), and Gao and Sharma (1994) introduced lattice-gas models for studying the flow through channels, random array of solid obstacles and/or regular arrays of cylinders. Ginsbourg and Adler (1994), Genabeek and Rothman (1999) and Bernabe and Olson (2000) produced detailed results related to boundary location and to the influence of surface topography on the flow rate, for flow inside channels with rough surfaces.

Lattice-Boltzmann method was applied for the reconstructed three-dimensional porous microstructure by Ferreol and Rothman (1995) and by Singh and Mohanty (2000).

Boolean models are *lower level* models with respect to lattice Boltzmann. In present application, the use of Boolean models is proposed for the prediction of intrinsic permeability. In contrast with lattice-Boltzmann method, which is free from intrinsic noise, Boolean models present, nevertheless, very attractive advantages from a computational point of view, as simulations are performed with Boolean variables needing less resident memory capability and reducing running time. Taking into account that intrinsic permeability is a *global property, spatial averages* can be performed considering the whole pore space. In addition, ergodic hypothesis enables the use of unrestricted *time averages.* In this presentation it is shown that, as a consequence of spatial and time averages, reduction of intrinsic fluctuations of Boolean models, leads to the prediction of very stable values of intrinsic permeability, when simulation is performed considering a sufficiently great number of time steps.

2.5.1 Simulation scheme

Evolution equation, Eq. (2.2), is the basic algorithm used for simulating flow. At time t=0, lattice particles are randomly distributed on the lattice sites. For each pre-collision configuration s, post-collision configuration s' is randomly chosen between those states s' with the same mass and momentum. This is performed by using a transition table located in computer resident memory and constructed, previously to simulation, following the particular LGA model used. Model is based on a FCHC lattice with b=24 (d'Humières *et al*., 1986). In propagation step, each particle at direction i is propagated to the neighbor site $\mathbf{X}+\mathbf{c}_i$.

At solid boundaries, particles that reached boundary sites are bounced back, at the next time step. This is the *bounce-back* condition that is frequently used, avoiding flow slipping at the boundary, and assuring adherence condition, $\mathbf{u}=\mathbf{0}$.

In simulating incompressible flows, considering the state equation relating pressure to density, pressure gradients can, only, be promoted, associated to density gradients, which must, in turn, remain small. This problem is eschewed in conventional simulation by using Navier-Stokes *low Mach number, M, approximation*. Nevertheless, in LGA simulation, flow is the result of *billiard balls* collisions and incompressibility can only be assured by working with small $|\mathbf{u}|$ meaning small $M=u/c_s$,. It can be show that $c_s=0.7071$, in present Boolean model. In this way, in LGA simulation a pressure gradient can only be created associated to a density gradient, which must be small, assuring M<<1 and avoiding compressibility effects.

Three-dimensional representations of porous structure were obtained from two-dimensional sections by using Liang *et al.* (1998) reconstruction method. Reconstruction is based on a truncated gaussian stochastic simulation that preserves the first two moments of phase-function $Z(\mathbf{r})$, i.e., porosity ε and auto-covariance function $R_Z(u)$.

For simulating flow, present scheme uses periodic conditions and a *pumping zone* at the beginning of the domain (Figure 4). Periodic conditions assure that particles that escape from the end of lattice-domain are re-introduced at its beginning. In *pumping zone*, momentum is added to the particles, forcing them to the flow domain. In this way, model tries to mimic the real conditions related to a *real* hydraulic closed looping.

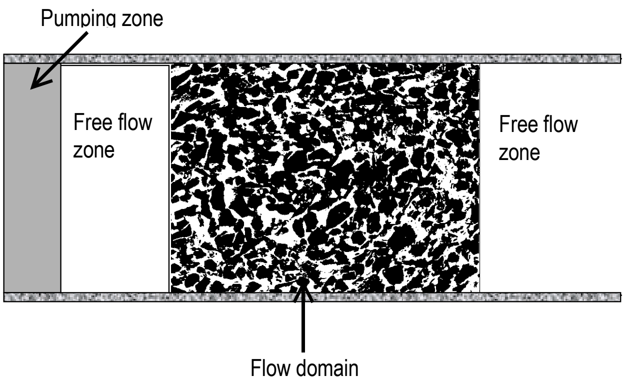
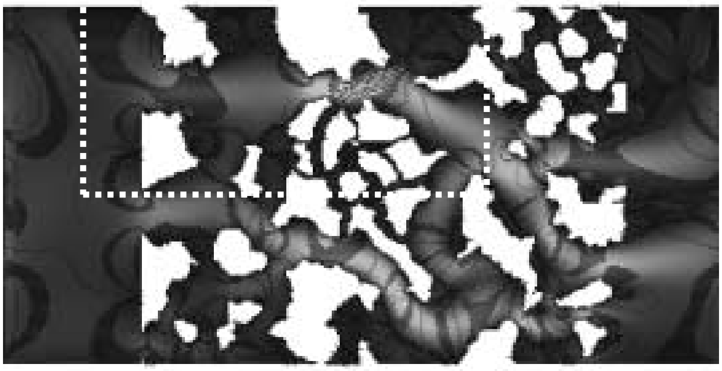


Figure 4 LGA simulation sheme.

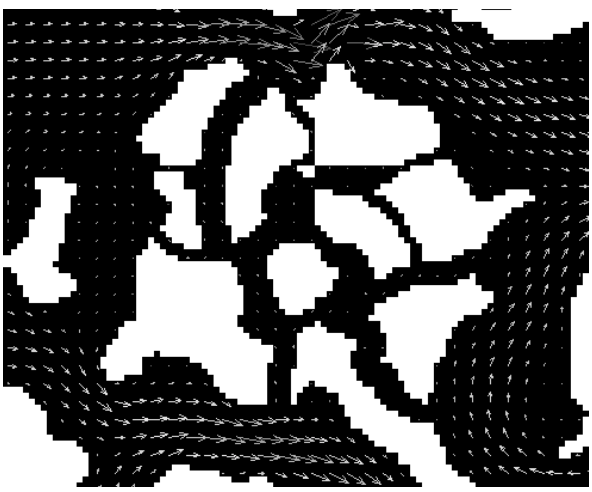
2.5.2 Simulation process, computer requirements

Figure 5 presents a LGA flow simulation through an artificially constructed , two-dimensional, porous structure, with two main pore-scales. Although pore-space is connected in both scales, fluid flows, preferentially, through the larger scale space, circulating around the porous grains and creating stagnant and/or vortex flows inside the lower pore-scale space.

Figure 6 presents a typical simulation result, showing permeability evolution for a 300^3 representation of a sandstone, with a reported experimental permeability of 69 mD. Simulation starts from zero-velocity initial conditions and is established after, around, 15000 time steps.



(a)



(b)

Figure 5. LGA flow simulation through an artificially constructed , two-dimensional, porous structure, with two main pore-scales: a) lighter gray is related to higher velocity modulus; b) zoom view showing circulation around the porous grains and stagnant and/or vortex flows inside the lower pore-scale space: velocity is represented by arrows

LGA model has intrinsic fluctuations due to Boolean occupation of a discrete lattice with a finite number of directions. In this way, mean flow rate was evaluated, at each time step, by considering all the sites located inside the porous domain and time fluctuations were reduced by performing time averages. Figure shows simulation results for time averages, using, respectively, 100, 200, 400, 800 and 1600 time steps.

Collision table takes 65Mb of computer resident memory, independently of 3D representation size. A 32 bits variable is used for describing the micro-state n_i (**X**,T) of each site. Process takes, around, 1Mb for storing the binary representation, 64 Mb for storing the micro-states, needing a total of 130 Mb for simulating a 200^3 representation. In this way, it was possible to run the most part of processes on ordinary Pentium processors. A 1 GHz processor takes about 2 μs to accomplish a single time-step, for each site. As collision and propagation steps are, only, performed on the sites located in porous phase, processing-time is, about, 3s for a 200^3 representation for a single time step and 12 h for 15000 time steps.

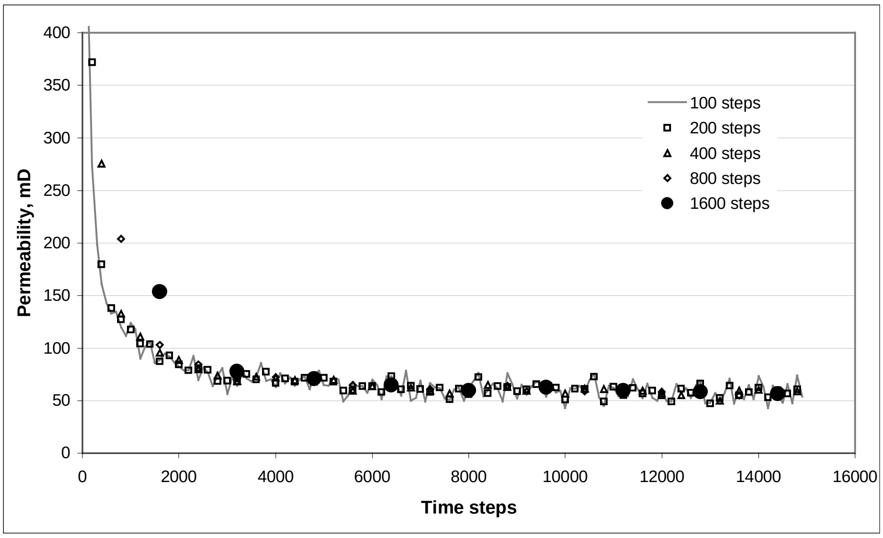


Figure 6. LGA permeability calculation for d18-connected 300^3 three-dimensional representation of a rock sample, using different time steps averaging. Experimental permeability value is 69 mD.

3. Mesoscale Models Based on LBE Relaxation Equation

Although very suitable from a computational point of view, the use of Boolean variables has very serious overcomes that may be summarized by considering: i) non-physical terms in macroscopic equations due to exclusion principle, ii) excessive noise, produced by excessively drastic transitions and requiring spatial averages and iii) high transport coefficients.

In fact, considering that populations in the lattice b directions follow a Bernoulli distribution, it may be shown that, for scalar quantities, fluctuations σ in Boolean simulation are given by (see, also Boghosian *et al.,*1997)

$$\sigma \sim \sqrt{\frac{1}{fmn}} \tag{3.1}$$

where f is the number of bits per site, m is the number of sites considered for spatial averages and n is the number of independent realizations.

Considering molecular-chaos hypothesis, mesoscale models start from the ensemble average of Boolean evolution equation, i.e., the lattice-Boltzmann equation (LBE) and proceeds actualizing distribution N_i at each **X**, T in two simulation steps:

collision step

$$N'_i(\mathbf{X},T) = N_i(\mathbf{X},T) + \Omega_i(N_o, N_1, ..,N_{bm}) \,, \tag{3.2}$$

propagation step

$$N_i(\mathbf{X}+\mathbf{c}_i, T+1) = N'_i(\mathbf{X},T) \tag{3.3}$$

for $i=0,1,...b_m$.

It is easy to see that, although drastically reducing fluctuations, mesoscale models have the same remaining overcomes of Boolean models when collision term Ω_i is taken as the ensemble average of Boolean transitions (Eq. 2.5). Particularly, model would predict a non-physical Fermi-Dirac distribution at equilibrium, due to exclusion principle (McNamara and Zanetti, 1988).

Nevertheless, avoiding its Boolean nature, a collision model can be written for Ω_i with the condition that it leads to a physically consistent equilibrium distribution. In this way, Higuera and Jimenez (1989) proposed a linear approximation to Ω_i

$$\Omega_i(N_o, N_1, ...N_{bm}) \sim \Omega_i(N_o^{eq}, N_1^{eq}, ...N_{bm}^{eq}) + \sum_{k=0}^{bm} \Lambda_{ik}(N_k^{eq} - N_k) \tag{3.4}$$

which reduces to BGK collision term (Bhatnagar *et al.*, 1954) when

$$\Lambda_{ik} = \frac{1}{\tau}\delta_{ik} \tag{3.5}$$

giving a *single* relaxation time model, Chen *et al.* (1991.c), Qian *et al.* (1992). Recently, He and Luo (1997), demonstrated that BGK relaxation equation

$$N_i(\mathbf{X}+\mathbf{c}_i, T+1) - N'_i(\mathbf{X},T) = \frac{N_i^{eq} - N_i}{\tau} \tag{3.6}$$

can be deduced from its well-known counterpart BGK equation from continuous kinetic theory, by using a discrete set of velocities.
In Eq. (3.6), a 2nd order velocity polynomial form is given to equilibrium distribution, which coefficients are obtained by considered lattice isotropy and by requiring

$$\rho = \sum_{i=1}^{b_m} N_i^{eq} + N_o^{eq} b_r \tag{3.7}$$

$$\mathbf{u} = \frac{1}{\rho}\sum_{i=1}^{b_m} N_i^{eq}\mathbf{c}_i \tag{3.8}$$

$$\Pi_{\alpha\beta} = \sum_{i=1}^{b_m} N_i^{eq} c_{i\alpha} c_{i\beta} = c_s^2 \rho \delta_{\alpha\beta} + \rho u_\alpha u_\beta \tag{3.9}$$

Considering these requirements, it may be shown that

$$N_i^{eq} = \frac{\rho}{b} + \frac{D}{b_m c^2} c_{i\alpha}\rho u_\alpha + \frac{D(D+2)}{2b_m c^4} c_{i\alpha} c_{i\beta}\rho u_\alpha u_\beta - \frac{D}{2b_m c^2}\rho u^2 \tag{3.10}$$

$$N_o^{eq} = \frac{\rho}{b} + \frac{1}{b_r c^2}\rho u^2 \tag{3.11}$$

A Chapman-Enskog analysis for the model above described shows that correct Navier-Sokes equations are retrieved for non-compressible flows. In fact, *non-physical* compressible effects appear, proportional to M^2, in momentum equation for compressible flows.

3.1 Thermodynamic consistent mesoscale models

To give thermodynamic consistency, lattice models must consider energy relaxation. A pioneer work on a 2D thermal lattice BGK model was published by Alexander *et al.* (1993), generalized by McNamara *et al.* (1993) to include different relaxation times for the stress and energy. Chen *et al.* (1994) used a 4th-order velocity expansion for the equilibrium distribution, up-grading the lattice symmetry to ensure isotropy for the sixth-rank velocity-moment tensor, using a single relaxation time model,

$$N_i(\mathbf{X}+\mathbf{c}_{ik}, T+1) - N_{ik}(\mathbf{X},T) = \frac{N_{ik}^{eq} - N_{ik}}{\tau} \tag{3.12}$$

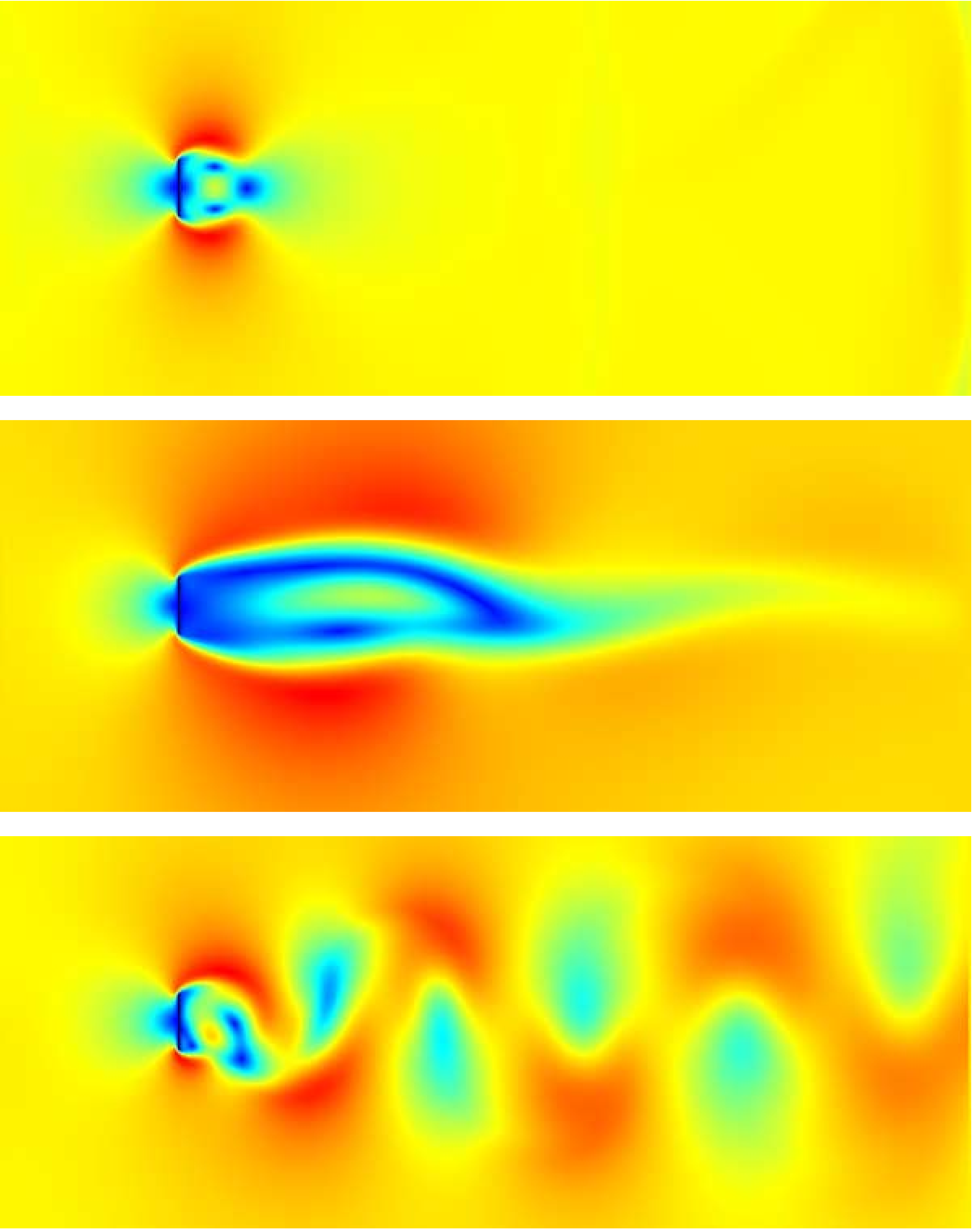
where k designates an energy level for velocity vector $\mathbf{c}_{ik}$, giving the modulus of $\mathbf{c}_{ik}$. Equilibrium distribution was written as

$$\begin{aligned} N_{ik}^{eq} = {} & A_k + M_k c_{ik,\alpha}u_\alpha + G_k u^2 + J_k\left(c_{ik,\alpha}u_\alpha\right)^2 + Q_k\left(c_{ik,\alpha}u_\alpha\right)u^2 + H_k\left(c_{ik,\alpha}u_\alpha\right)^3 \\ & + R_k\left(c_{ik,\alpha}u_\alpha\right)^2 u^2 + S_k u^4 \end{aligned} \tag{3.13}$$

Chen and co-workers model retrieves the correct thermo-hydrodynamic equations for compressible flows, with the only, visible limitation in the Prandtl number due to the use of a *single* relaxation time.

3.2 Sample application: von Karman vortex

Figure 7 presents simulation results for the two-dimensional flow past a rectangular plate. Simulation domain is 400X600, although only a part of simulation domain is shown in the figure. Simulation starts by supposing fluid at rest at t=0 and by imposing a constant velocity to the right at all four boundaries. Figure shows the formation of downstream von Karman vortices, from an asymmetry-generated instability source, degenerating the flow field to periodic waves, behind the plate. Reynolds number was Re=92.



(a)

(b)

(c)

Figure 7 Flow against a rectangular plate showing the formation of von-Karman vortices. Simulation domain is 400X600. Results are related to the first (a): 800, (b): 10000 and (c): 20000 simulation steps. Re=92.

4. A Boolean Model for Immiscible Fluids Based on Field Mediators

Long-range attraction between particles of the same kind promotes particles separation, being responsible for interfacial tension, which acts as a potential barrier at the interface between fluids. Boolean models for simulating immiscible fluids flow were, firstly, proposed by Rothman and Keller (1988). In this model, long range attraction between particles of the same kind was modeled by modifying the collision step, introducing an additional *separation* step between particles of different kinds, based on the information of the populations at first neighbors of site X, at step T, p_j (X+c_i, T), i=1,..., b_m, j=0_1,..,0_{br}, 1,..., b_m, where p designates the kind of particle found at direction j of site X +ci and 0_1,..,0_{br} is used for bit occupation when model allows b_r rest particles. Output site configuration is decided after a maximization step for the color flux in accordance with a color gradient at site **X**.

Rothman and Keller's model is computer expensive when processing time needs are considered. Chen and co-workers (Chen *et al.*, 1991.a) proposed a two-bits local model, where the time expensive neighbors survey of Rothman and Keller is avoided, by introducing colored holes. Colored holes are null-mass particles representing the *memory* of the kind of a given particle, and moves in the same direction that particle moved before collision. The state of a given site is represented by a two-bit Boolean variable (f_i (**X**,T), n_i (**X**,T)), where f_i=1 designates red and f_i=0 represents blue, n_i=1 represents a particle and n_i=0 represents a hole. In this way, in separation step, red (blue) particles are deviated to the direction from where red (blue) holes were originated, simulating *long range attraction*, by using, only, *local* rules. This is achieved by maximizing color flux $\mathbf{q} = \sum_i (2f_i - 1) n_i \mathbf{c}_i$ in accordance with hole flux $\mathbf{q}_H = \sum_i (2f_i - 1)(1 - n_i)\mathbf{c}_i$, at a given site **X**.

Field mediators are null-mass particles moving with light speed, introduced in electromagnetic theory for the quantum description of long-range fields. They were firstly introduced in LGA theory by Appert *et al.* (1993) when modeling phase transition. In this presentation, we introduce *field mediators* for simulating the flow of fluid mixtures presenting arbitrary miscibility, with the following, distinguishing, main features: i) like in Chen's model, present *field mediators* move with the greatest lattice speed, c=1, enabling the use of *local* rules in separation step, ii) emission and interference of present *field* mediators follow distinct rules when compared with Chen's *memory* mediators (holes), enabling to *avoid* Chen's optimization step, iii) field *strength* and *interaction* distance is simulated by introducing distinct emission, P_e, and extinction, P_a, probabilities, for field mediators, allowing to *control* interfacial tension and transition thickness and the degree of mixing between different fluids. These features were achieved by introducing a four-bit Boolean model, described in the next section.

4.1 Model (Santos *et al.*, 2001.b.c)

The state of a given site **X** at time T is given by a four bits Boolean variable (r_i(**X**, T), b_i(**X**, T), m^r_i (**X**, T), m^b_i(**X**, T)) where r, b, m^r and m^b designate, respectively, r particles, b particles, r mediators and b mediators. Model allows simultaneous r_i and b_i bit occupation, but exclusion principle is maintained between particles of the same kind. Particles of kind r are considered to have the same, unitary, mass.

Microdynamics has the following steps:

i) *Collision*. Collisions are responsible for mixing particles of different kinds, in the transition region, being related to binary species diffusion coefficient D_{rb}. Microdynamics equation relating post-collision Boolean variable p_i' to p_i can be written as:

$$\begin{aligned} p_i'(X,T) - p_i(X,T) &= \omega_i^p(r_{01},\ldots,r_{0b_r},r_1,\ldots,r_{b_m},b_{01},\ldots,b_{0b_r},b_1,\ldots,b_{b_m}) \\ &\equiv \omega_i^p(r^*(X,T),b^*(X,T)) \end{aligned} \tag{4.1}$$

where $p^* = (p_{oo},\ldots,p_{ob_r},p_1,\ldots,p_{b_m})$ designates a pre-collision p particles configuration of site X, at time T, oo,...,ob$_r$ is related to Boolean occupation for b_r allowable rest particles in a lattice with b_m directions. A Boolean variable with 2b bits, b_t =b_r + b_m, is used to designate an arbitrary particle state $s = (s^r{}_{01},\ldots,s^r{}_{0b_r},s^r{}_1,\ldots,s^r{}_{b_m},s^b{}_{01},\ldots,s^b{}_{0b_r},s^b{}_1,\ldots,s^b{}_{b_m})$ of the lattice model in a $[(b_r+1)\,2^{b_m}]^2$ dimensional B_r X B_p space (rest particles are undistinguishable). Collision operator,

$$\begin{aligned} \omega_i^p : B_r X B_b &\rightarrow \{-1,0,1\} \\ (r^*,b^*) &\rightarrow \omega_i^p(r^*,b^*) \end{aligned} \qquad p= r,b \tag{4.2}$$

$\forall$ i = oo,...,ob$_r$,1...,b$_m$, maps a 2^{2b} dimensional space on the set {-1,0,1}, respectively, *eliminating*, *leaving unaltered*, or *adding* a particle of kind p to direction i. Collision term can be written as

$$\omega_i^p(r^*,b^*) = \sum_{s,s'} \alpha_\xi(s,s')(s_i^{p'} - s_i^p)\prod_j r_j^{s_j^r}(1-r_j)^{1-s_j^r} b_j^{s_j^b}(1-b_j)^{1-s_j^b} \tag{4.3}$$

where $\alpha_\xi(s,s')$ is the transition matrix, $\xi = \xi(X,T)$ is a random variable attributed to site X, at time T. Transition matrix must assure , mass and momentum conservation:

$$\sum_i \omega_i^p = 0 \quad p = r, b \tag{4.4}$$

$$\sum_i c_i(\omega_i^r + \omega_i^b) = 0 \tag{4.5}$$

in collisions. In addition, let A(s, s') be the ensemble average of $A(s,s') = \langle \alpha_\xi(s,s') \rangle$. As usually, considering the set Ξ of the all possible ξ random values, transition matrices α_ξ, $\xi \in \Xi$, must be written so as to satisfy the conservation of probability and semi-detailed balance condition:

$$\sum_{s'} A(s,s') = \sum_s A(s,s') = 1 \tag{4.6}$$

as sufficient conditions for satisfying H-Theorem in describing irreversibility of diffusion processes.

ii) *Interference with field mediators*. In this step, particles of kind p, at site X, are subjected to long-range attraction from particles of the same kind. In present model, this is simulated locally, by inverting the momentum of each p-particle when a) it finds a p-mediator in the same direction and b) opposite direction is free from p-particles. Defining $n_p = \sum_i p_i$ p=r,b and n= n_r+n_b, this step is, only, performed when $0<(n_p)/n<1$, assuring a null-effect of long-range fields and preserving single fluid state equation inside each phase. Interference step can, thus be written as,

$$\begin{aligned} p_i''(X,T) = {} & p_i'(X,T)m_i^p(X,T)\big(1-p_{-i}'(X,T)\big) \\ & + p_{-i}'(X,T)m_{-i}^p(X,T)\big(1-p_i'(X,T)\big) \end{aligned} \quad \text{when } 0<n_p/n<1 \tag{4.7}$$

$$p_i''(X,T) = p_i'(X,T) \quad \text{otherwise} \tag{4.8}$$

where i=1,...,b_m and –i = i+b_m (mod b_m).

iii) *Emission of field mediators*. Considering an elementary volume ϑ located inside a mixture of two real gases, ϑ acts an attractive center for p molecules when n_p/n is above some critical value $(n_p/n)^*$, with a potential strength that depends on the kind of r-r, b-b and, consequently, r-b, interactions. In present LGA model, site X will be a source of p mediators when $(n_p/n)>(n_p/n)^*$, with a given emission probability P_e that depends on particle-p concentration n_p/n on site X, at time T. Emission probability is, thus, related to potential strength in the transition region, giving the interfacial tension σ_{rb}. When P_e=0, independently of n_p/n, fluids r and b will mix without long-range field restriction. On the other extreme, P_e=1, $\forall$ n_p/n, represents a mixture of two, ideally, immiscible fluids.

Emission step can be written,

$$m_i^{p'} = \begin{cases} 1 \text{ when } n_p/n > (n_p/n)^* \text{ and } P \le P_e \\ m_i^p \quad \text{otherwise} \end{cases} \tag{4.9}$$

where P is a random variable, $0 \le P \le 1$, attributed to site X, at time T. In this way, b mediators found at an attractive site for r particles are preserved, during emission step, allowing the site to propagate this information to its neighboring sites, in propagation step

iv) *Extinction of field mediators*. In addition to field strength, interaction length is an important parameter, contributing to transition layer thickness. In present model, interaction length is related to an extinction probability P_a. Thus, for a field mediator m^p (X,T) to be annihilated two conditions are imposed: a) n_p(X,T)=0 and b) P(X,T)$\le P_a$, $0 \le P \le 1$. These conditions assure that a field mediator p will be never destroyed in the transition region although r-mediators will be found inside b-phase, trying to rescue r-particles moved to b-phase by collisions. After extinction step:

$$m_i^{p''} = \begin{cases} m_i^{p'} \text{ when } n_p = 0 \text{ and } P \le P_a \\ 0 \text{ otherwise} \end{cases} \tag{4.10}$$

v) *Propagation*. In propagation, particles and mediators are propagated to next neighbors, in the same manner as in conventional LGA models:

$$p_i(X + c_i, T + 1) = p_i^{''}(X, T) \qquad p = r, b \tag{4.11}$$

$$m_i^p(X + c_i, T + 1) = m_i^{p''}(X, T) \qquad p = r, b \tag{4.12}$$

vi) *Boundary conditions*. Wetting/non-wetting properties of a pair of fluids with respect to solid surfaces are a macroscopic result of differential, long-range attraction between solid and fluid molecules. At equilibrium, this preferential attraction can be summarized by the formation of a well defined contact angle, θ, between fluid interface and the solid wall, which depends on the pair of fluids and on the solid surface. In present model, preferential attraction of solid wall, with respect to a given fluid p, is simulated by reflecting back p-mediators at boundary sites X_b, with a given probability P_r, related to θ. Non-wetting fluid mediators at boundary sites are not reflected, being annihilated at these sites. This condition may be written as,

$$m_i^p(X_b, T + 1) = \begin{cases} m_{-i}^{p''}(X_b, T) \text{ when } P(X_b, T) \le P_r \\ 0 \qquad \text{otherwise} \end{cases}$$

$\forall i$, pointing outward the solid surface, when p is the wetting fluid with respect to solid surface and where P, $0 \le P \le 1$, is a random variable attributed to site X_b, at time T.

vii) *External forces: forcing step.* Forcing step is performed before collision step, above described. Labeling by k the lattice direction parallel to external field direction, external forces $\mathbf{g}_p$ are simulated by reversing the momentum of particles r and b, located at direction -k opposed to $\mathbf{g}_p$, when direction, k, is free from particles of the same kind. Probability $P_{g,p}$ to this reversion, represents the force strength g_p on component p. Microdynamic equation describing *forcing step* can be written as:

$$p_k^{'}(X, T) = \begin{cases} p_{-k}(X, T)(1 - p_k(X, T)) \text{ when } P(X, T) \le P_{g,p} \\ p_k \qquad \text{otherwise} \end{cases} \qquad p=r,b \tag{4.12}$$

where P(X,T) is a random variable attributed to site X, at time T, $0 \le P \le 1$.

At equilibrium, using ergodic hypothesis when considering the whole lattice domain, the mean effect of forcing step on p-particles can be calculated by,

$$\mathbf{f}_p = 2P_{g,p} \frac{\langle n_p \rangle}{b_t} \left(1 - \frac{\langle n_p \rangle}{b_t} \right) \mathbf{c}_k \,, \tag{4.13}$$

which is the force, related to the momentum the $\langle n_p \rangle$ particles per site are expected to gain in direction k during a lattice time-step.

4.2 Sample application: Droplet formation under the action of gravity (Santos *et al*, 2001.c)

Although very interesting from a physical point of view, droplet formation from a dropper is a very difficult problem, when we consider classical discrete methods of fluid mechanics. Droplet formation is pictured in Figure 2 (Adamson, 1990) showing a sequence of drawings based on high speed photographs.

From a macroscopic point of view droplet's shape time evolution is linked to the competition it is subjected between gravity action, viscosity of the droplet fluid and interfacial tension. In this way, interfacial forces hold the droplet until break-off, as droplet weight increases. Break-off starts with the development of a throat, which becomes thinner in time and from where droplet fluid is pulled downward against the droplet and redistributed horizontally by viscous forces, giving an almost ellipsoidal shape to the falling droplet, with a major axis oriented along horizontal direction.

From a microscopic point of view, during the first moments of droplet fall, r-particles at droplet surface are subjected to long-range forces from r-phase inside the dropper, maintaining the integrity of r-phase in despite of gravity action. Droplet break-off starts when combined action of gravity and downward long range attraction from the created droplet increases with respect to upward long range attraction from r-phase inside the dropper, giving raise to the formation of droplet throat. During and after break-off, r-particles in the throat are pulled against the droplet, where these particles are redistributed inside the droplet by *r-r collisions* (related to the viscosity of droplet fluid).

Figure 8 shows a sequence of simulation results using present field-mediators model. Comparison of Figures 2 and 8 shows a very good qualitative agreement between simulation and experimental results.

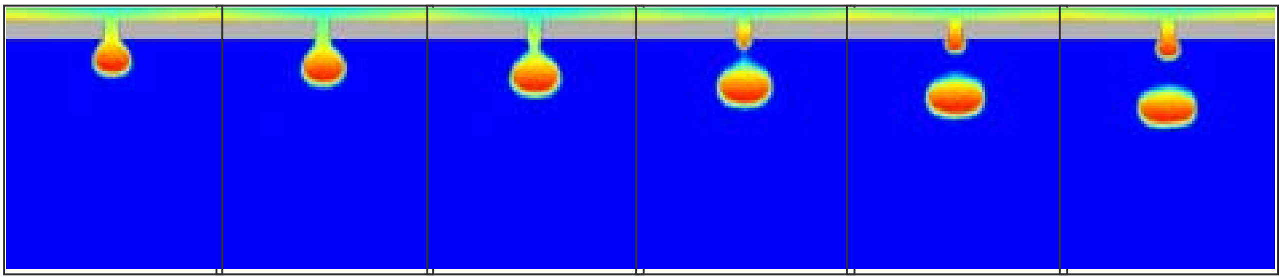


Figure 8 Simulation of droplet formation under gravity action.

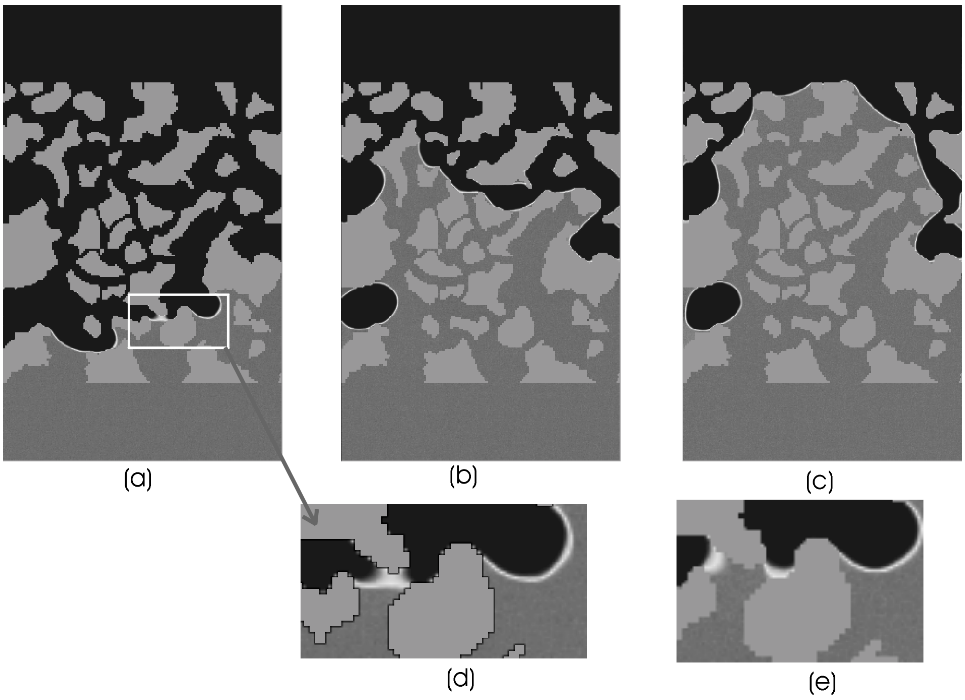


Figure 9 Simulation of capillary invasion of a wetting fluid, with a 0^o contact angle into a two-dimensional porous structure. Figure shows time averages for wetting fluid density taken for each 500 time steps (a) 45500 time steps, (b) 164000 time steps and (c) 240000 time steps. (d)-(e) Details of film break-up.

Figure 9 shows simulation results for a wetting fluid invasion process into a two-dimensional porous structure, for 45500, 164000 and 240000 time steps, respectively. Invasion proceeds by capillary forces, only. Pressure of the wetting and non-wetting fluids at the two chambers are kept the same and constant, during the process. Wetting fluid invades the capillaries, in contact with it, by piston displacement and reaches solid boundaries, which do not present any physical contact with it, by the action of long-range forces between solid surfaces and the fluid, giving rise to break-up mechanisms (see details d-e in the figure). Break-up of liquid films is an important physical mechanism, during wetting-fluid invasion and has been, intensively, studied during the last decades, although the most part of related works deal with mechanical equilibrium stability in simple cylindrical tubes (Everett and Haynes, 1972, Gauglitz and Radke, 1988). Break-up is the main factor producing non-wetting fluid trapping (Figure 9.b).

5. A Boltzmann Model for Miscible Fluids

At a given site, in a mixture of two fluids r and b, transition term can be written as a sum of two relaxation terms. Collisions between particles of the kind r, try to impose equilibrium distribution $R^o_{i,rr} = R^0_i(\rho_r, \mathbf{u}_r)$, with a relaxation time τ_r, specific to kind r particles and related to r-fluid viscosity. Collisions between particles of kind b, try to impose a *different* equilibrium distribution, with a relaxation time τ_b. Collisions between particles r and b try to impose a common distribution to both kinds of particles $R^o_{i,rb} = R^o_i(\rho_r, \mathbf{u})$ with a relaxation time τ_{rb}, related to binary species diffusion coefficient. In this way, collision model can be thought as a sequence of two relaxation process $R_i \to R^o_{i,rr}$ and $R^o_{i,rr} \to R^o_{i,rb}$

$$\Omega^r_i = \frac{R_i - R^o_{i,rr}}{\tau_r} + \frac{R^o_{i,rr} - R^o_{i,rb}}{\tau_{rb}} \tag{5.1}$$

for i=0,1,...b_m.

Equilibrium distribution for pure fluid r is, usually, required to satisfy:

i) mass conservation of r and b particles

$$\sum_{i=1}^{b_m} R_i + R_o b_r = \rho_r \tag{5.2}$$

ii) momentum conservation

$$\sum_{i=1}^{b_m} R_i \mathbf{c}_i = \rho_r \mathbf{u}_r \tag{5.3}$$

iii) preservation of momentum flux at first order

$$\Pi^r{}_{\alpha\beta} = \rho_r u_{r\alpha} u_{r\beta} + \frac{b_m c^2}{bD} \rho_r \tag{5.4}$$

In this way, equilibrium distribution for r-particles, considering r-r collisions, can be written as

$$R^o_i = f_r \left[1 + \frac{bDc_{i\alpha} u_{r\alpha}}{b_m c^2} + \frac{D(D+2)}{2c^4} \frac{b}{b_m} \left(c_{i\alpha} c_{i\beta} - \frac{c^2}{D} \delta_{\alpha\beta} \right) u_{r\alpha} u_{r\beta} + \frac{b}{c^2 b_m} u_r{}^2 \right] \tag{5.5}$$

$$R^o_o = f_r b_r - f_r \frac{b}{c^2 b_m} u_r{}^2 \tag{5.6}$$

and, similarly, for b particles.

Collisions of r and b particles lead to equilibrium distributions $R^o_{i,rb}$ and $B^o_{i,rb}$. These distributions are required to satisfy:

i) mass conservation of r and b particles

$$\sum_{i=1}^{b_m} R_{i,rb} + R_{o,rb} b_r = \rho_r \qquad \sum_{i=1}^{b_m} B_{i,rb} + B_{o,rb} b_r = \rho_b \tag{5.7}$$

ii) momentum conservation

$$\sum_{i=1}^{b_m} R_i \mathbf{c}_i + \sum_{i=1}^{b_m} B_i \mathbf{c}_i = \rho \mathbf{u} \tag{5.8}$$

iii) preservation of momentum flux at first order

$$\Pi_{\alpha\beta} = \rho u_\alpha u_\beta + \frac{b_m c^2}{bD} \rho \tag{5.9}$$

It is a simple matter to verify that

$$R^o_{i,rb} = f_r \left[1 + \frac{bDc_{i\alpha} u_\alpha}{b_m c^2} + \frac{D(D+2)}{2c^4} \frac{b}{b_m} \left(c_{i\alpha} c_{i\beta} - \frac{c^2}{D} \delta_{\alpha\beta} \right) u_\alpha u_\beta + \frac{b}{c^2 b_m} u^2 \right] \tag{5.10}$$

$$R^o_{o,rb} = f_r b_r - f_r \frac{b}{c^2 b_m} u^2 \tag{5.11}$$

and, similarly for b-particles.

6. A Boltzmann Model Based on Mediators for Immiscible Fluids

When fluids are immiscible, long-range attraction between particles of the same kind, will try to separate the two fluids. In this way, in addition to r-r and r-b collisions in the transition layer, there will be a separation effect due to long-range fields.

In its simplest form M^r_i ($\mathbf{X}$,T) can be written, in emission step, as proportional to $\rho_r(\mathbf{X},T)$. This information is propagated to next neighbors sites, where interference is to be produced. The presence of a long-range field from r-particles in a given site **X** is related to a non-null value of mediators distribution $M^r_i(\mathbf{X},T)$ and its action will try to move red particles to the direction from where mediators came

$$\mathbf{v}_r = -\sum_{i=1}^{b_m} M^r_i \mathbf{c}_i \tag{6.1}$$

i.e, to the direction r- particles are dominant.

Similarly, b-mediators will impose

$$\mathbf{v}_b = -\sum_{i=1}^{b_m} M^b_i \mathbf{c}_i \tag{6.2}$$

Defining a separation velocity $\mathbf{v}_s = f_\sigma(\mathbf{v}_r - \mathbf{v}_b)$, where f_σ is to be related to the interfacial tension between the fluids, mediators action will try to impose an equilibrium distribution to r-particles given by

$$R^o_{i,\sigma} = f_r \left[\begin{array}{l} 1 + \dfrac{bDc_{i\alpha}\left(\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha \right)}{b_m c^2} + \dfrac{D(D+2)}{2c^4} \dfrac{b}{b_m} \left(c_{i\alpha} c_{i\beta} - \dfrac{c^2}{D} \delta_{\alpha\beta} \right) \left(\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha \right) \left(\dfrac{\rho_b}{\rho} v_{s\beta} + u_\beta \right) \\ + \dfrac{b}{c^2 b_m} \left(\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha \right)^2 \end{array} \right] \tag{6.3}$$

$$R^{o}_{o,\sigma b} = f_r b_r - f_r \frac{b}{c^2 b_m}\left(\frac{\rho_b}{\rho} v_{s\alpha} + u_\alpha\right)^2 \tag{6.4}$$

and to b-particles

$$R^{o}_{i,\sigma} = f_r \left[\begin{array}{l} 1 + \dfrac{bDc_{i\alpha}\left(-\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha\right)}{b_m c^2} + \dfrac{D(D+2)}{2c^4}\dfrac{b}{b_m}\left(c_{i\alpha}c_{i\beta} - \dfrac{c^2}{D}\delta_{\alpha\beta}\right)\left(-\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha\right)\left(-\dfrac{\rho_b}{\rho} v_{s\beta} + u_\beta\right) \\ + \dfrac{b}{c^2 b_m}\left(-\dfrac{\rho_b}{\rho} v_{s\alpha} + u_\alpha\right)^2 \end{array} \right] \tag{6.5}$$

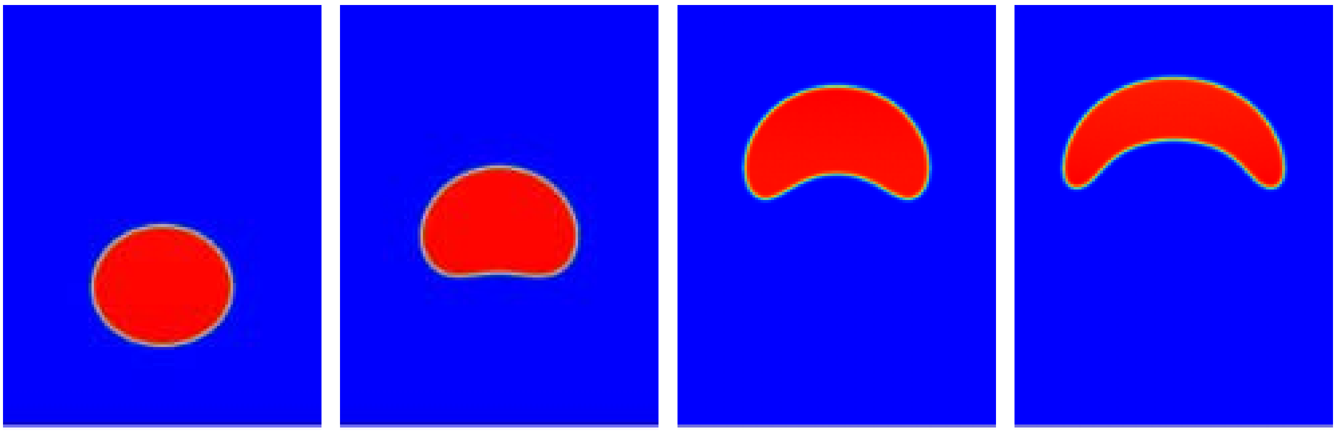
$$R^{o}_{o,\sigma b} = f_r b_r - f_r \frac{b}{c^2 b_m}\left(-\frac{\rho_b}{\rho} v_{s\alpha} + u_\alpha\right)^2 \tag{6.6}$$

assuring that, at this state, r and b particles will leave the site without breaking momentum conservation:

$$\sum_{i=1}^{b_m}\left(R^{o}_{i,\sigma} + B^{o}_{i,o}\right)\mathbf{c}_i = \rho\mathbf{u} \tag{6.7}$$

6.1 Sample application: bubble ascension dynamics, under gravity action

Figure-10 shows results of a two-dimensional simulation for bubble ascension against gravity. Simulation domain is 400X800. and bubble is considered as a circle at t=0, with an initial diameter of 100 lattice-units. Simulation uses a 2D FCHC lattice. Very interesting dynamical effects produce bubble deformation during ascension.



(a): t=0 (b): 2000 time steps (c): 3000 time steps (e): 4000 time steps

Figure 10. Bubble ascension against gravity. 400 x 800. Initial bubble diameter corresponds to 100 lattice units.

7. Conclusions

Lattice gas automata concepts appear to be very suitable for explaining *complex* macroscopic effects, based on *simple* models of fluid behavior at molecular level.

In this paper, Boolean model and Boltzmann models were presented for simulating single-phase flows and the flow of immiscible fluids. Field mediators were introduced for representing the action of long-range fields, but with an interference step described by local rules. Emission and extinction probabilities enable to *control* interfacial tension and transition thickness.

Considering their inherent simplicity, simulation results, apparently, confirm the adequacy of presented models to study fluid flow and physical phenomena related to fluid flow that require lower-level description scales.

References

A. Adamson, Physical Chemistry of Surfaces (5th Ed.), John Wiley & Sons (1990).

Alexander, F. J. Chen S., Grunau, D. W. (1993) "Lattice Boltzmann thermohydrodynamics" *Phys. Rev.* B 48, 634.

Appert, C. and Zaleski, S. (1990). "A lattice gas with phase transition" Phys. Rev. Lett., **64**, 1-4

Bernabe, Y., Olson, J.F. (2000). "The hydraulic conductance of a capillary with a sinusoidally varying cross-section" Geo. Res. Leters **27** (2), 245-248.

Boghosian, B.M., Yepez, J., Alexander, F.J., Margolus, N.H. (1997), " Integer lattice gases", Phys Rev E 55(4) 4137-4147

Chapman, S., Cowling, T.G., The Mathematical Theory of Non-Uniform Gases (3rd Ed.), Cambridge University Press (1970)

Chen, S., Doolen G. D., Eggert, K., Grunau, D. and Loh, E. Y. (1991.a). "Local lattice-gas model for immiscible fluids" *Physical Review A* **43,** 7053-7056.

Chen, S., Diemer, K., Doolen, D., Eggert, K., Gutman, S., Travis, B.J. (1991.b). "Lattice gas automata for flow through porous media" Physica D **47**(1-2), 72-84.

Chen, S., Wang, Z., Shan, X., Doolen, G. (1991.c). "Lattice Boltzmann computational fluid dynamics in three-dimensions", J. Stat. Phys, **68**, 379-400.

Chen, Y., Ohash, H., Akyama, M. (1994). "Thermal lattice Bhatnagar-Gross-Krook model without nonlinear deviation in macrodynamic equations", Phys. Rev. E, 50, 2776-2783.

Everett, D. H., Haynes, J.M. (1972) "Model studies of capillary condensation I. Cylindrical pore model with zero contact angle", J. Coll. Int. Science, 38, 125-137.

Ferreol, B., Rothman, D.H. (1995) "Lattice-Boltzmann simulations of flow through Fontainebleau sandstone". Transp. Porous Media 20(1-2), 3-20.

Frisch, U., Hasslacher B., Pomeau Y. (1986). "Lattice-Gas Automata for the Navier-Stokes Equation" *Physical Review Letters* **56**, 1505-1508.

Frisch, U., d'Humières, D., Hasslacher, B., Lallemand, P., Pomeau, Y. and Rivet, J. (1987). "Lattice Gas Hydrodynamics in Two and Three Dimensions" *Complex Systems* **1**, 649-707.

Gauglitz, P.A., Radke, C. J. (1988) "An extended evolution equation for liquid film breake-up in cylindrical capillaries", Chem. Eng. Science, 43, 1457-1465.

Gao, Y., Sharma, M.M. (1994). "A LGA model for fluid-flow in heterogeneous porous-media." Transp. Porous Media **17** (1), 1-17.

Genabeek, O., Rothman D. (1999). "Critical behavior in flow through a rough-walled channel". Phys. Letters A, **255**, 31-36.

Ginzbourg, I., Adler, P.M. (1994) "Boundary flow condition analysis for the three-dimensional latice-Boltzmann model" Journal de Physique II **4**, 191-214.

Hardy, J., Pomeau, Y., de Pazzis O. (1973). "Time Evolution of a Two-Dimensional Model System. I. Invariant States and Time Correlation Functions" *J. Math. Phys.* **14,** 1746-1759.

He, X., Luo, L.S. (1997). *Phys. Rev.* E 55, 6333.

Higera, F. and Jimenez, J. (1989). "Boltzmann approach to lattce gas simulations." *Europhys. Lett.* **9**, 663-668.

d'Humières, D., Lallemand, P., and Frisch, U. (1986). "Lattice gas models for 3D hydrodynamics" *Europhys. Lett.* **2**, 291-297.

Kohring, G.A. (1991a). "Fffect of finite grain-size on the simulation of flow through porous-media" Journal de Physique II 1 (2), 87-90.

Kohring, G.A. (1991b). "Calculation of permeability of porous-media using hydrodynamic cellular automata" J. Stat. Phys. **63** (1-2), 411-418.

Kohring, G.A. (1992). "The cellular automata approch to simulating fluid-flows in porous media" Physica A **186** (1-2), 97-108.

Liang, Z. R., Fernandes, C. P., Magnani, F.S. and Philippi, P. C., (1998) "A Reconstruction Technique of 3-D Porous Media by using Image Analysis and Using Fourier Transform", *Journal of Petroleum Science and Engineering***, 21**, 3-4, 273-283

McCarthy, J.F. (1994). "Flow through arrays of cylinders: lattice-gas cellular automata calculations" Phys. Fluids **6**(2), 435-437.

McNamara G. G. and Zanetti G. (1988). "Use of the Boltzmann Equation to Simulate Lattice-Gas Automata" *Physical Review Letters* **61**, 2332-2335.

McNamara, G., Alder, B. (1993). "Analysis of lattice Boltzmann treatment of hydrodynamics", Physica A, 194, 218-228.

Qian, Y. H., d'Humières D., Lallemand P. (1992) "Lattice BGK Models for Navier-Stokes Equation" *Europhys. Lett.* **17**, 479-484.

Rothmann, D. H., Keller, J. M. (1988). "Immiscible Cellular-Automaton Fluids" *J. Stat. Phys*. **52**, 1119-1127.

Rothman, D. H., Zaleski, S. (1997). *Lattice-gas cellular automata: simple models of complex hydrodynamics* (Cambridge University Press).

Santos, L.O. E., Philippi, P.C., Damiani, M.C., Fernandes, C.P. (2001.a). "Using three-dimensional reconstructed microstructures for predicting intrinsic permeability of reservoir- rocks based on a Boolean lattice gas method". *Journal of Petroleum Science and Engineering (submitted).*

Santos, L.O.E., Philippi, P.C. Bertoli, S.L. Facin. P.C. (2001.c). "Lattice-gas Models for Single and Two-Phase Flows: Application to the Up-Scaling Problem in Porous Microstructures", Comp.Appl.Mathematics (submitted)

Santos, L.O.E., Philippi, P.C., (2001.b). "A lattice-gas model based on field mediators for immiscible fluids", Phys. Rev. E (submitted)

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

ELEMENTS OF INDUSTRIAL HEAT TRANSFER PREDICTIONS

Florian Menter
AEA Technology GmbH
Staudenfeldweg 12
D-83624 Otterfing/Germany
E-mail: Florian.Menter@cfx-germany.com

Thomas Esch
AEA Technology GmbH
Staudenfeldweg 12
D-83624 Otterfing/Germany
E-mail: Thomas.Esch@cfx-germany.com

Abstract. *Different aspects of the accurate prediction of industrial heat transfer problems are discussed. The authors list advanced elements required for the simulation of turbulent boundary layer flows with heat exchange. The paper covers the different aspects of turbulence model selection, near-wall treatment and transition modelling.*

Keywords. *SST turbulence model, scalable wall functions, automatic near-wall treatment, transition modeling based on local variables*.

1. Introduction

Heat transfer plays a major role in many technical devices, ranging from heat exchangers to gas turbine blades or rocket engines. However, the number of CFD simulations aimed at heat transfer predictions is still relatively low. It is estimated that less than 1 % of all industrial CFD simulations target the prediction of heat transfer to and from solid walls. This indicates that thermal designs are still largely based on correlations rather than on three-dimensional CFD computations.

While engineers are using simple correlations, there has been a significant number of publications on CFD methods for heat transfer predictions. Many of these publications have concentrated on the develop-ment of advanced turbulence models including second moment closure models (SMC) for the turbulent stress tensor and the turbulent heat flux vector (Dol and Hanjalic, 1997). The application of a full SMC model to three-dimensional flows requires the solution of six additional equations for the turbulent stresses and an additional three equations for the turbulent heat flux vector. Practical experience has shown, that even for fluid dynamics simulations without heat transfer, SMC models impose a high burden on numerical schemes and therefore require significant computer resources. While there are flows, where the solution of a full SMC model offers significant advantages like buoyancy or swirling flow, SMC models have not consistently produced results of higher quality than simpler turbulence closures. The present authors concentrate therefore on eddy-viscosity models for the momentum equations in combination with an eddy-diffusivity model for heat transfer. This level of closure is considered to offer a good balance between the accuracy, robustness and cost requirements imposed by industrial projects. Based on the experience for flows with momentum exchange, it is expected that this level of closure will be able to give adequate results for a wide variety of industrial applications, provided that an optimal choice is made concerning the different options available.

In addition to the closure level, there are other elements of the turbulence model formulation, which have a significant effect on the accuracy of heat transfer simulations, like the turbulence scale equation and the near-wall treatment. Turbulence models are typically based on the ε-equation to determine the turbulent length scale. The choice in the wall boundary condition is either wall functions (Viegas and Rubesin, 1983) or integration to the surface using a low *turbulent* Reynolds number (low-Re) formulation (Jones and Launder, 1972). Experience has shown that the choice of the scale equation has a significant effect on the prediction of turbulent flows, and that the ε-equation has severe limitations in the near-wall region. It is well known (Rodi and Scheuerer, 1986) that models based on the ε-equation lead to an overprediction of the turbulent length scale in flows with adverse pressure gradients, resulting in high wall shear stress and high heat transfer rates. In combination with low-Re number extensions, the ε-equation has proven to be numerically stiff, leading to a significant reduction in numerical robustness. In addition, these models require a very fine near-wall resolution, which is typically one order of magnitude higher than for other one- and two-equation models.

In order to alleviate the deficiencies of the ε-equation in the near-wall region, the concept of a two-layer formulation (Patel et al, 1985) is frequently used in industrial CFD simulations. In this formulation, the ε-equation is only solved in the outer part of the boundary layer, whereas the inner portion of the logarithmic layer and the viscous sublayer is treated by a mixing length formulation. This method is available in numerous CFD methods, and has also been tested extensively for heat transfer predictions (Bredberg et al., 2000). The experience of the current authors is that the model works well for simple geometries and equilibrium flows. For complex flows, the coupling between the mixing length and the ε-equation becomes problematic and the solutions depend strongly on the specification of the matching location. As this location is usually determined by user input, the uniqueness of the solution is not guaranteed. More elaborate schemes, which base the matching point on a solution variable, like the turbulent Reynolds number, have a tendency to oscillate or even diverge in cases where the matching criterion is not satisfied (laminar regions, etc.). For modern CFD codes using unstructured meshes and scalable parallelisation, the two-layer model is no longer compatible with the data structures of the code. The study of Bredberg, et al (2000) has further concluded that the two-layer model is " … inadequate for heat transfer predictions in rib-roughed channels through its faulty length-scale representation".

Considering the numerous deficiencies of the standard ε-equation near the wall, alternative formulations, both of the scale-equation and of the near-wall treatment are required. An alternative to the low-Re number formulation of the k-ε model has been proposed by Durbin, (Parneix et al., 1999). The V2F model avoids explicit low-Re number terms in the ε-equation by using an elliptic relaxation equation near the wall. In addition, an equation for the fluctuating turbulent stress component normal to the wall, $\overline{v'^2}$ is introduced. The model requires the solution of four partial differential equations compared to two equations in the standard k-ε model, but improved results for a number of test cases have been reported by the authors.

An alternative to the ε-equation is the ω-equation in the form developed by Wilcox (1993). Instead of the equation for the turbulent dissipation rate, ε, an equation for the turbulent frequency, ω, of the large scales is used. The ω-equation has significant advantages near the surface and accurately predicts the turbulent length scale in adverse pressure gradient flows, leading to improved wall shear stress and heat transfer predictions. Furthermore, the model has a very simple low-Re formulation, which does not require additional non-linear wall damping terms. The correct sublayer behaviour is achieved through a Dirichlet boundary condition for ω. It could be argued that the sublayer behaviour of the ω-equation is in principle similar to the elliptic relaxation used in the V2F model. Near the wall, the convective terms in the ω-equation are zero and the equation is dominated by the elliptic diffusion terms, by non-linear source terms, and by the boundary conditions.

One of the main advantages of the k-ω model is its robustness even for complex applications, and the reduced resolution demands for an integration to the wall. It was pointed out by Menter (1992) that the main deficiency of the standard k-ω model is the strong sensitivity of the solution to free stream values for ω outside the boundary layer. In order to avoid this problem, a combination of the k-ω model near the wall and the k-ε model away from the wall has been proposed, leading to the SST (Shear-Stress-Transport) model (Menter, 1994). The SST model will be the basis for most of the other elements of heat transfer predictions described in this paper.

Another important element of heat transfer predictions is the treatment of the equations near a wall. It has already been pointed out that there are two main methods for the near-wall treatment. In most industrial CFD simulations, a wall function approach is used, which bridges the viscous sublayer and makes use of the available knowledge in the logarithmic profile region of the boundary layer. This method allows the use of much coarser near-wall grids, which also benefits the cell aspect ratio. On the downside, wall functions impose strict limitations on the grid generation process, as they impose an upper limit on the grid density in order to remain in the logarithmic part of the boundary layer. Experience has shown that this is a severe limitation, which has negatively impacted a large number of industrial CFD simulations. In order to avoid this limitation, a new approach to wall function treatment was introduced by Grotjans and Menter (1998), which has since been optimised. The method will be presented in Section 3.

In order to provide an even higher flexibility in the choice of a suitable grid, another step forward in the near-wall treatment will be discussed, namely the automatic near-wall treatment for ω-equation based models. This method switches automatically from a wall function to a low-Re formulation, based on the grid density provided by the user. It ensures therefore an optimal accuracy of the CFD solution on a given grid.

A frequently neglected element of heat transfer predictions is the transition from laminar to turbulent flow. The transition process can have an important effect on the wall temperature distribution of devices operating at relatively low-Re numbers, like gas turbine blades. Usually, the simulations are carried out as "fully turbulent", with the turbulent boundary layer starting in the nose region of the blade, or the leading edge of a plate. In reality, there can be a substantial portion of laminar flow, with significantly lower heat transfer rates. Some attempts have been made to predict this phenomenon with the help of low-Re k-ε turbulence models (see Savill, 1993). However, the conclusion is that low-Re models alone are not capable to accurately model transition. An alternative is the use of methods, based on experimental information in the form of correlations (Mayle, 1991; Abu-Ghannam and Shaw, 1980). While these models are more reliable, they have the disadvantage that their formulation is based on non-local variables, like the momentum thickness Reynolds number. Non-local models are not compatible with modern CFD codes using unstructured meshes and scalable parallelisation. A new method has been developed by the authors, which allows the combination of correlation based methods with a transport equation, using local variables. Aspects of this model will be presented in the last section of this paper.

All simulations in this paper are based on the commercial CFD software packages CFX-TASCflow and CFX-5, (AEA 2001).

2. The SST Turbulence Model

The basis for accurate heat transfer simulations is a reliable and robust turbulence model. It is the experience of the present authors that for boundary layer flows, higher order turbulence models offer little advantage over well calibrated eddy-viscosity models. The eddy-viscosity model does however have to satisfy a number of requirements in order to capture the main characteristics of the boundary layer development. One of the most important demands is that the model must be designed to avoid the build-up of turbulence in stagnation regions, as is frequently observed in standard two-equation models. Particularly for heat transfer predictions, these unphysically high turbulence levels have a strong influence on the heat transfer rates inside the boundary layer. Modifications of the standard formulation, like the model by Kato and Launder (1993), or the production limiter of Menter (1994) are available to overcome this deficiency.

A major boundary layer effect is the separation from a surface under adverse pressure gradient conditions. Separation has a strong effect on the near-wall turbulence and therefore on the turbulent heat transfer. The SST model, has demonstrated the capability of accurate separation predictions in numerous cases, (Bardina et al., 1997) and is used as the basis for heat transfer predictions by the present authors. The idea behind the SST model is to combine the best elements of the k-ε and the k-ω model with the help of a blending function *F1*. *F1* is one near the surface and zero in the outer part and for free shear flows. It activates the Wilcox model in the near-wall region and the k-ε model for the rest of the flow. By this approach, the attractive near-wall performance of the Wilcox model can be utilized without the potential errors resulting from the free stream sensitivity of that model. The SST model also features a modification of the definition of the eddy viscosity, which can be interpreted as a variable c_μ. This modification is required to accurately capture the onset of separation under pressure gradients. The formulation of the SST model is as follows:

$$\frac{\partial \rho k}{\partial t}+\frac{\partial \rho U_j k}{\partial x_j}=\tilde{P}_k-\beta^*\rho\,\omega\,k+\frac{\partial}{\partial x_j}\left(\Gamma_k\frac{\partial k}{\partial x_j}\right) \tag{1}$$

$$\frac{\partial \rho\omega}{\partial t}+\frac{\partial \rho U_j \omega}{\partial x_j}=\frac{\gamma}{\nu_t}P_k-\beta\rho\,\omega^2+\frac{\partial}{\partial x_j}\left(\Gamma_\omega\frac{\partial \omega}{\partial x_j}\right)+(1-F_1)2\rho\,\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial k}{\partial x_j}\frac{\partial \omega}{\partial x_j} \tag{2}$$

with

$$\Gamma_k=\mu+\frac{\mu_t}{\sigma_k},\quad \Gamma_\omega=\mu+\frac{\mu_t}{\sigma_\omega}\quad P_k=\tau_{ij}\frac{\partial U_i}{\partial x_j},\quad \tilde{P}_k=\min(P_k;c_l\varepsilon),\quad \mu_t=\rho\frac{a_1k}{\max(a_1\omega;S\cdot F_2)} \tag{3}$$

The coefficients, φ of the model are functions of F_1: $\varphi=F_1\varphi_1+(1-F_1)\varphi_2$, where φ_1,φ_2 stand for the coefficients of the k-ω and the k-ε model respectively:

$\sigma_{k1}=1.176$, $\sigma_{\omega 1}=2.000$, $\kappa=0.41$, $\gamma_1=0.5532$, $\beta_1=0.0750$, $\beta^*=0.09$, $c_l=10$

$\sigma_{k2}=1.000$, $\sigma_{\omega 2}=1.168$, $\kappa=0.41$, $\gamma_2=0.4403$, $\beta_2=0.0828$, $\beta^*=0.09$ (4)

with

$$F_1=\tanh(\arg_1^4);\ \arg_1=\min\left(\max\left(\frac{\sqrt{k}}{\beta^*\omega y};\frac{500\upsilon}{y^2\omega}\right);\frac{4\rho\sigma_{\omega 2}k}{CD_{k\omega}y^2}\right);\ CD_{k\omega}=\max\left(2\rho\,\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial k}{\partial x_j}\frac{\partial\omega}{\partial x_j};1.0e^{-10}\right) \tag{5}$$

$$F_2=\tanh(\arg_2^2);\ \arg_2=\max\left(2\frac{\sqrt{k}}{\beta^*\omega y};\frac{500\upsilon}{y^2\omega}\right) \tag{6}$$

$$\tau_{ij}=\mu_t\left(\frac{\partial U_i}{\partial x_j}+\frac{\partial U_j}{\partial x_i}-\frac{2}{3}\frac{\partial U_k}{\partial x_k}\right)-\frac{2}{3}\rho k\delta_{ij} \tag{7}$$

The only difference to the original formulation given by Menter (1994) is that the absolute value of the strain rate, S, is now used in the definition of the eddy viscosity instead of the vorticity in order to increase the generality of the method beyond aerodynamic applications.

In analogy to the modelling of the turbulent stress tensor, the turbulent heat flux vector is modelled with the help of a turbulent diffusivity:

$$\overline{u_i'T'} = -\varepsilon_h \frac{\partial T}{\partial x_i} = -\frac{\nu_t}{\Pr_t}\frac{\partial T}{\partial x_i} \quad \text{with} \quad \Pr_t = \frac{\nu_t}{\varepsilon_h} \tag{8}$$

In the absence of source terms, the temperature equation is therefore written as:

$$\frac{\partial \rho T}{\partial t} + \frac{\partial \rho U_j T}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\left(\frac{\mu}{\Pr} + \frac{\mu_t}{\Pr_t}\right)\frac{\partial T}{\partial x_j}\right] \tag{9}$$

The Prandtl number, Pr, is a fluid property, and the turbulent Prandtl number, $\Pr_t$, is set to a constant value, assuming an analogy between turbulent heat and mass transfer. Experiments and analysis, place the turbulent Prandtl number in the range of 0.9, which is also used in the present formulation.

Figure (1) shows simulations with the SST and the k-ε model for the flow in a plane diffuser (Gersten et al, 1987). The results demonstrate the superior performance of the SST model under adverse pressure gradient conditions. For more results see Menter (1994) and Bardina et al (1997).

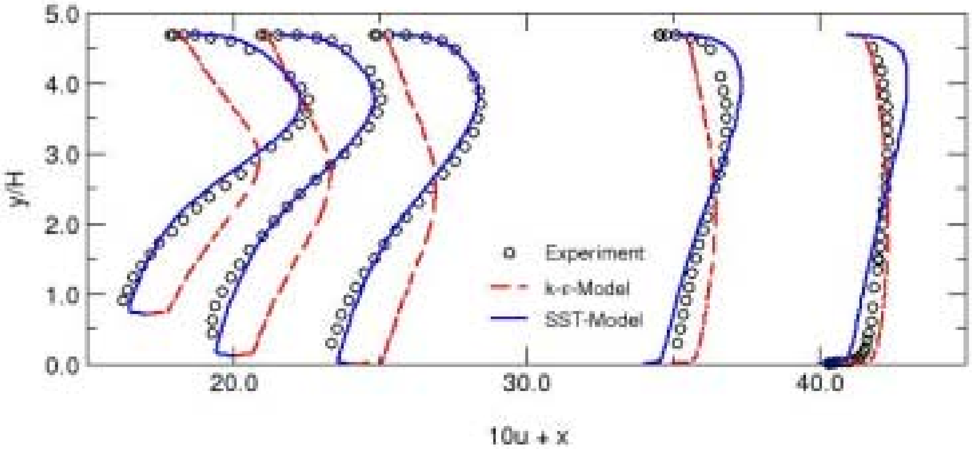


Figure 1: Comparison of computed velocity profiles for diffuser flow with experimental data.

Figure (2) shows the set-up for the experiment of Baughn et al (1984) of an axisymmetric pipe expansion with heat transfer.

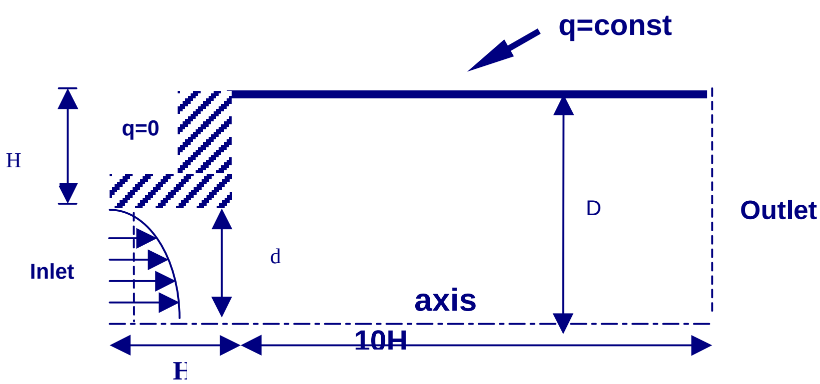


Figure 2: Experimental set-up for axisymmetric pipe expansion with heat transfer.

Figure (3) shows results for this case with different turbulence models. A fine grid of 100x240 points was used for the simulation downstream of the step (see also Section 3 on near-wall modelling). Grid convergence was ensured by a series of simulations on five different grids. The SST model captures the wall temperature well, even in the separated region. The computations have been performed with the low-Re number wall treatment of the model discussed in the following section. The k-ε model, on the other hand, leads to an overprediction of the heat transfer in the region of flow reattachment. This is a result of the overprediction of the turbulent length scale in this region. Note that in this test case, the separation is fixed by the sudden change in the geometry. This avoids the k-ε model's problems to predict the onset of separation as observed in Fig. (1). Included in Fig. (3) is also a computation with the two-layer k-ε turbulence model. This formulation significantly underpredicts the wall heat transfer rate. This is consistent with the observations in Bredberg et al. (2000). As the high Reynolds number versions of the two k-ε model simulations are identical, Fig.(3) also demonstrates the importance of the near-wall treatment of the equations.

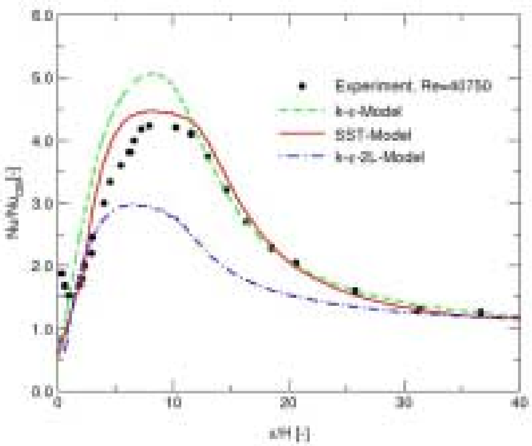


Figure 3: Comparison of experimental Nusselt number distribution with results from the SST and the k-ε model for sudden pipe expansion.

An important case for heat transfer predictions is the flow in a cooled driven cavity as shown in Fig.(4), (Metzger et al., 1989). Profiles of the velocity and turbulence field are specified at the inlet from a fully developed pipe flow. The temperature is constant at the inlet, and the lower wall of the cavity is cooled. Flows of a similar kind are present inside turbine blades with cooling ribs, and the prediction of the heat transfer is of high importance with respect to the thermal loading on the blades. The simulations have been carried out on a series of three different grids with a total number of 4.182, 16.362 and 64.722 nodes in the 2-D plane. The results are shown for the finest grid, which gave results in close agreement with the medium grid.

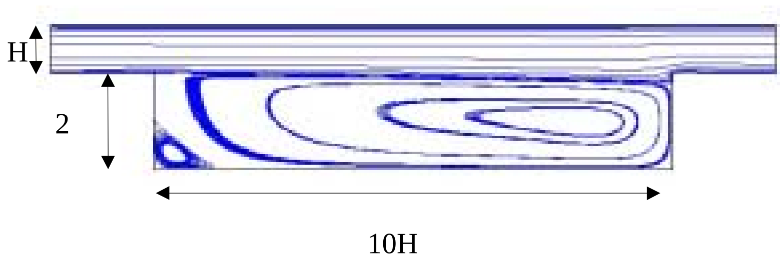


Figure 4: Experimental set-up and streamlines (SST model) for driven cavity flow with heat transfer.

Figure (5) shows a comparison of the simulations with the k-ε model with scalable wall functions, the k-ε two-layer model and the SST model against the experimental data. Apparently, this flow is not quite as sensitive to the turbulence model selection as the pipe expansion, but the SST model gives nevertheless the best agreement with the experimental data. The model picks up the increased heat transfer near the downstream sidewall of the cavity, which is not predicted accurately by the k-ε models. Again, the differences due to the change in the near-wall treatment for the k-ε model are significant and of the same order as the differences between the k-ε and the SST model.

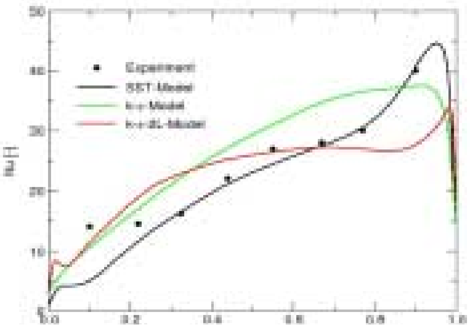


Figure 5: Comparison of experimental Nusselt number distribution with results from the SST and the k-ε model for a driven cavity.

3. Near-wall Treatment

3.1. Scalable Wall Functions

The idea behind the scalable wall function is to avoid the limitations imposed by standard wall functions in terms of near-wall grid resolution. Typically, the near-wall grid spacing has to satisfy requirements of the form $y^+ > Y_{low}^+$ where $Y_{low}^+ \approx 20$, depending on the numerical formulation. Particularly for flows at low device-Reynolds numbers, this is a severe limitation, as the boundary layer can be quite thin so that it cannot be resolved with a coarse near-wall grid. In the past a remedy was to generate one coarse cell near the wall and to refine the grid from the second cell on. This is a tedious process, which can only be applied in simple geometries. The goal of the present formulation is to allow the user to generate grids without imposing a lower limit from the wall functions.

The wall function formulation is most easily explained for a cell-centred discretisation, as shown in Fig.(6).

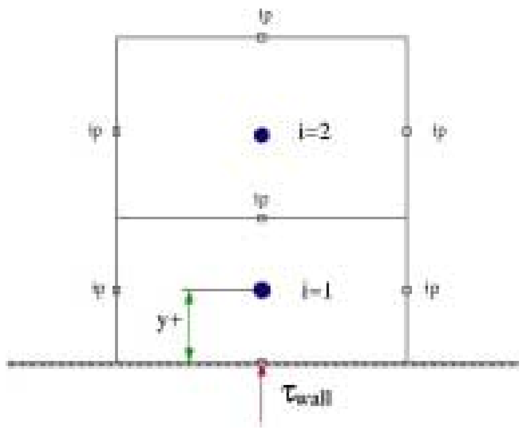


Figure 6: Near-wall grid for cell-centred discretisation.

The conservation equations are assembled by integration over control volumes. The fluxes are computed at the integration points ***ip***. The missing flux at the wall must be supplied by the boundary condition. In case of wall functions, this is usually achieved by the application of a logarithmic profile assumption. It implies that the first grid point (i=1) lies in the logarithmic part of the boundary layer.

$$u_\tau = \frac{U_1}{\frac{1}{\kappa}\log(y^+) + C}; \quad \tau_{wall} = \rho u_\tau^2; \quad y^+ = \frac{u_\tau y}{\nu} \tag{10}$$

This relation is only correct if the first grid point (i=1) lies in the logarithmic region. In case of a fine near-wall grid, this restriction is violated and the wall function accuracy will deteriorate. Eventually, the formulation will become singular as $y^+ \to 0$. The extent of the logarithmic region is shown in Fig. (7) for two different Reynolds numbers. For flows at high Reynolds numbers, the extent of the logarithmic region is quite large and it is easy to place the first grid point there, while at the same time satisfying the grid resolution requirements of the boundary layer. For the flow at low Reynolds numbers, the extent of the logarithmic layer is much reduced and it becomes increasingly difficult to place the first grid point in that region. Even if this can be achieved, it would in most cases result in an under-resolved boundary layer.

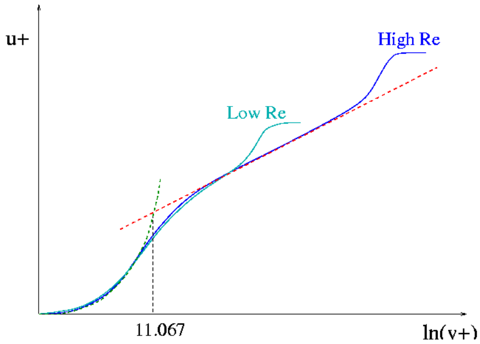


Figure 7: Near-wall scaling of turbulent boundary layer

In order to remedy this shortcoming of wall functions, the computation of the wall shear-stress is slightly altered to:

$$u_\tau = \frac{U_1}{\frac{1}{\kappa}\log(\tilde{y}^+) + C}; \quad \tilde{y}^+ = \max(y^+, Y^+_{\lim}); \quad \tau_{wall} = \rho u^2_\tau; \tag{11}$$

In contrast to the first formulation of the scalable wall functions, (Grotjans and Menter, 1998) only a lower limit is imposed on this new version. The limiting value of $Y^+_{\lim} = 11.067$ marks the intersection between the logarithmic and the linear profile.

On fine grids, the definition of $\tilde{y}^+$ is therefore decoupled from the grid spacing. The limiter prevents the first grid point to slide into the linear profile area. Note that there are no changes to the grid generation process, and arbitrarily fine grids can be used in the near-wall region. The physical interpretation for fine grids is that the wall is treated as if it would be the edge of the viscous sublayer. For grids, where the limiter is activated, the relationship between u_τ and U_1 becomes linear. As a linear profile is present even under severe non-equilibrium conditions, this results in a reduced dependence on the logarithmic profile assumption. The error introduced by this formulation is that the effect of the displacement of the viscous sublayer is not accounted for in the simulation. However, this is the case for all wall function formulations as the sublayer is never accurately resolved. As the relative thickness of the sublayer increases with decreasing device Reynolds number, this error can become significant under low device Reynolds number conditions.

The new formulation allows the use of arbitrarily fine near-wall grids without a violation of the logarithmic profile assumption. The specifics of the formulation depend on the numerical approach and will differ from code to code. The principal idea can however be applied to all wall function formulations, like in CFX-5 where a vertex based scheme is used.

For the treatment of the energy equation near the wall, an algebraic formulation is required to link the temperature and the heat flux. The formulation of Kader (1981) is used here:

$$\Theta^+ = \Pr\cdot y^+ \cdot e^{-\Gamma} + \left[2.12\ln(1+y^+) + \beta(\Pr)\right] e^{-1/\Gamma}; \quad \beta(\Pr) = (3.85\Pr^{1/3} - 1.3)^2 + 2.12\ln(\Pr) \tag{12}$$

with $\Gamma = \dfrac{0.01(\Pr y^+)^4}{1 + 5\Pr\cdot y^+}$, and the definition of the non-dimensional temperature:

$$\Theta^+ = \frac{T_w - T}{T_\tau}; \quad T_\tau = \frac{q_w}{\rho c_p u_\tau} \tag{13}$$

T_w is the wall temperature, q_w, the wall heat flux, c_p the heat capacity and ρ is the density. This formulation is valid through the entire y^+ range of the viscous sublayer and the logarithmic profile. For the current formulation, $\tilde{y}^+$ has to be used in order to be consistent with the momentum equations.

Figure (8) shows results for the simulation of the heat transfer for the sudden pipe expansion for three different grids (100x15, 100x30, 200x240 points past the step in the axial and normal directions, respectively). Figure (8) shows that the grid sensitivity is significantly reduced for the scalable wall functions. Note that the solution for the standard wall function deteriorates on the fine grids.

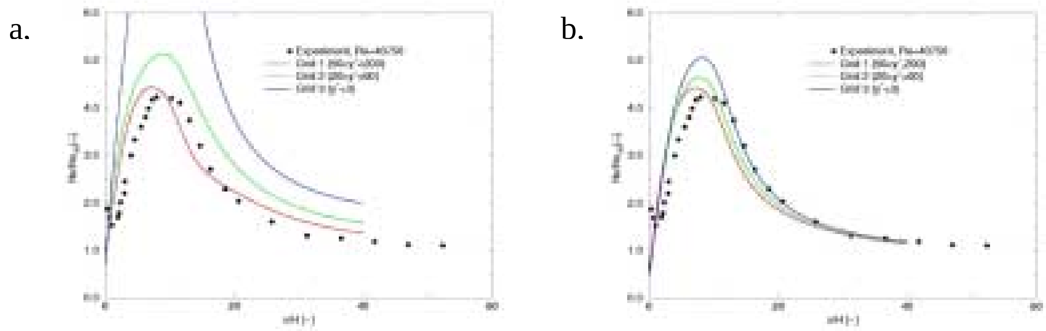


Figure 8: Grid sensitivity for wall function formulations. a: Standard wall function. b: Scalable wall

3.2. Automatic Wall Treatment

As pointed out in the last section, wall functions are not always desirable, as they neglect the influence of the viscous sublayer. Especially for flows at low device Reynolds numbers, the omission of the sublayer can have a significant effect on the solution. For example, the mass flow rate in a pipe or channel flow at low Reynolds numbers can be in error by 10 % and more, due to the relatively large influence of the viscous portion of the boundary layer. For external flows, the effect is usually smaller, as the displacement of the outer flow is then a second order effect.

Nevertheless, it is desirable to have the option of a robust and accurate viscous sublayer formulation to be able to solve the equations all the way to the surface. As already pointed out, low-Re k-ε based models do generally not satisfy the requirements imposed by complex industrial flow simulations. The current method is therefore based on the k-ω near-wall formulation. While the treatment of the ω-equations for fully resolved simulations is known (Wilcox, 1993; Menter, 1994) the optimal formulation would avoid the stringent grid resolution requirements of this low-Re formulation. The idea behind the automatic near-wall treatment is that the model shifts gradually between a viscous sublayer formulation and wall functions, based on the grid density. The ω-equation is well suited for this task, as it provides analytical solutions, both for the sublayer and the logarithmic region. A blending function depending on y^+ can therefore be defined.

The solutions for ω in the linear and the logarithmic near-wall region are:

$$\omega_{Vis} = \frac{6\nu}{0.075 y^2}; \quad \omega_{\log} = \frac{1}{0.3\kappa}\frac{u_\tau}{y} \tag{14}$$

They can be re-formulated in terms of y^+ and a smooth blending

$$\omega_1(y^+) = \left(\omega_{Vis}^2(y^+) + \omega_{\log}^2(y^+)\right)^{0.5} \tag{15}$$

can be performed. A similar formulation is used for the velocity profile near the wall:

$$u_\tau^{Vis} = \frac{U_1}{y^+}; \quad u_\tau^{\log} = \frac{U_1}{\frac{1}{\kappa}\ln(y^+) + C}; \quad u_\tau = \left[\left(u_\tau^{Vis}\right)^4 + \left(u_\tau^{\log}\right)^4\right]^{0.25} \tag{16}$$

This formulation gives the relation between the velocity at (i=1) and the wall shear stress. For the k-ε quation, a zero flux boundary condition is applied, as this is correct both for the low-Re and the logarithmic limit. For the energy equation, the formulation given by Eq.(12) is used. Again, the formulation depends to some extent on the discretisation method and on the available form of the wall functions, which has to be recovered on coarse grids.

Figure (9) shows the smooth shift of the wall treatment from a viscous sublayer to a wall function, based on the grid spacing. Figure (10) shows the dependence of the skin friction coefficient, c_f, on the grid spacing Δy^+. Figure (11) shows the variation of the heat transfer with grid resolution for a flat plate for a standard low-Re and the automatic near-wall treatment. The differences are in the opposite direction as for the wall functions, see Fig.(8). The standard low-Re results deteriorate for the coarser grids as expected, whereas the automatic wall treatment gives consistent results for all grids.

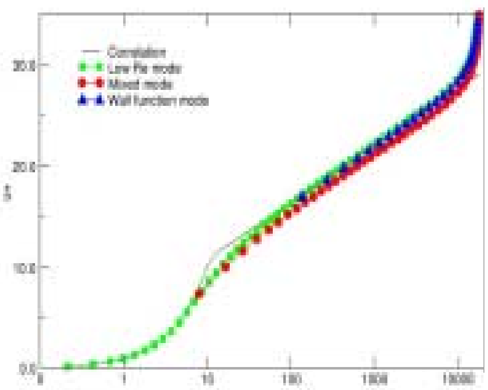


Figure 9: Automatic wall treatment for velocity profile

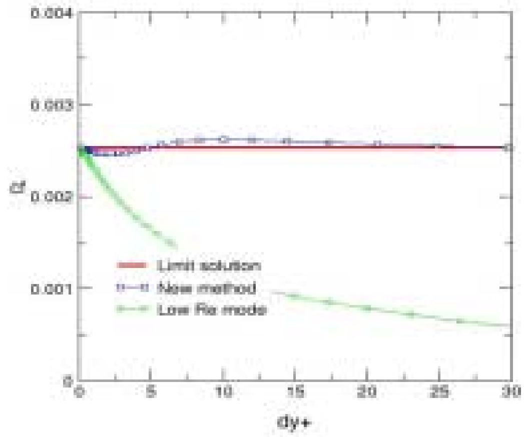


Figure 10: Grid sensitivity for different near-wall modes

a. b.

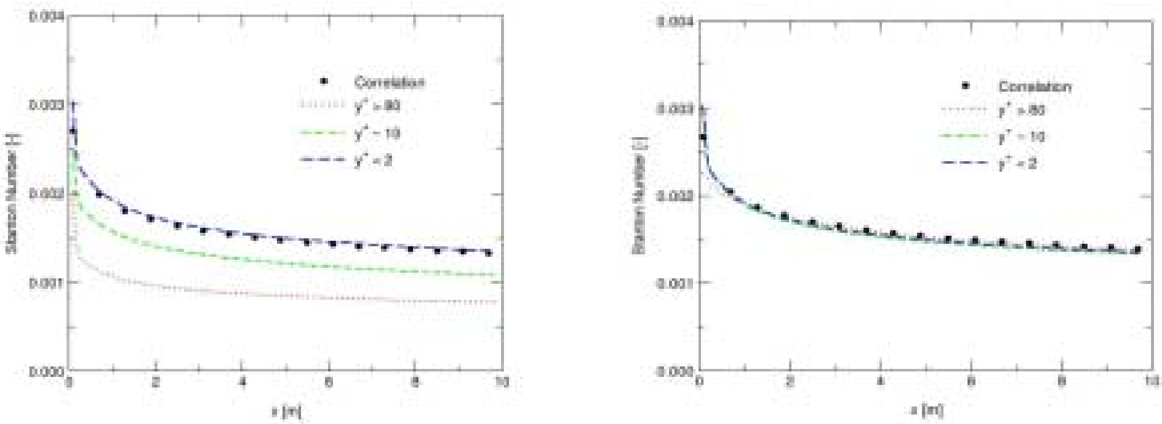


Figure 11: Grid sensitivity for integration to the wall. a, Standard low-Re model; b, automatic near-wall treatment

4. Transition Modelling

The location of the start and the extension of transition are of major important in the design and performance of many technical devices, where the wall-shear-stress or wall heat transfer is of interest. For turbomachinery flows, transition is usually imposed by high turbulence levels in the core flow. This leads to “by-pass” transition, where turbulence from the core flow is diffused into the laminar boundary layer. There are mainly two engineering methods to handle by-pass transition. The first is the use of low-Re turbulence models, with the hope that the low-Re extensions can capture the main effect. The second is based on experimental correlations, (Mayle 1991; Abu-Ghannam and Shaw 1980) which link the main cause of transition (free stream turbulence) to the transition location.

It was observed (Savill, 1993) that some low-Re k-ε models predict realistic transition locations for by-pass scenarios. However, the ability of a low-Re turbulence model to predict transition seems to be coincidental, as the calibration of the damping functions is based on the viscous sublayer behaviour and not on transition from laminar to turbulent flow. This is supported by the fact that only some models show this capability, although most models work well for the sublayer. The physical mechanism of low turbulence intensity in the sublayer and the transition region are quite different and it is not realistic to expect a consistent transition modelling capability from this approach.

The second approach, which is favoured by industry over the low-Re models is the use of experimental correlations. The correlations usually relate the free stream turbulence intensity, Tu, to the transition Reynolds number based on the momentum thickness. These transition criteria are usually linked with a two equation turbulence model by a modification of the turbulent production term. In the laminar regions, the turbulent production term is multiplied by zero and in the turbulent region by one. The transition model does therefore not affect the laminar or the fully turbulent solution. While this method proves sufficiently accurate, it poses numerical and programming challenges in Navier-Stokes methods.

The main problem is the necessity to compare actual momentum thickness Reynolds numbers with the critical value from the correlation. This is not an easy task in a Navier-Stokes solver environment, as the boundary layer edge is not well defined and the integration will therefore depend on the details of the implementation of the search algorithm. The use of correlation based transition criteria is also incompatible with modern CFD codes. As a result, one has to either abandon the correlation based transition methods, to stay with current solver technology, or to develop correlation based transition models which do not require non-local information.

A first step towards a transition model based on local variables has recently been made by the present authors. The model is formulated in terms of a transport equation for a non-linear disturbance, which can be linked to an intermittency function. The intermittency changes from zero to one in the region specified by the transition criterion. Due to space limitations, no complete account of the rationale behind the model is given here. Only the equations are listed. The model has been submitted for publication, (Menter 2002).

$$\frac{\partial(\rho f)}{\partial t} + \frac{\partial(\rho U_j f)}{\partial x_j} = c_{f1}\rho S F_{G1} f(1 - c_{f2} f) + \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_f}\right)\frac{\partial f}{\partial x_j}\right] \tag{17}$$

$$F_{G1} = \max(\tilde{\xi}_1(1 + c_{f3}\xi_2) - 1, 0)$$

$$\xi_1 = \frac{\Phi_1}{\mathrm{Re}_{\Theta t}\cdot 0.026} \qquad \Phi_1 = 80\frac{Sy^2}{\nu} \qquad \tilde{\xi}_1 = \frac{\xi_1}{\left(1 + (0.5\xi_1)^4\right)^{0.25}} \qquad \xi_2 = \frac{\Phi_2}{\mathrm{Re}_{\Theta t}\cdot 0.051} \qquad \Phi_2 = \frac{\nu_T}{\nu}$$

$$F_t = \frac{\tilde{f}^6}{\left(\tilde{f} - 0.01\right)^6 + 1} \quad \textit{with} \quad \tilde{f} = \max\left(\frac{f}{2} - 5, 0\right) \qquad \tilde{P}_k = F_t P_k$$

S is the strain rate, y is the distance from the surface and the model constants are:

$$c_{f1} = 0.5; \quad c_{f2} = 0.1; \quad c_{f3} = 5.0; \quad \sigma_t = 1.0; \tag{18}$$

The link to the transition criterion is through the momentum thickness Reynolds number at transition $\mathrm{Re}_{\Theta t}$. It is at least a function of the turbulence intensity, Tu, and can be computed from experimental correlations (Mayle, 1991).

Figure (12) shows solutions computed with this method for the VKI-MUR 247 turbine blade (Arts, 1991). The value of $\mathrm{Re}_{\Theta t}$ was specified from the transition location of the experiment. Note that the simulation is nevertheless non-trivial, as the model does not explicitly compute the momentum thickness Reynolds number, but works exclusively with local variables. Laminar flow is predicted on the lower surface, in agreement with the experimental data. It is clear that the accurate prediction of this flow requires the modelling of the transition process.

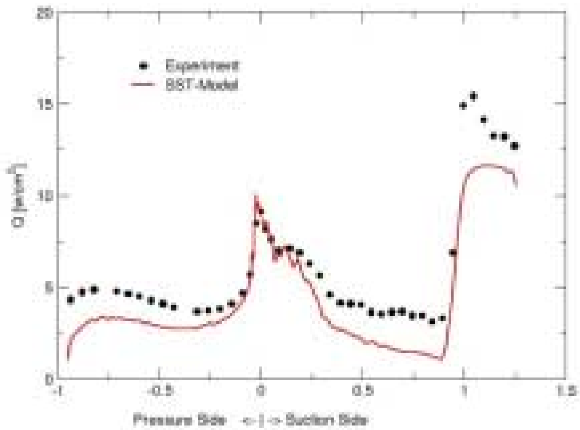


Figure 12: Prediction of heat transfer for turbine blade with local transition model.

5. Summary

Different elements of heat transfer predictions methods have been discussed. The emphasis was on advanced eddy-viscosity based models. It was shown that these models offer significant benefits for industrial heat flow simulations and that they can be used for a range of applications. The importance of the optimal selection of the different elements of heat transfer predictions, ranging from the turbulence model to the near-wall treatment of the equations to the inclusion of transition phenomena was demonstrated.

6. Acknowledgement

The authors want to thank Dr. G. Scheuerer for his valuable comments during the preparation of this paper.

7. Reference List

Abu-Ghannam, B.J. and Shaw, R., 1980, “ Natural transition of boundary layers - the effects of turbulence, pressure gradient and flow history”. J. of Mechanical Engineering Science, Vol. 22, pp. 213 - 228.

AEA Technology, 2001, CFX-TASCflow and CFX-5 Documentation. Harwell UK.

Arts, T., 1991, Data available via ERCOFTAC special interest group: (http://www.mecaflu.ec-lyon.fr/CONGRES/ERCOFTAC/database/register.html).

Bardina, J.E., Huang, P.G. and Coakley, T., 1997, “Turbulence Modeling Validation”, AIAA Paper 97-2121.

Bredberg, J., Davidson, L. and Iacovides, H., 2000, “Comparison of near-wall behavior and its effect on heat transfer for k-ω and k-ε turbulence models in rib-roughened channels”. In Nagano,. Y., Hanjalic, K., and Tsuji, T. eds: 3rd Int. Symposium on Turbulent Heat and Mass Tranfer, pp 381-388.

Baughn J., 1984, "Local Heat Transfer Downstream of an Abrupt Expansion in a Circular Channel With Constant Wall Heat Flux", Journal of Heat Transfer, Vol 106, pp. 789 – 796.

Dol, H. and Hanjalic, K., 1997, “ Development of a differential thermal second-moment closure using DNS data of the natural convection in a vertical channel”. In Hanjalic, K. and Peeters, T.W.J.: 2nd Int. Symposium on turbulence, heat and mass transfer, Delft University Press.

Gersten, K., Härten, A. and Pagendarm, H.G., 1987, In: “Optimierung von Diffusoren bezüglich der Diffusorströmung und der Diffusorwände”. VDI-Verlag Düsseldorf.

Grotjans, H., and Menter, F.R., 1998, “Wall functions for industrial applications”. In K.D. Papailiou, editor Computational Fluid Dynamics'98, Volume 1, Part 2, pages 1112-1117, Chichester. ECCOMAS, John Wiley Sons.

Jones, W.P. and Launder B.E., 1972, ”The prediction of laminarization with a two-equation model of turbulence”. International Journal of Heat and Mass Transfer, 15.

Kader, B.A., 1981, “Temperature and concentration profiles in fully turbulent boundary layers”. Int. Journal of Heat and Mass Transfer, 24, pp 1541-1544.

Kato, M. and Launder B.E., 1993, “The modelling of turbulent flow around stationary and vibrating cylinders”. In: Ninth Symposium on Turbulent Shear Flows, Kyoto, Japan.

Mayle, R.E., 1991, “The role of laminar-turbulent transition in gas turbine engines”. Journal of Turbomachinery, 113, pp 509-537.

Menter, F.R., 1992, “Influence of free stream values on k-ω turbulence model predictions”. AIAA Journal, 30(6), pp 1657-1659.

Menter F.R., 1994, “Two-equation eddy-viscosity turbulence models for engineering applications”. AIAA-Journal, 32(8), pp. 269-289.

Menter, F.R. 2002, “Transition modelling based on local variables”, submitted for publication to the 5th International Symposium on Engineering Turbulence Modelling and Measurements in Camp de Mar, Mallorca, Spain,

Metzger, D.E., Bunker, R.S. and Chyu, K.V., 1989, "Cavity heat transfer on a transverse grooved wall in a narrow channel", J. Heat Transfer, Vol. 111, pp 73-79.

Michelassi, V. and Shih, T.H. 1991, “Low Reynolds number two-equation modeling of turbulent flows”, ICOMP-91-06; CMOTT-91-01.

Patel, V.C., Rodi, W. and Scheuerer, G., 1985, “Turbulence models for near-wall and low Reynolds number flows - a review”. AIAA J. 23, 1308.

Parneix, S., Behnia, M. and Durbin, P.A., 1999,”Predictions of turbulent heat transfer in an axisymmetric jet impinging on a heated pedestal”. J. Heat Transfer 121, 43.

Rodi, W., and Scheuerer, G., 1986, “Scrutinizing the k – ε Turbulence Model Under Adverse Pressure Gradient Conditions”. J. Fluids Eng. 108.

Savill, A.M., 1993,” Some recent progress in the turbulence modelling of by-pass transition”, In: R.M.C. So, C.G. Speziale and B.E. Launder, Eds.: Near-Wall Turbulent Flows, Elsevier, p. 829.

Viegas, J.R.; Rubesin, M.W. 1983, “Wall-function boundary conditions in the solution of the Navier-Stokes equations for complex compressible flows”. AIAA-83-1694.

Wilcox, D.C., 1993, “Turbulence Modeling for CFD”. DCW Industries, Inc., La Canada, CA, 1993.

A GENERALISED FRACTIONAL DERIVATIVE APPROACH TO VISCOELASTIC MATERIAL PROPERTIES MEASUREMENT AND VIBRATION CONTROL DESIGN

José João Espíndola
Federal University of Santa Catarina - UFSC
Brazil

Viscoelastic materials are largely used as a means to provide damping to structures, thus mitigating resonant vibration responses. Devices made with viscoelastic materials such as isolators, dynamic vibration neutralisers (also called dynamic vibration absorbers), and structural links can be designed for highly efficient vibration control (Espíndola, J.J.; Bavastri, C.A.; 1995, 1997, 1998, 1999, Espíndola, J.J., Floody, S.E., 1999).

To properly design a vibration control strategy with viscoelastic materials, two basic dynamic properties must be measured; de material loss factor and the dynamic modulus of elasticity.

In the past the rheological model for viscoelastic material was based on the classical concept of derivative (with respect to time) of integer order. These constitutive equations contain too many parameters to be identified, which makes it computationally impractical.

To avert the direct computation of the model parameters, the vibrating sandwich beam method has been devised (Jones, D.I.G., 1979).

By this approach a three layer beam, in which the viscoelastic material forms the inner layer, is excited to vibration and the response characteristic measured. From the response characteristic of the beam the loss factor and the elastic modulus are extracted. The basic beam model is based on the RKU (Ross, D., Kerwin, E.M., Ungar, E.E., 1959) theory.

This is an extremely simplified model, leading to closed form solutions for both loss factor and dynamic modulus of elasticity. It bears too many simplifications that might eventually bring experimental difficulties. Nevertheless it has been used for many years up to now.

To improve this approach, advanced models have beam devised for the sandwich beam dynamics with great success (Espíndola, J.J., Brandon, J.A., 1996, Lopes, E.M.O., Brandon, J.A., Espíndola, J.J., 1996, Brandon, J.A., Lopes, E.M.O., Espíndola, J.J., 1999).

To obtain the loss factor and dynamic modulus of elasticity over a wide band of frequencies and temperatures, the beam test is performed inside a chamber with controlled temperature.

At LVA (Laboratory of Vibration and Acoustic of the University of Santa Catarina) the temperature inside the chamber can be set from -30ºC to +60ºC.

Over the last two decades the concept of fractional (or generalised) derivative (Ross, B., 1974) has gained the reputation of an extremely adequate tool to model viscoelastic materials (Bagley, R.L. and Torvik, P.J., 1979, Padovan, J., Guo, Y., 1988, Pritz, T., 1998).

The constitutive equation for viscoelastic material “in principle” also needs many parameters, but this is misleading. Fractional derivatives are such an intimate descriptor of rheological materials behaviour that only five, even four parameters are enough to accurately represent a particular material.

This fact bears enormous consequences as to the approach to measure the loss factor and the dynamic modulus, as well as to the optimum design of viscoelastic derives for vibration control.

In this lecture a new approach for the measurement of loss factor and dynamic modulus will be produced. This work is being carried out within the LVA.

A simple single degree of freedom viscoelastic system and its frequency response function is used to identify such proprieties in a wide band of frequencies and temperatures.

The quality and simplicity of the approach will be stressed throughout.

How this new model impacts on the design procedures of viscoelastic devices for vibration control will also be discussed.

REFERENCES:

- Espíndola, J.J., and Bavastri, C.A., “Reduction of Vibration in Complex Structures with Viscoelastic Neutralisers: A Generalised Approach”, Proceedings ASME Design Engineering Technical Conferences, Boston, 1995, Vol. 3, Part C, pp 761-766.
- Espíndola, J.J., and Bavastri, C.A., “ Modal Reduction of Vibrations by Dynamic Neutralisers in a Frequency Range – A Generalised Approach”, Proceeding of Diname 95, 1995, Caxambú, MG, pp. 214-217.
- Espíndola, J.J., and Bavastri, C.A., “An Efficient Definition of Transmissibility for a General Equipment Isolation System”, Paper DETC97-VIB 4120, ASME Design Conference, Sacramento – California, 1997 (in CD-ROM).
- Espíndola, J.J., and Bavastri, C.A., “Reduction of Vibration in Complex Structures with Viscoelastic Neutralisers: A Generalised Approach and a Physical Realisation”, Proceedings ASME Design Engineering Technical Conferences, Sacramento 1997, Paper DETC97/VIB-4187, in CD-ROM.
- Espíndola, J.J., and Bavastri, C.A., “Viscoelastic Neutralisers in Vibration Abatement: A Non Linear Optimisation Approach”, Journal of the Brazilian Society of Mechanical Sciences, Vol. XIX, no 02, March/1997, pp. 154-163 (Indexed by Applied Mechanics Reviews and Engineering Information, Inc.).
- Bavastri, C.A., Espíndola, J.J. and Teixeira, P.H., “ A Hybrid Algorithm to Compute the Optimal Parameters of a System of Viscoelastic Vibration Neutralisers in a Frequency Band”, Movic’98, Zurich, Switzerland, 1998, Vol.2, pp. 577-582.
- Espíndola, J.J., and Floody, S.E., “On the Modelling of Metal – Elastomer Composite Structures: A Finite Element Method Approach”, Applied Mechanics in the Americas, Vol. 8, pp. 1335 – 1342, Rio de Janeiro, Brasil, 4-8 January, 1999.
- D.I.G. Jones, “Two Decades of Progress Damping Technology”, Aircraft Engineering, pp. 9-16, January 1979.
- D. Ross, E.E. Ungar, and E.M. Kerwin, Jr. “Damping of plate flexural vibrations by means of viscoelastic laminate,” Structural Damping, ASME, New York, pp. 49-88, 1959.
- Espíndola, J.J., Brandon, J.A., and Lopes, E.M.O., “Numerical Conditioning in the Inverse Problem of Heterogeneous Sub-Structures”, Proceeding of the International

Congress MV 2, New Advanced in Modal Synthesis of Large Structures Non-Linear, Damped and non Deterministic Cases, Lyon-France, October/1995, Vol. 1, pp. 125-136, ed. By L. Jezequel.

- Lopes, E.M.O., Espíndola, J.J., and Brandon, J.A., " On the use SVD for Solving Inverse Problem in Sandwich Beams", Proceeding of the International Conference on Identification in Engineering Systems – Swansea – Wales, 27-29 March/1996, pp. 184-193.
- Lopes, E.M.O., Brandon, J.A., and Espíndola, J.J., "Some Recent Results in the Inverse Analysis of Sandwich Beams", ISMA 21 – 1996 International Conference on Noise and Vibration, Leuven, Belgica, 1996.
- Brandon, J.A., Lopes, E.M.O., and Espíndola, J.J., "Some Aspects of Experimental and Theoretical Reanalysis Using Polymeric Material", Second International Conference on Identification in Engineering, Swansea, U.K., 1999.
- B.Ross, "A Brief History and Exposition of the Fundamental Theory of Fractional Calculus", in Lecture Notes in Mathematics, Vol. 457, Springer-Verlag, New York, 1975, pp. 1-36.
- Bagley, R.L., Torvik, P.J., "A Generalised Derivative Model For a Elastomer Damper", Shock and Vibration Bulletin, 49, 1979.
- Padovan, J., Guo, Y. "General Response of Viscoelastic Systems Modelled by Fractional Operators", Journal of the Franklin Institute, Vol. 325, No. 2, pp. 247-275, 1988.
- Pritz, I., "Frequency Dependencies of Complex Moduli and Complex Poisson's Ratio of Real Solid Materials", Journal of Sound and Vibration (1998) 214 (1), 83-104.

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

THE STRUCTURAL DYNAMICS MID-FREQUENCY CHALLENGE: THE GAP BETWEEN FEA AND SEA

Arruda, José Roberto de França
Universidade Estadual de Campinas – UNICAMP/FEM
Brazil

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

THE CHALLENGES FOR MACHINE AND MECHANISM DESIGN AT THE BEGINNING OF THE THIRD MILLENIUM AS VIEWED FROM THE PAST

Marco Ceccarelli
Laboratory of Robotics and Mechatronics
DiMSAT – University of Cassino, Via Di Biasio 43, 03043 Cassino (Fr), Italy
ceccarelli@ing.unicas.it

Abstract. The purpose of this paper is to overview and discuss basic concepts and facts in historical development of Mechanism and Machine Science in order to outline and motivate the future challenges for Machine and Mechanism Design. The overview has been presented by using few examples of few significant authors/designers.

Keywords. *History of Engineering, History of Kinematics, History of Mechanism Design, Theory of Machines and Mechanisms.*

1. Introduction

In the coming Computer Age will the Mechanical Engineering and particularly Machine and Mechanism Design (MMD) have a fundamental role and successful application like in the past?

Two main considerations may give a positive answer:

- whatever Electronics, Informatics, Telecommunications and so on, will be enhanced and expanded in Technology, Mechanical Design will be always needed since a woman/man will always live and interact with the environment on the basis of mechanical phenomena of the human nature;
- enhancements in knowledge and Technology needs for human life and industrial production changed and evolved over time, also because of the evolution of MMD, requiring innovation that MMD was always able to achieve.

In this keynote paper the two above-mentioned observations have been addressed by looking at the historical development of Mechanical Engineering from the viewpoint of MMD.

The historical semantic meaning of Engineering and MMD has been reviewed to understand and stress the aim and goals of the design activity in the form of continuous challenging problems. In particular, MMD can be recognized as a substantial part of TMM (Theory of Machines and Mechanisms) as the modern field of MMS (Machine and Mechanism Science) was named until the terminology has been updated officially by IFToMM, the International Federation for the Promotion of Mechanism and Machine Science (formerly International Federation for the Theory of Machines and Mechanisms) in July 2000, (IFToMM 2000a).

In this paper, the historical evolution of MMD has been outlined by referring to few emblematic examples of few significant authors/designers. The topic of History of MMS addresses great attention not only in the Community of History of Science, but more and more even from technical viewpoint in the Community of Kinematicians and IFToMM at large, as shown from the recent publications in the Proceedings of 1999 IFToMM World Congress (Leinonen Ed. 1999) and HMM2000 Symposium (Ceccarelli Ed. 2000).

Of course, it is quite possible to be a good mechanical engineer without any knowledge of the History of MMS. However, from the point of view of the discipline as a whole it is believed that historical research is necessary. Historical research throws light on the identity of MMS, it helps us to show who deserves the credit for specific contributions in the field, and it can lead to a fuller understanding of what machines and mechanisms are and how they have been designed and used over the time.

There is a considerable literature on the general History of Technology, on the history of certain specific areas of technology and on the history of mathematics, mechanics, physics, etc. Publications on the History of TMM, however, are rather rare. This fact gives to the History of MMS the dignity of a topic requiring specific attention for specific papers. We do not refer to “humanistic” historical papers but to papers about the History of MMS with a technical content. Precisely because the subject has become very technical and in order to gather information from people who participated in the developments, a conscious effort has been made to stimulate members of the IFToMM community to write about the history of their discipline. The goal was to stimulate experts in MMS with some feeling for history, people who can understand, appreciate and refresh past works in MMS, to write a historical paper. These papers should be written by and for experts in MMS with the aim to make them aware of technical developments in the past and give them further motivation and ideas for their research.

This paper is an attempt to synthetically overview and discuss basic concepts and historical developments in the specific field of MMD with the aim to identify patterns for future investigation from the past experiences and determine main challenges for the future of MMD. In fact, one will discover in the proposed overview that most of the modern ideas and approaches may come from past experiences so that similarly we can learn from the past how to attach the future problems.

Many other authors have attached the problem of outlining the History of MMS at different level of content, in the past as for example Chasles (1837) and Reuleaux (1875), and recently as for example De Jonge (1943), Ferguson (1962), Hartenberg and Denavit (1964), Hain (1967), Nolle (1974), Crossley (1988), Dimarogonas (1993), Marchis (1994), Angeles (1997), Ceccarelli (1998), and even as bibliographic work only, as in the case of De Groot (1970).

Indeed, very rich reference lists can be found also in some historical overviews, as for example in (Hain 1967), (Nolle 1974), (Marchis 1994). Moreover, in some case the historical overview has been focused on specific topic or machines, as for example in (Kestell 1963) on boots and shoes manufacturing, and in (Koetsier 1997) on early programmable machines. In addition, overviews on recent State-of-Art on Mechanisms and related advanced application fields have been presented in Journals and Congresses with a view to future activity, in many countries, as for example in (Roth 1983), (Hunt 1984), (Dubowsky 1993), (Freudenstein 1993), (Ruggieri 1997), (Garcia de Jalon 1997), and (Ceccarelli 2000c) only to cite some. Recently, the International Symposium on History of Machines and Mechanisms HMM has been established within the technical IFToMM community to have a forum for discussion and publication of papers in the field of History of MMS, and particularly TMM, but from technical viewpoint. The Proceedings of the event HMM2000 has been published as (Ceccarelli Ed. 2000)

This paper can be considered as a complementary work to the above-mentioned sources (to which the reader is kindly suggested to refer for an in-depth historical information on MMS), not only for the space limits, but mainly for the aim of focusing the historical overview to outline future challenges of MMD.

2. Historical meaning of Machine and Mechanism Design

The term Engineer is a modern terminology that has been established clearly to indicate designers of industrial systems only at the beginning of XIXth century with the foundation of Schools of Engineering in Europe, starting with the Ecole Polytéchnique in Paris in 1794.

But the Engineering activity has been started since the beginning of Technology and particularly of Design activity for systems and structures. However, Schools of Engineering were established with specific aims when the demand of expert well-educated persons has been reached a certain level of need for the industrial development.

The term Engineer can be considered a further specialization of the competence of Architects. In fact, in Antiquity designers were named as Architects in agreement with the semantic meaning of the Greek word τεκτων that means constructor and therefore, by general extension, designer. Nevertheless, even in the Greek Classical Age a differentiation appeared when μηχανικοι were named peoples, who were specifically experts in the design of mechanical systems that were used mainly in the Theatre for play representations. Architects were named the Roman Engineers and the technical culture was addressed as Architecture both for civil constructions and mechanical systems (for war machines, automation and transportation). An emblematic example is the encyclopedic work by Vitruvius (who lived in 1st century B.C.) titled “De Architectura”, (Vitruvius 1511) and (Barbaro 1584), containing even a special chapter liber X on mechanical systems of any type.

Only in the late Roman Empire the word “ingenius” started to be commonly used with the aim to describe mechanical devices with very innovative design and/or operation. During the Middle Ages the term “magister ingeniosissimus” and more specific “engignier” was used mainly in South France and North Italy.

But only in 1496 Ludovico il Moro used the term “Ingeniarius” in an official document in Milan, Italy, to indicate designers whose professional competence was clearly differentiated from that one of Architects. However, the term was not always used, as the case of Leonardo Da Vinci (1452-1519) emblematically indicates since he was considered, even by himself, as an architect more than an Ingeniarius, although besides his architectural works he has had and has a great repute as designer of new mechanical devices.

At the end of Renaissance the growth of knowledge and need of specific technical competence produced a clear differentiation between Architects and Engineers, who were devoted mainly to the practice of design and construction of machines and mechanisms. But for a long time the mechanical devices were considered of secondary importance with respect to the architectural goals and therefore somehow the engineering activity was considered as a “dirty job”. Emblematic is the case of Filippo Brunelleschi (1377-1476), who is recognized as a great architect but only few know his “ingenium” of designing specific machines that he used for the construction of his architectural masterpieces.

During the XVIIth century more than engineers, the designers were considered as experts in Mechanics that reached great results both in Theory and Practice. However, the term was clearly established in Military field with the Ecole de Genié Militaire.

In conclusion, the term Engineer was clearly coined and used since the beginning of XIXth century in the form and meaning that we still use today. In the Mechanical Engineering the field of Machine and Mechanism Design reached a mature identity and great appreciation since the second half of XIXth century within the discipline named as Theory of Machines and Mechanisms.

TMM (Theory of Machines and Mechanisms) is often misunderstood even in the IFToMM Community, although it is recognized as the specific discipline of Mechanical Engineering related with mechanisms and machines.

The meaning of TMM can be clarified by looking at the meaning of the topic over time through few definitions by significant Authors as in the following, (Ceccarelli 1999b):

by Marco Pollione Vitruvius (he lived in 1st century B.C.) in De Architectura- liber X, translated and discussed by Daniele Barbaro in (1584): “A machine is a combination of materials and components that have the capability of moving weights”;

by Galileo Galilei in (1593): “A machine is a means by which a given weight will be transported to a given location by using a given force”;

by Paolo Branca in (1629): He described machines by stressing “the operation as consumption of motor power and changes of wieghts in functions of time history, cost and operator skill”;

by Jacob Leupold in (1724): He treated the description of machines and mechanisms referring to “their aim of modifying motion rather than just the construction of machinery”;
by Josè Maria de Lanz and Augustin de Betancourt in (1808): “In agreement with M. Monge, we consider as elements of machines the devices than can change the direction of the movements… the most complicated machines are only combinations of those capable of single movements”;
by Robert Willis in (1841): “I have employed the term Mechanisms as applying to combinations of machinery solely when considered as governing the relations of motion. Machinery as modifier of force”;
by Franz Reuleaux in (1875): “A machine is a combination of bodies capable of withstanding deformation, so arranged as to constrain the (mechanical) forces of nature to produce prescribed effect in response to prescribed input motions”.
by Francesco Masi in (1897): “Hence we name: as mechanism a kinematic chain that has been fixed on one of its components; as machine a mechanism whose components make mechanical work”;
by Gabriel Koenigs in (1905): “A machine is recognized as an assembly of resistant bodies that are constrained reciprocally and are under the action of natural forces. If you abstract from forces, the remaining of a machine consists of bodies and constraints. This is a mechanism”;
by Richard S. Hartenberg and Jacques Denavit in (1964): “The term machine is associated with the use and transformation of force, and although motion is varying degree is encountered in a machine, the idea of force dominates. Mechanism, on the other hand, definitely conjures up the idea of motion, and while forces do exist, they are relatively small and unimportant compared with the exploitation of motion”.

In addition, IFToMM terminology (IFToMM 1991) gives:

- Machine: mechanical system that performs a specific task, such as of material, and the transference and transformation of motion and force.
- Mechanism: system of bodies designed to convert motions of, and forces on, one or several bodies into constrained motions of, and forces on, other bodies.

Finally, the meaning for word “Theory” needs further explanation. The Greek word for Theory comes from the corresponding verb, whose main semantic meaning is related with examination and observation of existing phenomena. But, even the Classic language the word theory includes practical aspects of observation as experiencing the reality of the phenomena, so that theory means also practice of analysis results. In fact, this last meaning is what was included in the discipline of modern TMM as Gaspard Monge (1746-1818) established in the Ecole Polytechnique, (Chasles 1886), at the beginning of XIXth century (see for example the book by Lanz and Betancourt (1808), whose text include synthesis procedures).

In conclusion since the modern assessment, TMM has been considered as a discipline, which treats analysis, synthesis and practice of mechanisms and machines. This will be also in the future, since we shall always have mechanical devices related with life and working of human beings. These mechanical devices need to be designed and enhanced with approaches from mechanical engineering because of the mechanical reality of the environment where the human beings will always live, although new technology will substitute some components or facilitate the operation of mechanical devices.

The modern significance of TMM has found its expression on the international institutional level with the foundation of the International Federation for the Theory of Machines and Mechanisms (IFToMM) in 1969, Fig.1. The history of IFToMM is outlined by the past IFToMM Presidents in Chapter 2 of (Ceccarelli Ed. 2000).

Fig.1 A historical moment of the foundation of IFToMM, the International Federation for the Theory of Machines and Mechanisms, in Zakopane (Poland) on 27 September 1969, (Courtesy of IFToMM Archive) in which one can recognize: 1- prof. Ivan Ivanovic Artobolevskii (USSR); 2- prof. Adam Morecki (Poland); 5- prof. Nicolae I. Manolescu (Romania); 6- prof. Erskine F. Crossley (USA); 7- prof. Giovanni Bianchi (Italy); 8-prof. Aron E. Kobrinskii (USSR); 9- prof. Werner Thomas (Germany); 10- prof. Jan Oderfeld (Poland).

The main activities of IFToMM can be considered the publication of the Journal Mechanism and Machine Theory (six issues per year, started since 1969) and the organization of the World Congress (every four years; the 11th Congress will be held in 2003) with the aim “to promote research and development in the field of Machines and Mechanisms by theoretical and experimental methods, along with their practical applications”, (IFToMM 2000b). In addition IFToMM is active by means of many specific Commissions in the many field of Mechanical Engineering with particular attention to MMD.

3. Designs from Antiquity

Few original information are available on the design activity for MMD in Antiquity. One can deduce the level of designing activity for machinery and mechanical systems by looking mainly at the literature, monuments and archeological sites, but also by considering the life and culture in Antiquity.

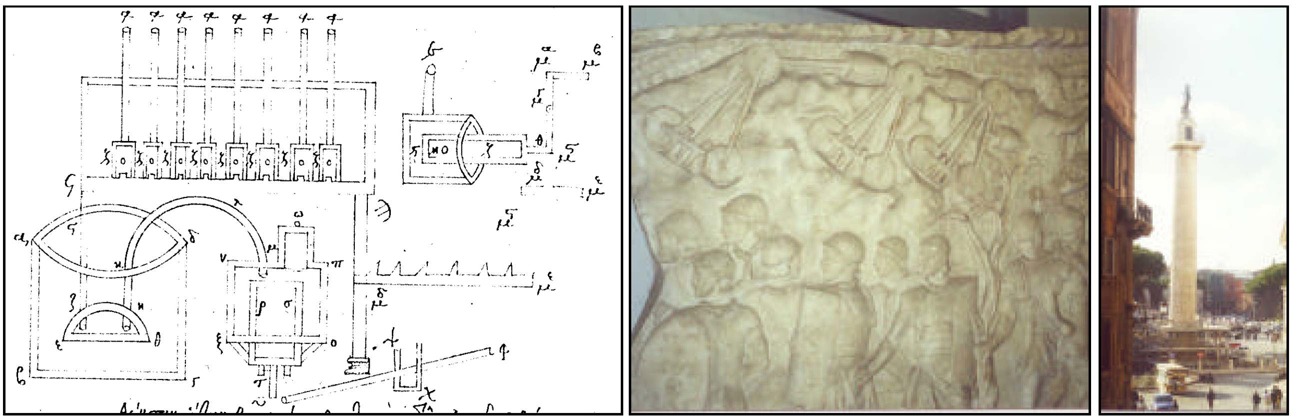
Very few technical documents have reached us today, even if they have been rewritten and translated during Middle Age and then Renaissance with renewed technical interest. Emblematic example is the work by the Roman Vitruvius as given in the references (Vitruvius 1511) and (Barbaro 1584).

The Literature can give description of mechanical designs, sometimes with additional considerations which give however indication of the potential of the technical culture, like today the Space Fiction proposes interplanetary trips with fantastic spatial ships, whose utopia is based on a hopeful evolution of the reality of the current Astronautics. Thus, for example in the Homer’s poetry (Iliad and Odyssey) of the VIIIth century B.C., which reports tales dated up to XIth century B.C., one can find the description of many intriguing mechanical devises (even robots ante litteram). But in the Greek culture since the VIth century B.C. the activity of mechanical designers is well recognized, as pointed out in (Dimarogonas 1993). The Greek tradition of a technical culture and practice reached highs in the IIIth century B.C. with the establishment of the School of Alexandria in Egypt where among the many activities, practical aspects of Mechanics and designing practice received specific attention both for investigation and teaching with a great reputation due to several personalities. First professional engineers/designers can be considered Ctesibius, Filon, Heron and Archimedes, who were formed and worked at the School of Alexandria in IIIth and IIth century B.C. They carried out activity of teaching, investigating and designing mechanisms and systems with various automatic operation features. They were famous in Antiquity for several automatic devises that they designed by using the acquired knowledge in Mechanics and other technical fields. Emblematic is the use of Pneumatics by Ctesibius and Heron, who wrote also several treatises on the subject. Figure 2a) is an illustrative example of the level of knowledge and expertise that they reached in designing complex systems with integration of components and design procedures, likewise nowadays we are attempting in the discipline of Mechatronics.

The great expertise of Roman engineers is well known, since the Roman Empire is recognized to have reached its goals somehow even because of the engineering activity mainly in transportation structures (like roads and bridges) and war machines. The work “De Architectura” by Vitruvius give a significant example of the skill of Romans in designing activity. Unfortunately, there is not original mechanisms and machinery still existing, but one can find documented information even in archeological sites. Emblematic is the example shown in Fig.2b) in which war machines are shown in the Traian’s column that is a kind of reportage of the war against the Dacian warriors. The three-wheeled war machines are illustrated as a significant machinery that was used successfully in the war. The technical representation is not fully understandable but it is significant to indicate the importance and level of this kind of technical means.

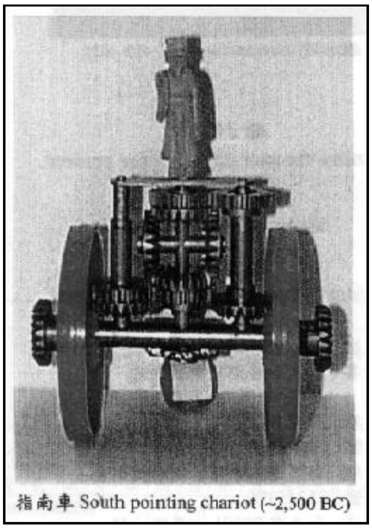
Of course, designing activity was carried out also in other civilizations, even if there were no contacts among them.

For example in China great experiences can be advised in designing activity of mechanical systems, even with very complicate architectures and features, as for the cases shown in Fig.3.

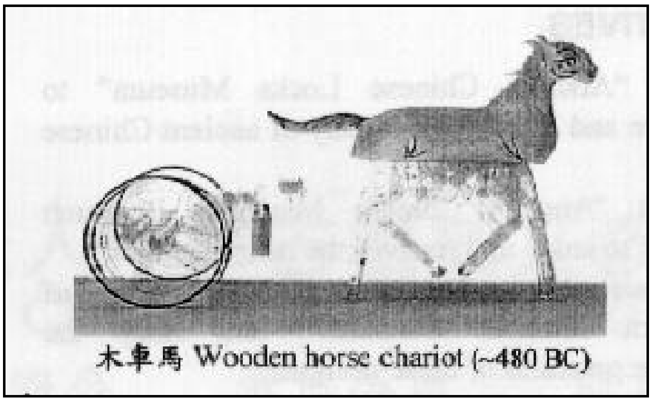


a) b)

Fig.2. Examples of mechanical designs from Antiquity: a) a mechanism of a hydraulic organ designed in the IIth century B.C. by Heron of Alexandria as redrawn in XVth century (Courtesy of Biblioteca Medicea Laurenziana in Florence, Italy); b) war machines represented in the Traian’s column that was built in 113 A.D in Rome (Courtesy of the Museum on Roman Civilization in Rome, Italy).



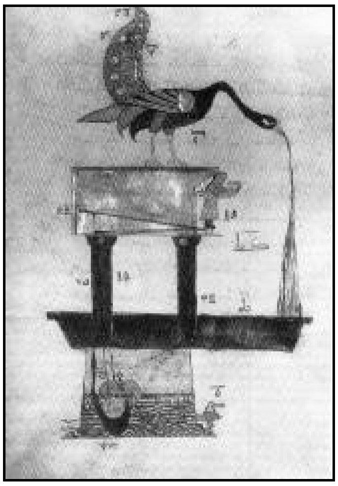
a)



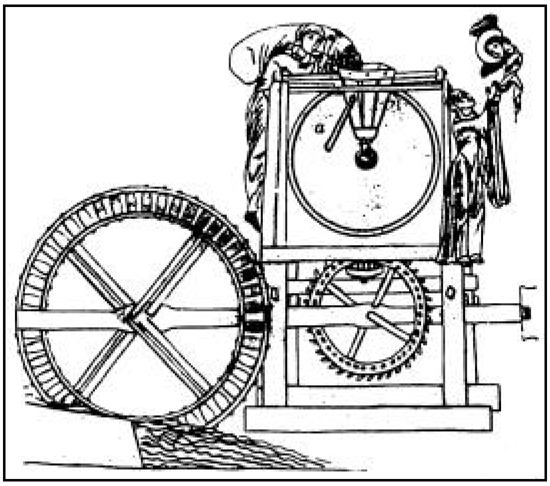
b)

Fig.3 Examples of Chinese designs from Antiquity: a) a modern reconstruction of the south pointing chariot whose design can be dated approximately in 2,500 B.C. (Courtesy of Ancient Chinese Machines Research Center in Tainan, Taiwan); b) a modern reconstruction of the mechanical cow whose design can be dated approximately in 450 B.C. (Yan 1999)

The highs in technical knowledge, and specifically in MMD, were not completely lost over time, although it can be difficult to find and understand this kind of heritage in some periods. For example during the Middle Ages it is usually said that the civilization went back even in technical aspects. But nevertheless the previous knowledge was passed and even enhanced through other civilizations, as for example with the Arabic culture as the design case by Al-Jazari in XIIth century shown in Fig.4a). In addition a kind of advances has been also experienced when one consider the more practical purposes of machine design, as for the case of Fig.4b).



a)



b)

Fig.4 Examples of preserving mechanical designs of the Antiquity in the Middle Ages: a) a hydraulic system designed by Al-Jazari in XIIth century by using the Mechanics by Heron, (Rosheim 1994); b) a flour milling machine redesigned by Herrad von Landsberg (1899) in XIIth century.

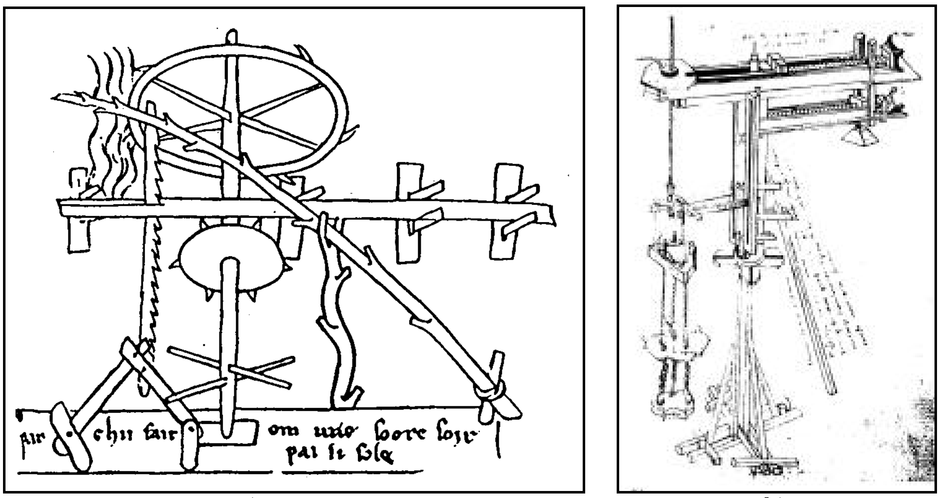
4. Early studies and designs

The modern evolution of MMD can be considered as started at the time of the Renaissance. But, since the time of Middle Ages mechanism design addressed specific attention and emblematic designs show a certain high level of knowledge and expertise by few persons, as in the examples of Fig.5. The growing need of technical means required better designs but mainly new mechanisms. Thus, in the scanty literature, one can find mechanical designs with better and better machines, as the case of Fig.4b). But also new machines were conceived with new solutions even for mechanism design. The example of Fig. 5a) gives an early automatic machine, which is based on a mechanism that is a very first example of body guiding by means of a coupler and even a first use of a five-bar linkage, if one will recognize this mechanism in the unclear drawing at the left bottom. Indeed, a certain unclear explanation of the machine design and operation can be advised as due to the will of a designer to ensure not only paternity (a king of patent copyright) but mainly the consulting. Thus, although the overall meaning of the design and operation of the automatic machine is clear, the practical running of machine required details that only the designer could give. Similarly, the crane device of Fig.5b) shows a complex mechanism containing an early design of a modern system that only recently has addressed great attention with the name of tendon driven parallel manipulator. In the case of Fig.5b) the mechanism design is more clear, since the goal of the designer was not the machine but an architectural masterpiece.

In fact, at an early stage of technical Renaissance one can recognize two different approaches in MMD: practical purpose and secondary means.

Practical purpose of MMD was started by the need to have machines for basic works and it grew with the growing needs of increasing production and variety of tasks, as shown also in the examples of Fig.6.

Secondary aim of MMD was established during the Renaissance when great architectural works required suitable machines to help and complete large constructions. Thus, the case of Brunelleschi is illustrative of this approach that considered the MMD as secondary to the architectural goal, although the design achievements were brilliant as in the case of Fig.5b). In fact, Brunelleschi never considered himself as a machine designer. The view of considering the designing of machinery and mechanisms as a secondary minor activity was persistent for long time and only at the end of XVIth century there was a certain repute of MMD as independent important discipline because of the work by Guidobaldo Del Monte (1543-1607) and mainly Galileo Galilei (1564-1642), being University professor.



a) b)

Fig.5 Examples of early mechanism designs with modern systems: a) an automatic wood sawing machine containing perhaps a five-bar linkage designed by Villard de Honnecourt in the XIIth century, (Bechmann 1991); b) a crane containing a wire parallel manipulator module designed by Filippo Brunelleschi (1420).

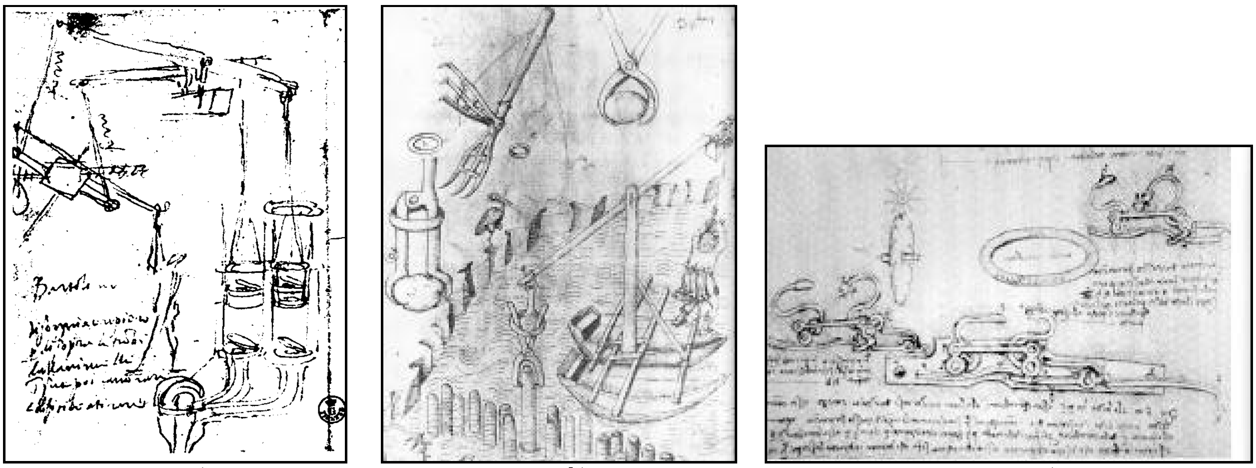
In addition, a limited teaching activity and circulation of the know-how restricted the knowledge and expertise on MMD to few group of persons who were working with a maestro in his bottega as pupils and collaborators at the same time, as discussed in (Ceccarelli 1998). Thus, the secrecy and unclear illustration of new designs were a common practice of early designers during Renaissance. But similar contradictorily behavior of designers is still persistent nowadays in the form of patent licensing and secret consulting.

At the time of Renaissance highs in MMD were achieved by several designers in several field of applications. The cases of Fig.6 are very few examples in which one can recognize: a first kinematic study of a pump operation through geometric schemes in Fig.6a); a first study of different two-finger grippers for several grasping applications in Fig.6b); a first specific attention to mechanisms for a specific automation application in Fig.6c) for gun firing.

The literature of mechanical designs during the Renaissance is very rich, and it is worth to remark that Leonardo da Vinci, although he conceived and designed several innovative machines, was only one of the many who devoted their professional activity partially to MMD.

Since the beginning of practice in mechanical design, an important issue for mechanisms design has been recognized in a classification of mechanisms. Several attempts were proposed in the past and still nowadays the classification of mechanisms is an open problem as concerning the derivation of exhaustive expert systems for computer use in design procedures. Classification of mechanisms has been proposed from two points of view, namely identification of basic simplest components and cataloguing of existing solutions.

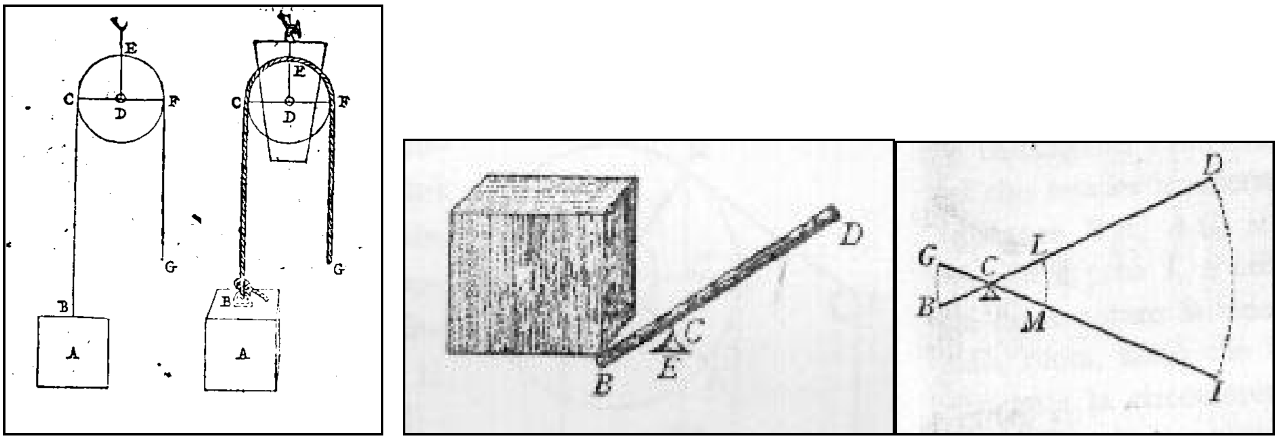
Archimedes was the first who attached the problem from an engineering viewpoint by studying the mechanics of basic machines and particularly the lever and screw. Vitruvius attempted a catalogue of machines for several purposes by using the view of basic machines. Much later, the problem of studying the basic machine components for design purposes has been attached by Guidobaldo Del Monte in (1577) and by Galileo Galilei in (1593) as young University professor. In both texts one can find detailed drawings, models and formulation that can be used to understand the operation of the mechanical design and even to design specific solutions of the basic machines: lever, pulley, wheel and axle, wedge (inclined plane), and screw. Del Monte and Galilei addressed great attention to the lever, in agreement with an Archimedian approach, and they formulated the mechanics of the lever in such a way that they applied it to all basic machines for analysis purposes.



a) b) c)

Fig.6 Early mechanical designs of machinery in the Renaissance period (XV-XVI th centuries): a) by Antonio da Sangallo Il Giovane, (Frommel 1994); b) by Mariano di Jacopo (Il Taccola) (1969); c) by Leonardo da Vinci, (Cianchi 1984)

Considerable is the abstraction of the mechanical design to draw suitable model in which the basic design parameters are outlined in form of first kinematic sketches, as shown in the illustrative examples of Fig.7. In fact, in Fig.7 a) a pulley system is analyzed by Del Monte by using the basic geometry, which has been abstracted from the real components of the system shown in the right part of the drawing. Similarly, in Fig.7b) the motion properties of a lever are studied by Galilei from a suitable synthetic drawing.



a) b)

Fig. 7 Early schemes of basic machine elements: a) the pulley by Guidobaldo Del Monte in (1577); b) the lever by Galileo Galilei in (1593).

The increasing growth of knowledge, demand, and number of designers in XVIth and XVIIth centuries gave great stimulus for theoretical studies and publications of machine designs. Theoretical studies started with attention to the motion of rigid bodies since the Renaissance, as for example (Varronis 1584) and (Wallis 1670); and at the end of XVIIIth century they brought to the modern formulation of the equation of the motion of rigid bodies. The historical development to Screw Theory has been outlined in (Ceccarelli 2000a and 2000d) by reviewing main works up to the modern referenced papers, but stressing the work by Mozzi (1763) as the first rigorous treatise ,on helicoidal motion.

Specific attention to the mechanism characteristics was addressed in XVIIth century and a fundamental work is due to Philippe De La Hire, who first formulated the roulettes and applied them to mechanisms in (De La Hire 1706).

Emblematic is also the work (Roberval 1730) by Personier de Roberval (1602-1675), who attached the problem of path analysis with a certain view to mechanism application, whose an example is reported in Fig.8, which shows also a high level of abstraction from the reality of the body motion. Figure 8 gives also an idea of the level of knowledge that, because of the specific high abstraction level, required specific expertise in the specific field of mechanical design and theoretical mechanics. The Geometry addressed great attention even as a fundamental discipline for machine design, as illustrated for example in (De La Hire 1695; Ozanam 1720; Grandi 1739) that were texts for teaching Mechanics but with a view to machine design.

After the masterpiece works by Del Monte and Galilei the basic machine components were treated as a traditional subject during the XVIIth and XVIIIth centuries. Only in XIXth century they were re-considered with renewed attention, but mainly to deduce kinematic model and formulation that could be helpful for modern design and cataloguing purposes of mechanisms.

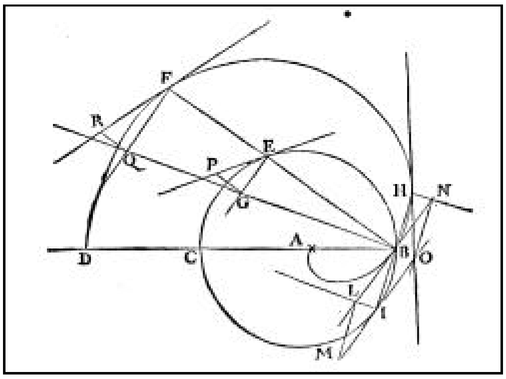
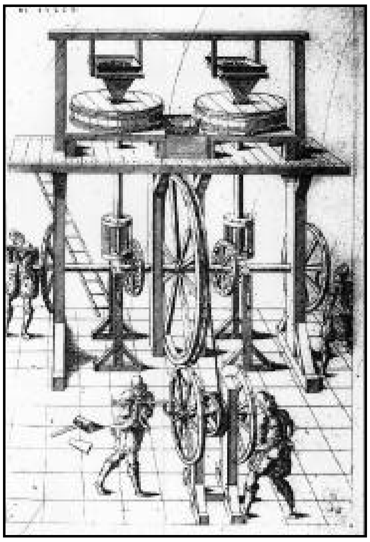


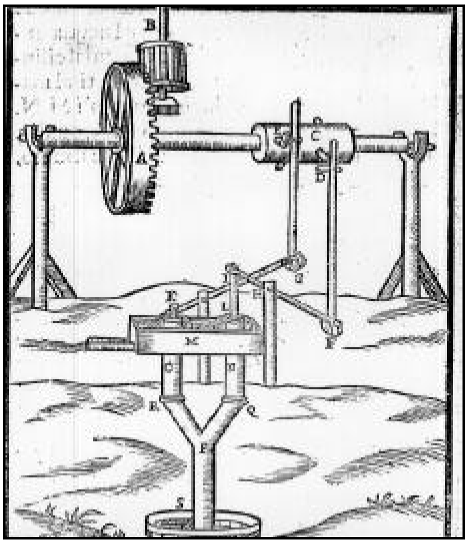
Fig.8 An example of early studies on mechanisms characteristics as a scheme for a rotation of a line by P. De Roberval at the end of XVIIth century, (Roberval 1730).

In the XVIIth century the increase of practice in MMD motivated the need of catalogues of machine designs of common use to help both designers and users as in (Ramelli 1588). Indeed, the use of machine handbooks (at the time named as Theatrum machinorum) was aimed to facilitate the circulation of a technical culture but the use of machines.

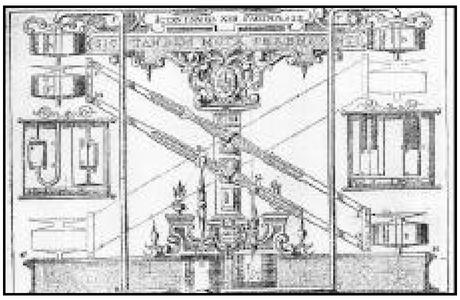
Figure 9 shows examples taken from few catalogues of machines but gives an example of the evolution of the machine complexity before the beginning of the Industrial Revolution. Thus, Fig.9a) shows a gear machine design with large size at the end of XVIth century; Fig. 9b) shows a hydraulic pump powered by a combination of gears and mechanisms, at the beginning of XVIIth century; Fig. 9c) shows a mechanism design that was proposed for perpetual motion as an example of supposed highs achieved by MMD at the end of XVIIth century; Fig.9d) shows a spatial mechanism used at the beginning of XVIIIth century; Fig.9e) shows a complex mechanism for the watches at the end of XVIIIth century.



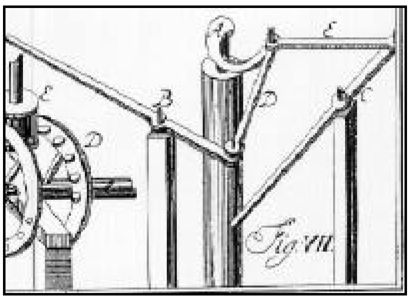
a)



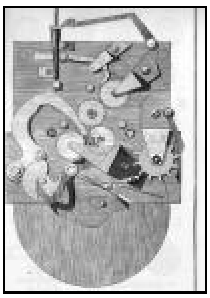
b)



c)



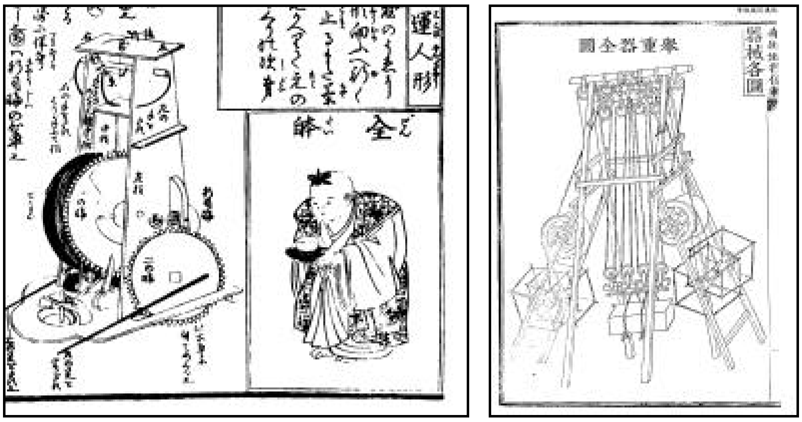
d)



e)

Fig.9 Examples of mechanical designs in early catalogues/handbooks of machines and mechanisms: a) by J. Besson in (1578); b) by G. Branca in (1629); c) by G. Schott in (1664); d) by J. Leupold in (1724); e) by J.B. D'Alembert and D. Diderot in (1774).

Technical evolution was experienced not only in European countries, but like in the past, other civilizations developed their own knowledge in MMD, although since the Renaissance contacts started to be more frequent and even with exchanges of products and knowledge. An example of independent relevant technical evolution can be advised in the Asiatic countries. Figure 10a) gives illustrative example of highs in Japan since XVIIth century in the field of automatic machines through a very advanced design of a gear-mechanism automatic (robotic) system. Figure 10b) shows a complex crane mechanism that was used in Korea in XVIIth century.



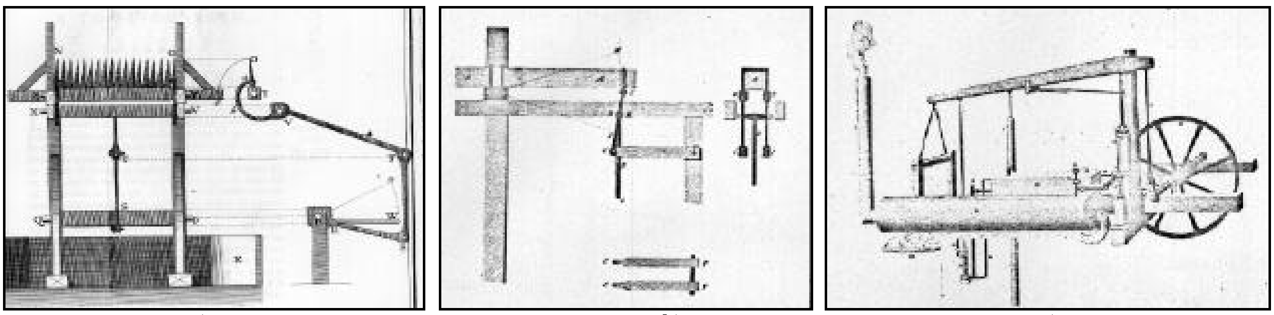
a) b)

Fig.10 Examples of early modern mechanical designs in other countries: a) the automatic tea servant built in Japan since the XVIth century (Rosheim 1994); b) a crane mechanism used for construction of the Korean Royal Palace in XVIIth century. (Courtesy of the Kyujanggak Archive at Seoul National University, Korea).

At the end of XVIIIth century the advent of Industrial Revolution in Europe gave a further great stimulus in advancing MMD since the enlarged demand of machines with larger and larger power and faster and faster operation required to deepen in MMD and mainly design new mechanisms and machines. The early activity was mainly aimed to design mechanisms and machines with practical/experimental application of acquired knowledge. However, the increasing of the need of better and better designs required a deeper and deeper knowledge and more and more rational design procedures. This demands was the motivation of the foundation of Schools of Engineering everywhere and specifically the establishment of TMM and more in particular modern MMD at the beginning of XIXth century. Indeed, the XIXth century can be considered as the Golden Age for TMM and particularly MMD.

But before the establishment of rigorous Academic teaching of TMM and particularly MMD, there was an intense designing activity and practice with new and old mechanisms at the early beginning of Industrial Revolution before the end of the XVIIIth century. This activity required also a certain ordering and rules that led to be the foundation of Patent offices, that were first established in Great Britain and France.

Examples of this new approach for mechanism design are illustrated in Fig.11: Fig. 11a) shows the mechanical design of an automatic sawing machine together with details of an early kinematic study of the operation of the main mechanism; Fig.11b) shows the mechanical design of the famous Watt 4-bar straight-line mechanism as deposited by James Watt (1736-1819) with indication of used material but detail in construction and motion performances; Fig.11c) shows a machine design by the American Oliver Evans (1755-1819), who worked similarly to many others (Watt included) in the Industrializing World to enhance machinery and mechanisms with practical solutions. The drawings and schemes of Fig.11 appear in some extent similar to those of nowadays and if one compares them with previous designs of Figs. 5 to 10 he/she will recognize a great evolution but with main aspects as developed in the past yet.



a) b) c)

Fig.11 Examples of early mechanisms designs during the Industrial Revolution: a) a French early patent in (Gallon 1735); b) a patent by J. Watt in 1777, (Ferguson 1962); b) a machine designed by O. Evans, (Ferguson 1962).

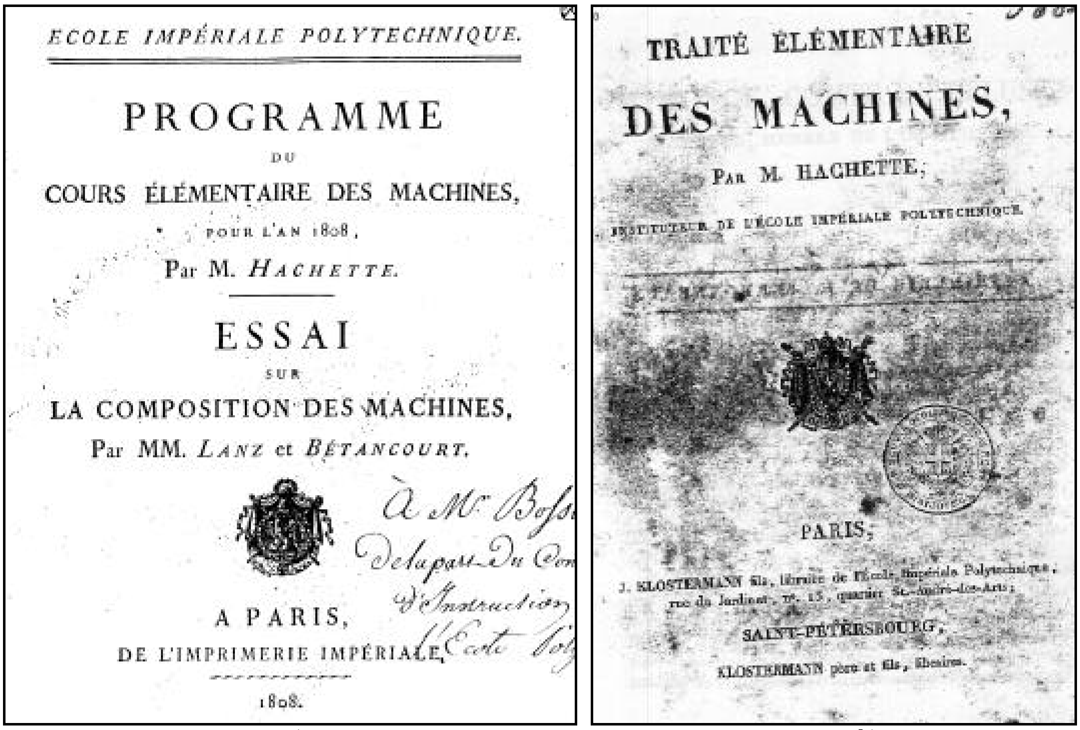
5. The establishment of modern Machine and Mechanism Design

The assessment of the modern TMM and MMD can be dated with the establishment of the Ecole Polytechnique in Paris in 1974. More in particular, Gaspard Monge (1746-1818) can be considered one of founding fathers of modern TMM since he first recognized the need of teaching mechanism analysis and design in specific classes. Since the foundation of the Ecole Polytehnique he proposed the teaching of mechanism theory within his course on Descriptive Geometry since he considered TMM as an application of the Geometry with additional motion characteristics. But he never gave these classes that were taught only starting in 1806 by Jean Nicolas Pierre Hachette (1769-1834), who was his pupil and successor, being the first Academic teacher of a complete course on MMD.

However, it is to note that early textbooks on MMD were written since XVIIIth century, even for a teaching activity, as for examples (Ozanam1720), (Frisi 1777) only to cite some. In these texts, main attention has been focused on general properties of machines, and even to details on mechanism design but from a descriptive viewpoint only.

The ideas for these first classes by Monge were collected and ordered under the supervision of Monge himself and Hachette by two pupils of the Ecole Polytechnique, namely Jose Maria Lanz and Agustin de Betancourt, in the book “Essai sur la Composition des machines”, (Lanz and Betancourt 1808), Fig.12a) that was published in 1808 and can be considered the first text of modern TMM and MMD, as proposed in (Ceccarelli and Cuadrado 1997). Just after, in 1811 Hachette himself published his textbook “Traité elementaire des machines“, (Hachette 1811), Fig.12b), giving the start of modern TMM and a tradition for a course on MMD, as pointed out in (Chasles 1886).

The modern approach of the above-mentioned milestone books on TMM can be recognized in two aspects, namely a first mathematical formulation for analysis and synthesis purposes as an application of geometric and kinematic properties of mechanisms, and a first classification of basic mechanisms with a cataloguing of motion variety. Both approaches can be recognized in Fig.13.


ECOLE IMPÉRIALE POLYTECHNIQUE.

PROGRAMME
DU
COURS ÉLÉMENTAIRE DES MACHINES,
POUR L'AN 1808,
Par M. HACHETTE.

ESSAI
SUR
LA COMPOSITION DES MACHINES,
Par MM. LANZ et BÉTANCOURT.

A PARIS,
DE L'IMPRIMERIE IMPÉRIALE.
1808.

TRAITÉ ÉLÉMENTAIRE
DES MACHINES,
PAR M. HACHETTE,
INSTITUTEUR DE L'ÉCOLE IMPÉRIALE POLYTECHNIQUE.

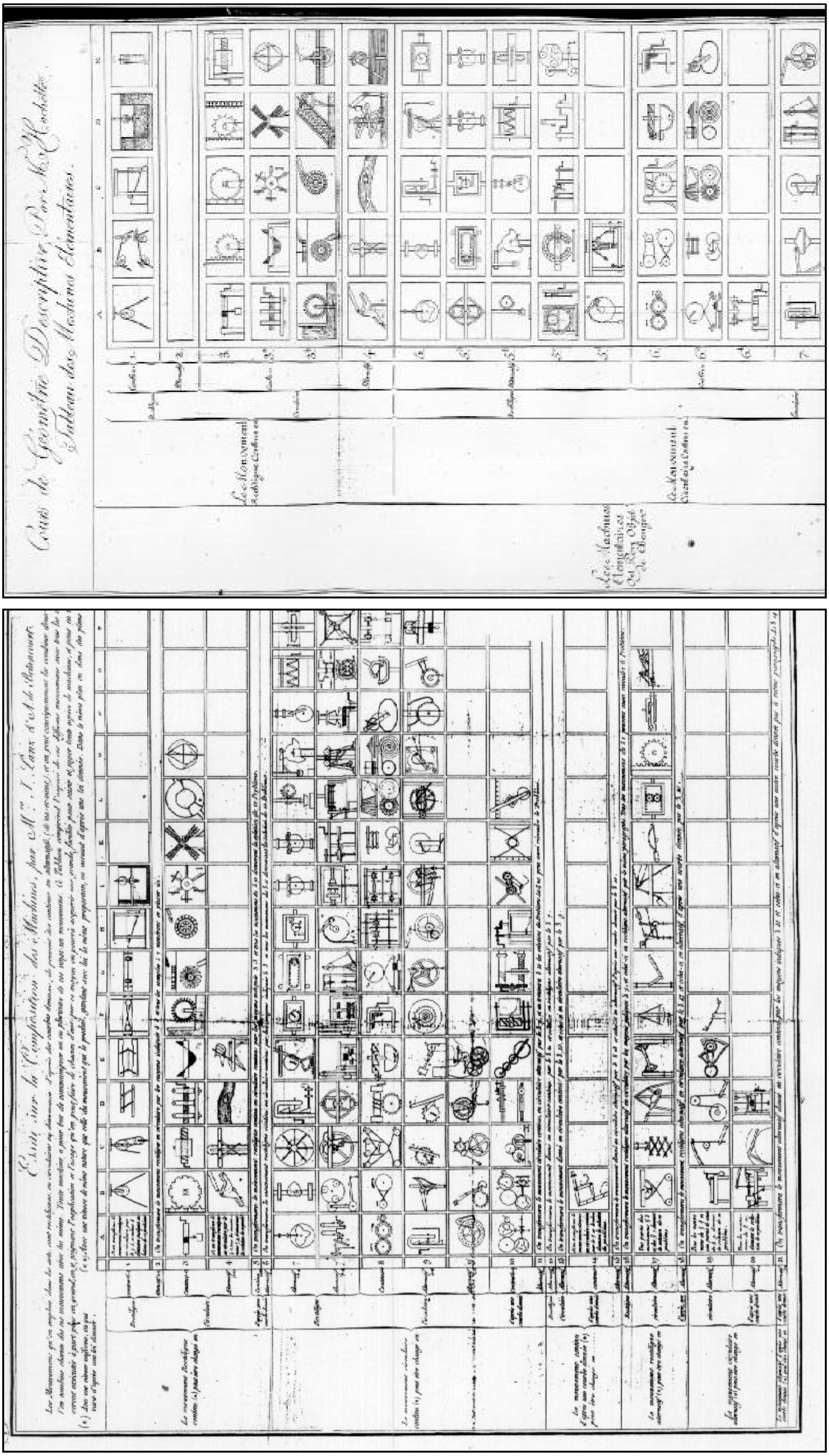
PARIS,
J. KLOSTERMANN fils, libraire de l'École Impériale Polytechnique, rue du Jardinet, n°. 13, quartier St.-André-des-Arts;
SAINT-PÉTERSBOURG,
KLOSTERMANN père et fils, libraires.

a) b)

Fig.12 Title pages of milestones for TMM: a) by J.M. Lanz and A. Betancourt in (1808); by J.N.P. Hachette in (1811).

Figure 13 shows tables of synthetic overview of mechanisms for motion transmission by Lanz and Betancourt (1808), Fig.13a), and by Hachette (1811), Fig.13b), according to the early modern classification that was originally thought by Monge. In both cases it is possible to appreciate an in-depth study of mechanisms in term of mechanical design and kinematic behavior.

The classification of Monge is based on the transformation of movements, which are possible by useful technical means. These movements were classified as the continuos straight-line motion, the intermittent straight-line motion, the continuos circular motion, and the intermittent circular motion. The combination of these motions for giving an output motion with similar motion characteristics gives ten categories of machines, as partially shown in Fig.13. This view for cataloguing mechanisms was very successful and extensively used even for modified approach like for example in the encyclopedic work by Gian Antonio Borgnis (1818-21), who in 9 volumes, written from 1818 to 1821, he treated all the existing machines as also a function of the working purpose in a form of an handbook that was used until the beginning of XXth century. Later, in (1841) Robert Willis (1800-1875) proposed an alternative classification that is based on the velocity ratio and the kind of contacts that are used to transmit and transform the motion. Many other authors have addressed the problem of an exhausting catalogue of mechanisms by combining and/or enlarging the above-mentioned fundamental classifications, and even by attempting new way to describe the mechanisms with different formalism, as the case in (Babbage1826), (Reuleaux 1875), (Masi 1897), and (Koenigs 1905) only to cite the most significant contributions in XIXth century. Still today the topic addresses attention, as documented by the encyclopedic work by I.I. Artobolevsky (1975-80).



a) b)

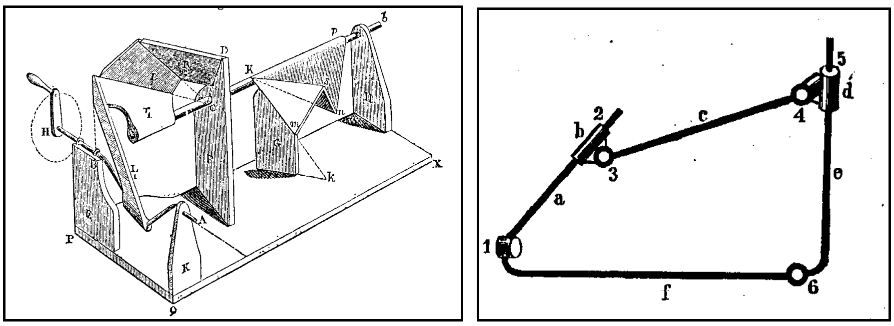
Fig.13 Early modern classifications of machines and mechanisms according to G. Monge: a) a table by J.M. Lanz and A. Betancourt (1808); b) a table by J.N. Hachette (1811).

TMM became more and more important so that a specific field, named as Kinematics of mechanisms, was recognized as an independent discipline and academic courses were started everywhere in Europe (for example, in 1838 at Sorbone University and in 1841 at the early established Technical University of Turin).

The name Kinematics was coined by A.M. Ampere in (1834) to identify the part of Mechanics that deals with the motion of rigid bodies regardless the actions making it. Kinematics of mechanisms can be thought as the application of Kinematics to mechanisms so that it addresses more specifically to subjects of TMM related to procedures for analysis and synthesis of mechanisms. Specific attention was addressed to mechanisms and specific bodies, and it brought to the publication of several fundamental books that today are sometimes forgotten, but with a content of current interest yet. Indeed, they could be subject of specific investigation to re-address attention to the topic and publications yet. Main textbook works of XIXth century can be considered:
from France: (Belanger 1864), (Bobillier 1870), (Bresse 1885), (De La Goullepierre 1864), (Laboulaye 1861), (Mannheim 1880), (Resal 1862), (Villie 1888);
from Germany: (Burmester 1888), (Grashof 1883), (Redtenbacher 1857), (Reuleaux 1875), (Schoenfies 1872);
from Great Britain: (Goodeve 1876), (Kempe 1877), (Kennedy 1886), (Rankine 1887), (Roberts 1875), (Smith 1889), (Willis 1841);
from Italy: (Allievi 1895), (Borgnis 1818-21), (Cavalli 1882), (Giulio 1846), (Masi 1897), (Tessari 1890);
form North America: (Mac Cord 1883), (Hiscox 1899); (Robinson 1896);
from Russia: (Chebyshev 1899).

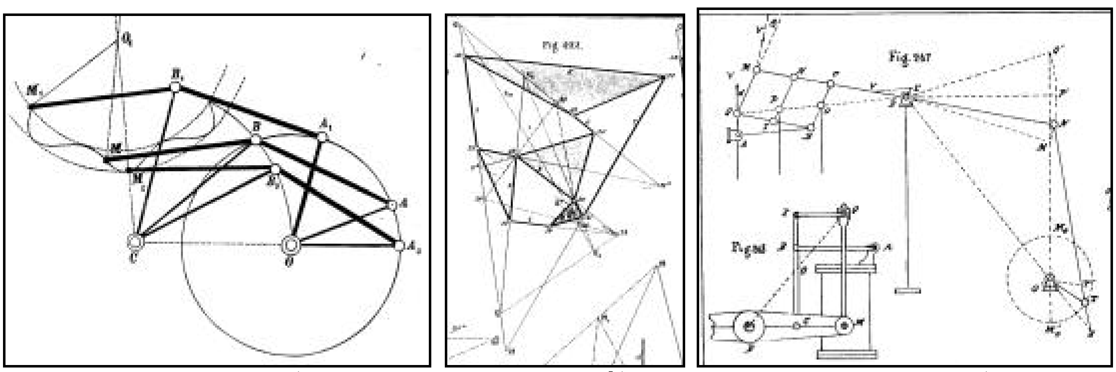
In Figs. 14 and 15 illustrative examples are reported to show how the mechanisms were modeled and analyzed to give formulation useful to teach, design, and investigate the Theory and Practice of mechanisms in the second half of XIXth century. Figures 14 and 15 show illustrative examples of attached problems and kinematic formulation from some very significant authors. Figure 14a) shows a kinematic sketch of a three-dimensional mechanism that Willis studied in some detail, although he did not provide any practical application for it, but he used it as an example to show how the Kinematics of mechanisms could attach new devices. In Fig.14b) a more synthetic sketch is reported from the work by Franz Reuleaux (1829-1905) with the aim to show essential parameters in a kinematic symbolism that we still use today. Figure 15 gives examples of a geometric formulation of mechanism geometry and kinematics for design purposes: in Fig.15a) Pafnutii L'vovich Chebyshev (1821-1894) outlined a computation scheme for the structural error of coupler paths whilst designing a circle-tracing four-bar linkage; in Fig.15b) Ludwig Burmester (1840-1927) drew a geometric construction for a design algorithm as based on the instantaneous properties of mechanisms; in Fig.15c) Francesco Masi sketched the essential operation of a mechanism through a basic kinematic interpretation.

The above-mentioned books were and still are fundamental in the field of TMM and MMD even for attaching problems that are considered only very recently. Further examples of the modernity of the TMM studies in the XIXth century are illustrated in Fig.16 in which one can recognize arguments of kinematics applied to Vision problems, Fig.16a), (Tessari 1880), and to grasping design, Fig.16b), (Masi 1897), as referring to Italian contributions only.



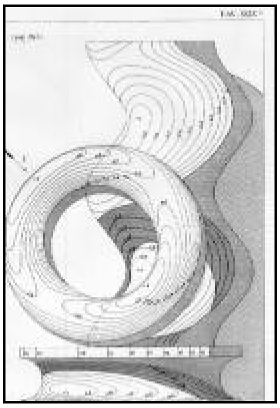
a) b)

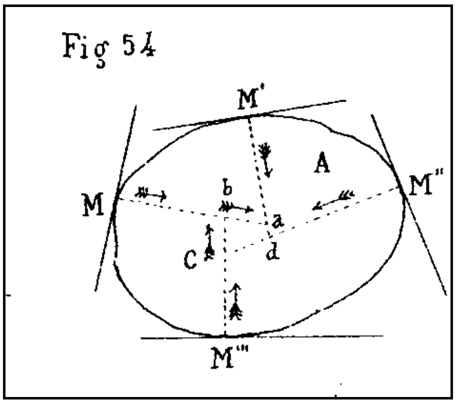
Fig.14 Early modern kinematic schemes for mechanisms: a) by R. Willis in (1841); b) by F. Reuleaux in (1875).



a) b) c)

Fig.15 Examples of early modern schemes for design purposes: a) by P. Chebyshev in (1899); b) by L. Burmester in (1888); c) by F. Masi in (1897).





a) b)

Fig.16 Examples of an early application of Kinematics knowledge: a) to Vision problems as determination of shadows and chiaro-schuro in an object by D. Tessari in (1880); b) to Mechanics of Grasp as analasying stable grasps by F. Masi in (1897).

Unfortunately the Italian development of TMM is not well known abroad, but, similarly to what occurred in other European countries, in Italy there was such as intense activity with very interesting contributions, that should be rediscovered and studied again but with a modern view and formulation, as attempted in (Ceccarelli 2000b).

At the beginning of XXIth century the attention was addressed to more and more complicate mechanisms, with more and more links, and even to spatial mechanisms as successful practical application of the Theory of Mechanisms and numerical developments of Kinematics.

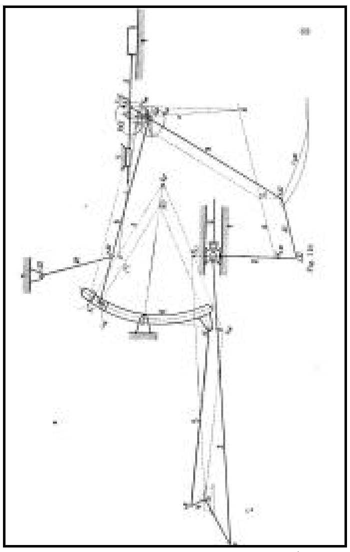
Thus, Fig.17a) shows an example of an in-depth study of the operation of a multi-body system from kinematic viewpoint by drawing of basic configurations of the links and paths of significant points during the motion so that they can be used for a detailed formulation in a mathematical procedure. Figure 17b) shows a sketch of the today so-called Screw Triangle to compute the helicoidal motion in open chain spatial mechanisms in a very similar way to the very recent treatments, as for example in the fundamental work (Roth 1966).

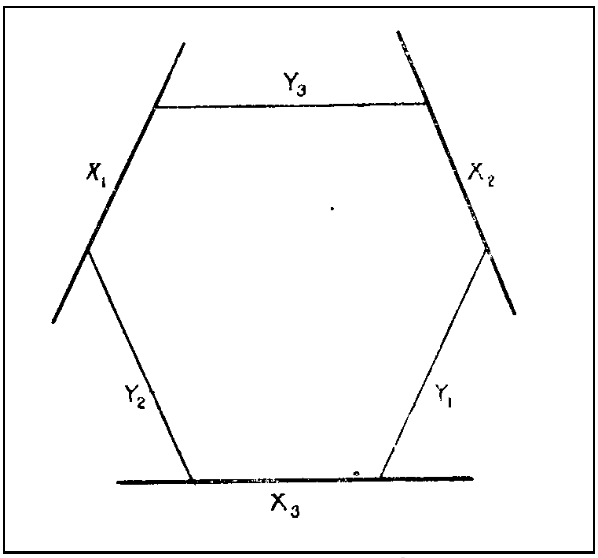
Beside the academic study and design of mechanisms, during the XIXth century an intense activity was experienced for the practical design and application of mechanisms that enhanced the Technology in all aspects.

In Fig.18 few examples are reported to illustrate the fecundity of mechanisms design in different fields of Mechanical Engineering: Fig.18a) gives a mechanical design of looms that were enhanced in automation and high speed operation because of better mechanisms; Fig.18b) shows an automatic sawing machine whose operation whose greatly improved by using better and better mechanisms; Fig.18c) shows an emblematic use of multi-link mechanisms to transmit power in early faster trains; Fig.18d) gives a sketch of an advanced design for typing and calculator machines.

But the great technical evolution started at the beginning of XXth century and many of the mechanical achievements can be considered the results of an intense teaching activity and practice, that were based mainly on the above-mentioned textbooks.

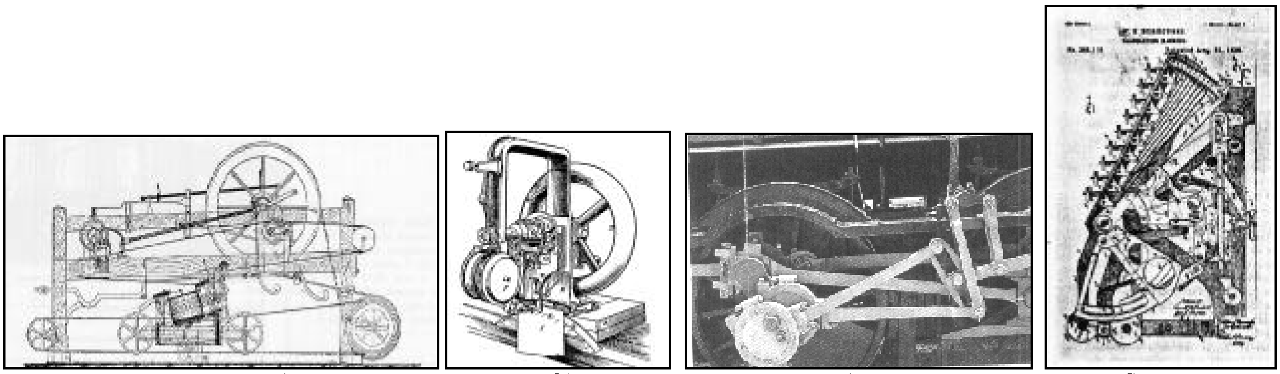
Nevertheless, MMD practice evolved from geometry based procedures to more mathematical formulation. A brilliant example of this can be recognized in the work by Lorenzo Allievi (1895) that can be considered as a re-formulation and mathematical application of much of the geometric theory by Burmester (1888).





a) b)

Fig.17 Early modern studies of complex mechanisms: a) by M. Grubler in (1917); b) by R. Bricard in (1927).



a) b) c) d)

Fig.18 Examples of early modern machines using mechanisms designs in XIXth century: a) in a Compton loom; b) Howe sawing machine for textile manufacturing; c) in power transmission in an early train; d) in a calculator machine.

In addition, many authors have successfully attempted to update the topics of TMM by considering the enhancements of Technology and acquired experiences of new mechanisms. They also gave contributions with novel analysis and design procedures. Example of this intense acticity are the works (Bourguignon 1906), (Crelier 1991), (Franke and Oldenbourg 1930), (Bricard 1927), (Grubler 1917), only to cite some.

But one can note a lack of significant publications (or at least not recognized outside each country) during and in between the two Worlds Wars. This is probably because most of the technical knowledge was kept secret for military reasons.

However, just after the 1950's a renewed interest on MMD arose mainly for the optimism addressed on the coming computers. Thus, the Kinematics of mechanisms was revised in view of using computer calculation. Relevant works can be considered those by Rosenauer and Willis (1953), Beggs (1955), Hall (1961), Hain (1967), Hartenberg and Denavit (1964), that can be also considered an expression of the coming leadership of North America in MMD by using the European tradition.

Thus, in the 1970's the practice of MMD has been established in a modern view and approach. Everywhere in the world Universities give courses on Kinematics of Mechanisms and MMD, by referring to texbooks written by local experts and professors even in local language so that a huge bibliography has been published since then.

6. Current practices and investigations

The current practices and investigations on MMD can be considered as based on the use of computers. Therefore, they are mainly oriented to development of procedures for analysis and synthesis rather than to new designs. Because of the huge number of peoples working in the field and different ways for attaching mechanism problems, it is rather impossible to indicate exhaustively all the specific subjects and applications, which the work and investigation are currently addressed to.

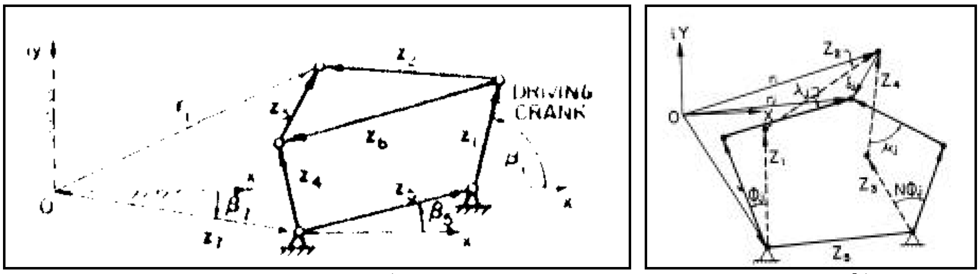
Nevertheless, main arguments of current interest can be identified from the Proceedings of the IFToMM World Congress, (Rovetta Ed.1995; Leinonen Ed.1999), and ASME Biennial Mechanisms Conference, (ASME Ed. 1998 and 2000), and they can be listed as: Kinematic and dynamic analysis of mechanisms (cams, planar linkages, spherical mechanisms, spatial mechanisms, geared mechanisms); dynamic analysis of multibody systems; analysis of mechanisms with higher pairs; analysis of mechanisms with flexible links; performance analysis; micro-mechanics; kinematic synthesis of mechanisms (by means of numerical or symbolic formulation, or through optimization problems); analysis of solutions for mechanisms synthesis or inverse kinematics; workspace of mechanisms; experimental evaluation of mechanisms; history and teaching of Theory of mechanisms; mobility analysis of mechanisms; analysis and synthesis of type of mechanisms; kinematic formulation of the motion of rigid bodies; vibration analysis; analysis and synthesis of robotic systems; design methodologies; path planning; biomechanics; measurements in mechanisms; theoretical mechanics; robotics in medical applications; safety and reliability of mechanisms; high-speed and high-accuracy mechanisms and robots; redundant manipulators; overconstrained mechanisms; kinematics for CAD/CAM; error analysis; topology of mechanisms; man-machine systems; walking machines; mobile robots; mechanics of grasping; grasping systems; identification and diagnostics; bearings; balancing; industrial applications; control of mechanisms; machine elements; micro- mechanisms; mechatronics computational kinematics; software development; teaching methods; experimental methods; micro-electrico mechanical systems; durability; parallel mechanisms; compliant mechanisms; design optimization; design modeling; chains and belts; rapid prototyping of mechanisms; virtual prototyping; Inverse kinematics; web-based design; conceptual design; crash analysis; tolerance analysis; concurrent design; CAD in education; design decision theory; spring mechanisms; synthesis of complex mechanisms.

Most of the attention is addressed to formulate procedures and problems (mainly theoretical ones, for analysis or design aims) in a general way but for numerical simulation by using computers. Even the attempt of deriving closed-form expressions is directed for an use of computer algorithms, in symbolic or numerical procedures.

Emblematically, the use and development of computer algorithms have required and still requires a suitable modeling of mechanisms that in some extent is even more schematic than in the past, since only very basic characteristics are need for formulating efficient computations. Further mechanism properties can be computed by suitable numerical parameterization or additional simple computations so that a user/designer may even not have a very advanced knowledge in Kinematics.

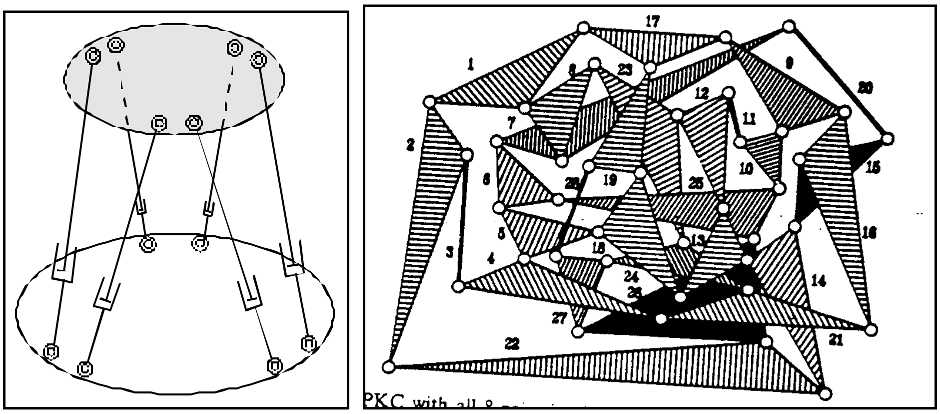
Illustrative examples are illustrated in Fig.19 referring to two early works (but still of current validity) that can represent the beginning of the current computer oriented approach to mechanisms. In Fig.19a), (Freudenstein and Sandor 1959) the four-bar linkage has been sketched through suitable vectors that were used for a formulation that can be considered the first work for computer computation in mechanism design. The model of Fig.19b), (Roth and Freudenstein 1963), has been developed for investigating numerical aspects of the computer computation for mechanisms. In fact, the use of computers has renewed interest on the modeling, formulation, and design of mechanisms with new and old solutions and approaches at the same time.

The great power of computation of modern means has given great stimulus to study complex and multi-link kinematic architectures in such a way that new mechanisms have been of current interest and even rapidly applied in machines. Emblematic is the case of parallel manipulators, whose general chain is represented in Fig.20a) as a Gough-Stewart Platform. The parallel manipulator architecture was first applied in some novel industrial applications with the pioneer work by Stewart (1965), but only in the late 80's it has addressed considerable attention. Numerical investigations have given further industrial implementation in robotic field, although parallel manipulators are still intensively investigated both for computation purposes and prototyping research, even by considering hybrid chain as a combination of parallel and serial chains.



a) b)

Fig.19 Early schemes for mechanism design by means of computer oriented algorithms: a) by F. Freudenstein and G. Sandor in (1959); b) by B. Roth and F. Freudenstein in (1963).



a) b)

Fig.20 Examples of complex mechanisms of current interest: a) the paralell architecture of Gouh-Stewart Platform; b) a 28-link 1 dof planar mechanism.

Similarly, more and more complex mechanisms have been attracted the interest of kinematicians, even sometimes to solve very intriguing mechanisms as the case of Fig.20b), which has been studied mainly for numerical kinematic purposes and no practical applications.

A novel field for MMD has been recently identified as Mechatronics (but not so new if you considers the approach for Fig.2a) !) to study a mechanism as a multi-body system as a whole together with the other components of a modern machine, like actuators, sensors, and control equipment, by considering contemporaneously all the aspects of operation and systems of the different components. However, the mechatronic approach for mechanisms considers Kinematics and MMD as a fundamental part of the investigation for a successful design and practice of mechatronic systems.

Specific continuos conference activity on TMM has started only in 1950's, as documented by the First Conference on Mechanisms held at Purdue University in October 1953, (Machine Design Ed. 1953).

Since the modern re-establishment of the Kinematics of machinery in the 1960's, even the teaching aspects have addressed specific attention for the determination of suitable curriculum of study for engineers, as documented by the first International Conference for Teachers of Mechanisms, (Crossley 1961).

As already mentioned in the Introduction, the field of TMM has been universally established with the foundation of IFToMM, and several conferences are today organized and successfully held in the many aspects of MMS and particularly MMD in national and international frames. Even series have been established and it is noteworthy to list the last events of the main international activity: IFToMM World Congress (Leinonen Ed. 1999); ASME Biennial Mechanisms Conference (ASME Ed. 2000); Romansy Symposium (Morecki and Bianchi Eds. 2000); ARK Worshop (Lenarcic and Husty Eds. 2000); CK Workshop (Park and Iurascu Eds. 2001). Although they can be considered specific conferences on MMD, they include topics from many other fields, and mainly Robotics can be recognized as a fecund area for application of MMS and specifically MMD. Thus, in many activities in Robotics and Automation one can find subjects of MMS and MMD, both from theoretical and practical results.

The conference activity can be considered very important, beside the Journal publications, since at a conference one can present even preliminary results of her/his work; one can exchange experiences and opinions (but now e-mail communication is more and more essential and efficient); one can establish personal contacts. In fact, the practice of the conference activity has brought enlargement of the MMS community but even it is a significant source for innovation and design from viewpoint of both inspiration and presentation of new designs and studies.

Very recently laboratory experimental activity is again of basic importance, after that since 1960's the computer simulations have given the possibility to avoid main of the preliminary laborious experimental validations and even the presumption that experimental activity cannot be necessary. But prototyping validation is often required when new solutions and new operations are conceived and models are not well tested with past experiences. In addition, laboratory experimental activity can be considered yet as an important formation phase for young engineers, but even for expert investigators in the field of MMD.

Summarizing, the current activities in MMD are mainly focused on computer oriented formulation and procedures for analysis, design, and practice purposes. But experimental activity is now emerging with more and more attention mainly for investigations involving new aspects and designs, high-speed operation, and micro-scaled systems.

7. Challenges for future activity

Preoccupation for the future of Mechanisms Design has been raised recently, mainly after that in some countries the University formation has been revised with a considerable reduction of attention and teaching space for the fields of Kinematics and consequently Mechanisms Design.

Emblematic is the case of U.S.A. where the study of Kinematics and MMS has been strongly reduced in the University curricula and even research funding has been reduced since MMD has been not considered as strategic as in the past for the advance of machinery, as pointed out in the specific report to NSF (Shaha 1996). Therefore, there is a great attention and even several efforts are made to re-address attention to the field as illustrated in several actions of ASME, like the publication (ASME International 1998).

However, since the beginning of TMM in XIXth century the teaching of mechanisms has been an object of discussion and it has been re-shaped several times with updated means and goals in agreement with the re-formulation of the curriculum of study for mechanical engineers. Nevertheless, a certain considerable basis has been definitively established after the pioneering teaching at the early XIXth century and a common teaching of TMM and MMD has been persistent so far in the Schools of Engineering, as it is documented by the several textbooks published around the world.

Since 1950's the emerging new Technologies have also speeded up the process of a re-thinking on the teaching of MMS and MMD. Emblematic is the need of even conference activity that is carried out at national and international levels. Historically significant is the International Conference for Teachers of Mechanisms, (Crossley 1961). In the Proceedings one can read most of the arguments of the todays discussions for the future of teaching MMD !

The teaching activity can be considered of primary interest and importance for the future development and successful practice of a discipline since it gives proper rational means and rigorous formation to the peoples, but is also a need to produce peoples and personalities, who will enhance the practice and research of a field.

Indeed, the possible lack of future success of MMD field can be also due to the fact that the pure teaching aim of teachers has been overpassed by the more and more practice of teachers in applications design. This is to say that the primary goal of University (Institutions for formation) should be, like in the past, the formation of engineers and not the application of the research results only. The research activity can be considered also as fundamental for the formation of good teachers, who, because of their research activity, are aware of the current issues and have in-depth experiences of the engineering practices so that they can teach a live subject. Nevertheless, it should be taken into account that the teaching practice and means should be updated to the modern needs and culture.

Beside the teaching challenge, one can address importance for the future MMD mainly in the following topics:

- efficient computation of analysis and synthesis algorithms
- high-speed light mechanisms
- multi-dof spatial mechanisms
- integration of mechanisms design with design procedures of other components of a (mechatronic) system
- miniaturization
- easy operated low-cost advanced mechanisms
- rapid prototyping or even no prototyping

The computational efficiency has been recognized as a fundamental feature of the designing activity since the first uses of computers in engineering activity. The greater and greater power of computation of the computing systems permits but also requires to deepen the computation skill of designers in order to achieve better and better simulation and design procedures. This the fact that today but even more and more in the future a mechanical engineer must have a basic knowledge of computing systems and an expertise in computational procedures in term of both numerical algorithms and programming techniques.

The demand of higher and higher speed productions requires and will require more and more the use of high-speed mechanisms for automatic and non-automatic machines. A high speed mechanism can be designed wit a light weight design but even with a particular attention to the dynamics for the high-speed operation. Both the issues can be achieved by designing new mechanisms but also by a better analysis of the Kinematics and Dynamics of high-speed operation of mechanisms. These topics will require also an intense experimental activity, as already started recently.

Great enhancements are expected in many fields by using more complex mechanisms and/or multi-dof (degree of freedom) mechanisms in order to perform more demanding tasks. Thus, it can be thought that spatial mechanisms and multi-dof mechanisms will be more extensively applied. This demand requires also practical expertise in using spatial mechanisms and several actuators contemporaneously.

But more than the above-mentioned topics, the mechatronic practice will be asked to engineers, who should manage the integration of mechanical systems with the essential counterparts with electric and electronic natures. This is because the success and efficiency of a mechatronic system strongly depends of a full understanding of the mechatronic design and operation. Consequently, the pure mechanical design will be overpassed by design procedures that take into account other design characteristics from other engineering disciplines. This will require also to rethink and re-formulate the formation of engineers with an in-depth study of other disciplines.

In XIXth century one challenge was to design and build large scale system, much powerful and larger than the first industrial machines of XVIIth century. In previous centuries the mechanical systems were conceived mainly with the size of human beings since they were aimed to help and even substitute human workers. But in XXIth century, as already started at the end of the XXth century, a great new demand of miniaturized mechanical systems of a size of few millimeters and even less (as for nano-machines) will be asked to designers. Micro-systems and even more nano-systems will require to rethink and re-investigate on mechanisms since their behavior at the micro/nano scale is affected by other phenomena than gravity and friction only. The manufacturing and operation cannot be obtained with the accuracy of normal sized systems and usual performances that involve the consideration of phenomena of micro-scale with a specific novel study of Kinematics and Dynamics. Important experimental activity is expected and already undergoing in several laboratories around the world in order to get the necessary practice for working with micro/nano systems as well as to deepen the knowledge of the phenomena and operation in industrial oriented designs.

Although the greater and greater demand of innovative designs, a basic aim of the future activity of mechanical engineers can be considered to be related with the development of low-cost systems but with easy operation in order to have wider and wider application of the systems, even by users who are not technical experts. These issues require an in-depth rethink of many designs and even to reformulate the design and operation of well established systems, but the components too.

Thus, a fundamental future issue for MMD can be considered the updating of MMD itself in its many aspects but even in the basic philosophy of designing activity.

Finally, the most important challenge for MMD in the future can be considered to have capability to attach problems and develop solutions in shorter and shorter time in new and new fields of application that will be defined suddenly and suddenly in the future, probably without the chance that we have today to study and practice with prototypes in research centers. A significant consequence can be advised in the fact that approximated solutions and mainly simple few-link mechanisms will have again great interest and application.

Summarizing, the fundamental challenge for the future of MMD can be recognized in the capacity and flexibility to update and adjust the knowledge and expertise to new updating design requirements and systems.

8. Conclusions

Machine and Mechanism Design MMD is an important field of Machine and Mechanism Science MMS, formerly known as TMM (Theory of Machines and Mechanisms). The historical development of Mechanical Design has occurred with deeper and deeper attention to Mechanisms Design, but only at the beginning of XIXth century TMM has been established as an independent mature discipline. In fact, MMD have had a great role in Mechanical Engineering in the last two century as outlined in this paper through few significant examples. Indeed MMD has been enhanced to a so high level that new advances seem to be difficult to achieve. Therefore, a certain crisis of identity for the future is today felt by the MMD community. Nevertheless, the past experiences give a certain optimism, also because human beings will always live in a mechanical environment so that mechanical systems will be always required and designed but with new and new requirements and applications.

9. Acknowledgement

The author wishes to thank the library of the Department of Mechanics and Aeronautics of "La Sapienza" University of Rome (Italy), the library of the Department of Mathematics of Politecnico di Milano (Italy), the library of

Politecnico di Torino (Italy), the University library of Bologna (Italy), the University library of Padua (Italy), the University library of Pisa (Italy), the University library "Boaga" of Roma (Italy), the Library of the "Castelnuovo" School of Mathematics at the La Sapienza University of Roma (Italy), the University library of Torino (Italy), the Apostolic Library in Vatican, and Library of Montecassino Abbey for the help in the search of original material.

10. References

Allievi L, 1895, "Cinematica della biella piana", Regia Tipografia Francesco Giannini & Figli, Napoli.

Ampere A.M. , 1834, "Essai sur la philosophie des sciences", Paris.

Angeles J., 1997, "A Fin-de-Siecle View of TMM", Proc. of Int. Conference on Mechanical Transmissions and Mechanisms, Tianjin.

Artobolevsky I.I., 1975-80, "Mechanisms in Modern Engineering", Mir Publ. Moscow, 5 Vols.

ASME International, 1998, "Focus on the Future", ASME Brochure.

ASME (Ed.), 1998, 26^{5h} Biennial Mechanisms Conference, in CD-Rom Proceedings of 1998 ASME Design Engineering Technical Conferences, Sept. 1998, Atlanta.

ASME (Ed.), 2000, 26^{th} Biennial Mechanisms and Robotics Conference, in CD-Rom Proceedings of 2000 ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Sept. 2000, Baltimore.

Babbage C., 1826,"On a method of expressing by signs the action of machinery", Philosophical transactions of the Royal Society, London, vol. 116, pp. 250-265.

Barbaro D. , 1584, "I Dieci Libri dell'Architettura di M. Vitruvio", Venezia.

Bechmann R. , 1991, "Villard de Honnecourt, Le pensée technique au XIII^{e} siècle et sa communications", Paris.

Beggs J.S. , 1955, "Mechanisms", Mc Graw Hill, New York.

Belanger, 1864, "Traitè de Cinèmatique", Paris.

Besson J. , 1578, "Théatre des instruments Mathématiques et Mécaniques...", Lion.

Bobillier E.E., 1870, "Cours de Geometrie", Paris.

Borgnis G.A. , 1818-21, "Traitè complet de mecanique appliquée aux arts", Bachelier, Paris, 9 Vols.

Bourguignon P. , 1906, "Cours de Cinematique Theorique et Appliquée -II. Cinematique Appliquée", Paris.

Branca G. , 1629, "Le machine", Roma.

Bresse J.A., 1885, "Cours de Mecanique et Machines", Ecole Polytecnique, Paris.

Bricard R., 1927, " Lecons de Cinematique", Gauthier-Villars, Paris. 2 Vols.

Brunelleschi F., 1420, "Zibaldone", Firenze.

Burmester L. , 1888, "Lehrbuch der Kinematik", Leipzig.

Cavalli E. , 1882, Elementi di cinematica teorica - ad uso delle scuole di applicazione per gli ingegneri, Hoepli, Milano.

Ceccarelli M., 1998, "Mechanism Schemes in Teaching: A Historical Overview", ASME Journal of Mechanical Design, Vol.120, pp.533-541.

Ceccarelli M., 1999a, "Cinematica della Biella Piana by Lorenzo Allievi in 1895", Xth IFToMM World Congress on Theory of Mechanisms and Machines, Oulu, Vol.1 , pp. 37-42.

Ceccarelli M., 1999b, "On the meaning of TMM over time", Bulletin IFToMM Newsletter, Vol.8. Nr.1.

Ceccarelli M., 2000a, "Screw Axis defined by Giulio Mozzi in 1763 and Early Studies on Helicoidal Motion", Mechanism and Machine Theory, Vol.35, pp.761-770.

Ceccarelli M. , 2000b, "Italian Kinematic Studies in XIXth Century", International Symposium on History of Machines and Mechanisms - Proceedings of HMM2000, Kluwer, Dordrecht, pp.197-206.

Ceccarelli M., 2000c, "An Overview of History of Robotics with an Eye to the Future", 2000 IEEE International Conference on Intelligent Engineering Systems, Portoroz, Plenary Paper 4, pp.25-33.

Ceccarelli M., 2000d, "Preliminary Studies to Screw Theory in XVIIth Century", Ball Conference, Cambridge, CD Proceedings, July 2000.

Ceccarelli M. (Ed.), 2000, International Symposium on History of Machines and Mechanisms - Proceedings of HMM2000, Kluwer, Dordrecht.

Ceccarelli M., Cuadrado I., 1997, "Sobre el Essai sur la Composition des Machines por Jose Maria De Lanz y Augustin de Betancourt en 1808", 3° Congresso Iberoamericano di Ingegneria Meccanica, CD Proceedings, La Habana.

Chasles M., 1837, "Apercu historique sur l'origin et le développement des méthodes en géométrie ...", Mémoires couronnés par l'Académie de Bruxelles, Vol.11. (2^{nd} Ed., Paris, 1875).

Chasles M., 1886, "Exposé historique concernant le cours de machines dans l'enseignement de l'Ecole Polytechinique", Gauthier-Villars, Paris.

Chebyshev P.L., 1899, "Ouvres de P.L. Tchebychef", Imp. de la Academie de Sciences, St. Petersbourg.

Cianchi M. , 1984, "Le Macchine di Leonardo", Becocci Publ., Firenze.

Crelier L., 1911, "Systemes Cinèmatì ques", Paris.

Crossley E.F.R. (Ed.), 1961, " Proc. of the Int. Conference for Teachers of Mechanisms", Yale University.

Crossley E.F.R., 1988, "Recollections from Forty Years of Teaching Mechanisms", ASME Jnl of Mechanisms, Transmissions and Automation in Design, Vol.110, pp.232-242.

D'Alembert J.B., Diderot D., 1774, "Recueil de Planches, sur les Sciences, les Arts Liberaux, et les Arts Mechaniques",

3° Edition, Livourne.
De Groot J. , 1970, "Bibliography on Kinematics", Eindhoven University, Eindhoven.
De la Goullepierre H. , 1864, "Traité des mecanì smes", Paris.
De Jonge A.E.R., 1943, “A Brief Account of Modern Kinematics”, Transactions of the ASME, August, pp.663-683.
De La Hire P., 1695, “Traitè de Mecanique”, Paris.
De La Hire P., 1706, “Traitè des Roulettes”,in Memoires de l’Academié des Sciences, Paris.
Del Monte G. , 1577, "Mechanicorum liber", Pesaro.
Dimarogonas A.D., 1993, “The Origins of the Theory of Machines and Mechanisms”, in Modern Kinematics – Developments in the Last Forty Years Ed. by A.G. Erdman, Wiley, New York, pp.3-18.
Dubowsky S., 1993, “The Future Role of the Science of Mechanisms in Design”, in Modern Kinematics – Developments in the Last Forty Years Ed. by A.G. Erdman, Wiley, New York, pp.575-.
Ferguson E.S., 1962, “Kinematics of Mechanisms from the Time of Watt”, Contributions from the Museum of History and Technology, Washington, paper 27, pp. 186-230.
Franke R., Oldenbourg R. , 1930, "Eine Vergleichende Schalt und Getriebelehre - NeuWege der Kinematik", Munich.
Freudenstein F., 1993, “Critical Research and Unsolved Problems in Kinematics and Related Areas”, in Modern Kinematics – Developments in the Last Forty Years Ed. by A.G. Erdman, Wiley, New York, pp.585-588.
Freudenstein F., Sandor G.N., 1959, “Synthesis of Path-Generating Mechanisms by Means of a Programmed Digigtal Computer”, ASME jnl of Engineering for Industry, Vol.81, pp.159-168.
Frisi P. , 1777, Instituzioni di Meccanica, d'Idrostatica, d'Idrometria e dell'Architettura Statica, e Idraulica, Galeazzi Regio Stampatore, Milan.
Frommel C.L. , 1994, "The architectural drawings of Antonio da Sangallo the younger and his circle", The MIT Press, New York.
Galilei G., 1593, “ Le Meccaniche”, reprinted in Opere di Galileo Galilei Edited by F. Brunetti, Torino, 1964.
M. Gallon (Ed.) , 1735, Machines et Inventions approuvèes par l’Academies Royale de Science, Tome Premier – Depuis 1666 jusqu’en 1701, Paris.
Garcia de Jalon J., 1997, “La Ingenieria Mecànica en la Sociedad del Conocimiento”, Keynote lecture, Spanish National Congress on Mechanical Engineering, Bilbao.
Giulio C.I. , 1846, "Sunti delle Lezioni di Meccanica applicata alle arti", Tipografia Pomba, Torino.
Goodeve T.M., 1876, “The Elements of Mechanism”, London.
Grandi G. , 1739, "Instituzioni Meccaniche", Firenze.
Grashof F. , 1883, "Theorie der Getriebe und der Mechanischen Messinstrumente", Hamburg.
Grubler M. , 1917, "Getriebelehre", Springer, Berlin.
Kempe A.B., 1877, “How to Draw a Straight Line”, London.
Kennedy A.B.W., 1886, “The Mechanics of Machinery”, MacMillan, London.
Kestell T.A., 1963, “Evolution and Design of Machinery Primarly Used in the Manufacture of Boots and Shoes”, Proc. Institution of Mechanical Engineers, 1963-64, Vol.178, Pt 1, n.24, pp.625-683.
Koetsier T., 1997, “On the Prehistory of Programmable Machines: Musical Automata, Looms, Calculators”, International Workshop on Robotics in Alpe-Adria-Danube Region RAAD’97, Cassino, pp. 65-70.
Koenigs G., 1905, “Introduction a une Théorie Nouvelle des Mécanismes”, Librarie Herman, Paris.
Hachette J.N.P. , 1811, "Traitè elementaire des machines", Paris.
Hain K., , 1967, "Applied Kinematics", McGraw-Hill, New York.
Hall A.S.jr, 1961, "Kinematics and Linkage Design", Prentece-Hall, Englewood Cliffs.
Hartenberg R.S. and Denavit J. , 1964, "Kinematic Synthesis of Linkages", Mc Graw-Hill, New York.
Herrad von Landsberg, 1899, "Hortus deliciarum: texte explicatif commenc pour le chanoine A, Straub", Strasburg.
Hiscox G.D., 1899, “Mechanical Movements”, NewYork.
Hunt K.H. , 1982, "Kinematic Geometry of Mechanisms", Springer-Verlag, Berlin.
IFToMM, 1991, “IFToMM Commission A. Standard for Terminology”, Mechanism and Machine Theory, Vol.26, n.5.
IFToMM, 2000a, “Minutes of the 2000 Meeting of the IFToMM Executive Council”, IFToMM Archive, CISM, Udine.
IFToMM, 2000b, “IFToMM Constitution and By-Laws 2000”, IFToMM web page: www.cim/mcgill.ca/~iftomm.
Laboulaye C., 1861, “Traitè de Cinèmatique ou theoriè des mecanismes”, Paris.
Lanz J.M. and Betancourt A. , 1808, "Essai sur la composition des machines", Paris.
Leinonen T. (Ed.), 1999, "Proceedings of Xth IFToMM World Congress on the Theory of Machines and Mechanisms", Oulu, 7 Vols.
Lenarcic J. And Husty M. (Eds.), 2000, Advances in Robot Kinematics, Kluwer, Dordrecht. (held in Portroz)
Leupold J., 1724, “Theatrum Machinarum”, Leipzig.
Machine Design (Ed.), 1953, Transactions of the First Conference on Mechanisms, Purdue University, October 12-13, 1953, Machine Design, Vol.25, dec 1953, pp.173-220 (7 papers).
MacCord C.W., 1883, “Kinematics”, Newyork.
Mannheim A., 1880, “Cours de Gèometrie Descriptive Contenant les èlements de la Gèomètrie Cinèmatique, Paris.
Marchis V. , 1994, "Storia delle Macchine - Tre millenni di cultura tecnologica", Ed. Laterza, Milano. (in Italian).
Mariano di Jacopo (il Taccola), 1969, "Liber tertius de ingeniis ac aedifitiis non usitatis", edited by J.H. Beck, Milan.

Masi F. , 1897, "La teoria dei meccanismi", Bologna.

Morecki A. and Bianchi G. (Eds.), 2000, 13th CISM-IFToMM Symposium on Theory and Practice of Robots and Manipulators Ro.Man.Sy.'2000, Springer-Verlag, Wien. (held in Zakopane, Poland)

Mozzi G. , 1763, “Discorso matematico sopra il rotamento momentaneo dei corpi”, Stamp. di Donato Campo, Napoli.

Nolle H., 1974, "Linkage Coupler Curve Synthesis: A Historical Review –I and II" , IFToMM Journal Mechanism and Machine Theory, Vol.9, n.2, pp.147-168 and pp.325-348.

Ozanam M. , 1720, “La Mechanique", Paris.

Park F.C. and Iurascu C.C. (Eds.), 2001, 2nd Workshop on Computational Kinematics CK2001, Seoul.

Ramelli A., 1588, "Le diverse et artificiose machine", Paris.

Rankine M.W.J., 1887, “Manual of Machinery and Millwork”, London.

Redtenbacher, 1857, “Die Bevegungs Mechanismen”, Heidelberg.

Resal H., 1862, “Traitè de Cinèmatique”, Paris.

Reuleaux F., 1875, “Theoretische Kinematic”, Braunschweig.

Roberts S., 1875, “On Three Bar Motion in Plane Space” Proc. London Math. Soc., Vol.7, pp. 14-23.

Roberval P., 1730, "Observation sur la composition des mouvements, et sur le moyen de trouver les touchantes del lignes courbes", in “Memoires de l'Academie Royale des Sciences. Depuis 1666 jusq'a' 1699”, Paris.

Robinson S.W., 1896, “ Principles of Mechanisms, NewYork.

Rosenauer N., Willis A.H. , 1953, "Kinematics of Mechanisms", Sidney.

Rosheim M.E.., 1994, “Robot Evolution”, Wiley, New York.

Roth B., 1966, “ On the Screw Axes and Other Special Lines Associated with Spatial Displacements of a Rigid Body”, ASME jnl of Engineering for Industry, Vol.88, pp.1-9.

Roth B., 1983, “Robots – State of Art in Regard to Mechanisms Theory”, ASME Jnl of Mechanisms, Transmissions, and Automation in Design, Vol.105, pp.11-12.

Roth B., Freudenstein F., 1963, “ Synthesis of Path-Generating Mechanisms by Numerical Methods”, ASME jnl of Engineering for Industry, Vol.81, pp.1-7.

Rovetta A. (Ed.), 1995, "Proceedings of Xth IFToMM World Congress on the Theory of Machines and Mechanisms", Milan, 4 Vols.

Ruggieri G., 1997, “La Teoria dei Meccanismi”, Keynote Lecture, Meeting of TMM Italian Group, Taormina.

Schoenfies A., 1872, “Geometrie der Bevegung”, Leipzig.

Schott G. , 1664, "Technica Curiosa", Norimberga.

Shah J.J. (Ed.), 1996, “Research Opportunities in Engineering Design – Final Report to NSF”, NSF Strategic Planning Workshop, ASME DETC, Irvine.

Smith R.H., 1889, “Graphics or the Art of Calculating by Drawing Lines”, London.

Stewart D., 1965, "A Platform with Six Degrees of Freedom", *Proc. of the Inst. of Mech. Eng.*, London, Vol.180 (1965), pp.371-386.

Tessari D., 1880, “La Teoria delle Ombre e del Chiaro-Scuro”, Tip. Camilla e Bertolero, Torino.

Tessari D. , 1890, La cinematica applicata alle macchine, Loescher, Torino.

Wallis J., 1670, “Mechanica:sive De Motu, Tractatus Geometricus”, London.

Varronis M., 1584, “De Motu Tractatus”, Iacobi Stoer, Geneve.

Villiè E., 1888, “Traitè de Cinèmatique”, Paris.

Vitruvius P. M. , 1511, "De architectura" edited by Fra Giocondo, Verona, (reprinted in 1513, 1522 and 1523).

Willis R., 1841, "Principle of Mechanisms", 1870- 2nd Ed., London.

Yan H.-S.,1999, “A Design of Ancient China’s Cattle Machine”, Xth World Congress on Theory of Machines and Mechanisms, Oulu, Vol. 1, pp 57-62.

OPTIMIZATION OPPORTUNITIES IN THE BRAZILIAN AEROSPACE INDUSTRY

Hugo Borelli Resende
EMBRAER
hresende@embraer.com.br

Optimization has been for a long time considered as a very important activity of the Product Development process. In engineering, this is generally easier stated than performed: despite the huge amount of knowledge accumulated throughout the last decades, particularly in the structural field, the practical use of optimization requires both skilled people and robust, reliable optimization tools/methods, in order to obtain a design that fulfills the requirements specified for the product. Nevertheless, these resources are not always available at the right time and place during the development process. Thus the need for a more systematic approach involving optimization in the Product Development Process: its insertion as early as possible in the activities, including here even the formulation of the requirements, its extensive application through the conceptual and detailed design phases, and the understanding, particularly from a managerial point of view, of its essential role in the overall process. This approach is illustrated through the description of some opportunities of application of optimization in the Brazilian aerospace industry.

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

ON THE THERMAL ANALYSIS OF MANUFACTURING PROCESSES

Ranga Komanduri and Zhen Bing Hou
Mechanical & Aerospace Engineering
Oklahoma State University
Stillwater, OK 74078, U.S.A.
Phone: (405) 744-5900; Fax: (405) 744-7873; and e-mail: ranga@ceat.okstate.edu

Abstract. *The authors have recently addressed thermal aspects of various manufacturing processes, such as arc welding [1,2], laser surface transformation hardening of gears [3], metal cutting [4-6], polishing [7-9], and high-speed machining [10] as well as tribological applications, such as sleeve bearing [11], and sliding contact [12] using Jaeger's classical heat source method [13]. This method was originally developed for the analysis of moving heat source problems and for the determination of the temperatures at sliding contacts. However, when two bodies -- one stationary and the other moving relative to the heat source, are involved in sliding contact, the heat partition between them is invariably involved. The heat partition, however, is not a constant along the interface as the temperature distributions are different for the stationary and moving bodies. However, such a problem can be addressed using Blok's ingenious heat partition technique [14-17] combined with Chao and Trigger's functional analysis approach [18] that takes into account variable heat partition. In this paper, their pioneering contributions are briefly reviewed. It may be noted that the solutions developed based on these principles for various manufacturing and tribological problems are general solutions in that both transient and steady state analyses can be conducted [19]. They can also be used to determine the temperature distribution not only at the surface but also with respect to depth which again is a very important consideration for most manufacturing and tribological applications as it affects the subsurface deformation, metallurgical changes, hardness variation, and residual stresses. The analysis can also determine the maximum and average temperatures within the area of the heat source. Thus, the analytical procedure developed is comprehensive and the results obtained are exact compared to other methods using FEM or FDM methods. In this paper, Jaeger's classical heat source method together with Blok's heat partition and Chao and Trigger's functional analysis approach will be used to illustrate the analysis of the temperature distribution in the workmaterial, chip, and cutting tool in machining..*

Keywords*: manufacturing processes, moving heat source, heat partition, metal cutting.*

Nomenclature

a thermal diffusivity, cm²/sec
c specific heat, J/g. $^\circ$C
l half-width of the band heat source, cm
l_i location of the differential small segment of the band heat source dl_i elative to the origin of the coordinate system and along its width, cm
L width of the tool-chip interface band heat source, cm
M any point in the medium where the temperature rise is concerned
N_{Pe} Peclet number, $N_{Pe} = vl/2a$
q_l heat liberation intensity of a moving line heat source, J/cm $\cdot$ sec
q_{pl} heat liberation intensity of a moving plane heat source, J/cm² $\cdot$ sec
R distance between the moving line heat source and the point *M*, where the temperature rise is concerned, cm
r chip thickness ratio, $r = \frac{t_c}{t_{chip}}$, or $r = \frac{V_{ch}}{V_c}$
t_c depth of cut, or undeformed chip thickness, cm
v velocity of a moving plane heat source, cm/sec
v_c cutting speed, cm/sec
v_s shear velocity, cm/sec
v_{ch} chip velocity, cm/sec
w width of cut, cm
X, z the coordinates of the point where the temperature rise is concerned in the moving coordinate system, cm
α rake angle, degrees
θ temperature rise, $^\circ$C
θ_M temperature rise at point M, $^\circ$C
λ thermal conductivity, J/cm $\cdot$ sec $\cdot$ $^\circ$C
ρ density, g/cm³
ϕ shear angle, degrees
φ oblique angle, degrees
B fraction of the heat conducted into the workmaterial
$(1-B)$ fraction of the heat conducted into the chip

1. Introduction

Various manufacturing processes used in the production of a wide range of products account for a significant fraction of GNP. Several hundred billion dollars per year are consumed in the U. S. alone for these activities [20]. The plastic deformation and/or frictional energy involved in the above operations are manifested as heat energy generating high temperatures as well as high temperature gradients that can have detrimental effects on the finished parts as well as on the tools. For example, the high temperatures can accelerate wear of tools (or limit the cutting speed capability in the case of material removal operations) and the high temperature gradients can cause detrimental surface and near surface modifications, such as microstructural modifications, subsurface deformation, heat affected zone (HAZ), chemical modifications of the material, thermal cracking, hardness variation, and residual stresses all of which come under a generic term known as surface integrity of the part. A fundamental analysis of the thermal aspects of various manufacturing processes and tribological applications (friction, lubrication, and wear) can shed light on the physics of the process and enables optimization of process parameters, selection of appropriate tools (materials, geometry etc.), and lubricants for improved performance, reliability in service, and ultimately the economics of the operation.

2. Various Approaches to Thermal Problems

There are basically three approaches to address thermal problems in manufacturing and tribology. They are (1) analytical (or mathematical), (2) numerical (or simulation), such as finite difference method (FDM), finite element method (FEM), boundary element method (BEM), and (3) experimental. Each method complements the other two and has to be checked for accuracy by comparing amongst them. Otherwise, the validity of any one method cannot be ascertained with certainty. This is because while each method has its own advantages, it has its own limitations. For example, the analytical approach may require some simplifying assumptions to solve complex partial differential equations, such as quasi-steady state conditions, simple geometry and boundary conditions, constant thermal properties, etc. Similarly, while variable thermal properties with temperature can be incorporated in the simulation methods, the accuracy of these techniques are somewhat limited by the size of the grid and the distances between the nodes. While finer grid gives more accuracy, it can be at the expense of inordinate computational time, stability, and accumulated errors. Similarly, experimental techniques are limited by the accuracy and limitations of the sensors and the instrumentation used. In this paper only the analytical approach is considered for the case of metal cutting. For other manufacturing processes, similar analytical approach can be made (see References 1-12, and 19 for details). The geometry of the system under consideration, the operating conditions, the initial and boundary conditions will vary depending on the type of manufacturing process under consideration. In view of limited space, a detailed account of them is considered outside the scope and interested readers are encouraged to refer to References 1-12, 19 for details.

3. Pioneering Contributions to the Thermal Analysis

Pioneering contributions to the thermal analysis of moving and stationary heat sources were made by Rosenthal [21-24], Blok [14-17], and Jaeger [13,25] during the mid-1930's to the mid-1940's. Rosenthal applied the moving heat source analysis to arc welding and introduced the moving coordinate system for addressing a wide range of welding problems. Blok considered moving and stationary heat source problems involved in tribology and introduced an ingenious heat partition method. Jaeger introduced the heat source method that formed the basis for much of the analysis that followed. Their combined work lead to the subsequent analyses of thermal problems in various manufacturing processes and tribology. For example, in the field of metal cutting, pioneering analytical work on the temperatures in machining began in the early 1950's by Chao and Trigger [26], Hahn [27], Trigger and Chao [28], and Loewen and Shaw [29]. Other important analytical contributions in this area include the work of Leone [30], Nakayama [31], Boothroyd [32],Weiner [33], Rapier [34], Dutt and Brewer [35], and Dawson and Malkin [36], to name some. Barrow [37] reviewed both theoretical and experimental techniques for assessing the cutting temperatures and may be referred to for details as well as on other contributions. Many consider the seminal contributions to metal cutting made during 1940's and 1950's as the "Golden Age of Metal Cutting"

Hahn [27] considered the shear plane heat source in cutting as an oblique band heat source moving at the cutting velocity in the direction of cutting to determine temperature distribution. This approach, which is adopted by the authors, does not require a priori the assumption of heat partition between the workmaterial and the chip. Regarding the temperature distribution due to chip-tool interface frictional heat source, Chao and Trigger [18] introduced the functional analysis approach to take into account variable heat partition at the interface due to a stationary (tool) and a moving (chip) body. In the following, their contributions will be briefly reviewed as they form the basis for the thermal analysis of various manufacturing problems and tribological applications in our investigations. Due to limitations of space, other significant contributions will not be discussed here; instead the readers are referred to the various papers published in this area by the authors [1-12, 19] where literature review is covered in more detail.

3.1. Rosenthal's Moving Heat Source and the Moving Coordinate System:

In many manufacturing processes, the heat source involved is not stationary but moves relative to the conduction medium. Examples of this include arc welding, surface transformation hardening of a gear tooth with a laser beam. Here, heat is supplied by external means and is transferred directly to the conduction medium (assuming losses due to convection and radiation to be small or negligible).

Rosenthal [21-24] made seminal contributions to the analytical treatment of the temperature distribution in a conducting medium due to moving heat sources. Starting from the Fourier partial differential equation (PDE) of heat conduction (Eqn.1), Rosenthal applied it to welding problems (assuming the heat source to be a moving point or moving infinitely long line heat source) by introducing a moving coordinate system.

$$\frac{\partial^2\theta}{\partial x^2}+\frac{\partial^2\theta}{\partial y^2}+\frac{\partial^2\theta}{\partial z^2}=\frac{1}{a}\frac{\partial\theta}{\partial t} \qquad (1)$$

When the origin of the moving coordinate system coincides with the moving heat source and moves along with it with the same speed (with its X-axis coinciding with the x-axis of the original absolute coordinate system), the relationship between the coordinates of the point where the temperature rise is concerned along X-(or x-) axis at any time t is given by $X = x - vt$. Substituting this in Eqn. 1, the general partial differential equation (PDE) of heat conduction in a moving coordinate system is obtained as:

$$\frac{\partial^2\theta}{\partial X^2}+\frac{\partial^2\theta}{\partial y^2}+\frac{\partial^2\theta}{\partial z^2}=-\frac{v}{a}\frac{\partial\theta}{\partial X}+\frac{1}{a}\frac{\partial\theta}{\partial t} \qquad (2)$$

Eqn. 2, even for a unidirectional heat flow, e.g., along the X-axis, involves three variables, namely, X, θ, and t. Hence, solution of this equation by the separation of variables or other similar techniques is not be feasible. To solve this problem, Rosenthal incorporated an experimental observation [21-24], namely, when the working time of the moving heat source is sufficiently long, the temperature rise distribution around the heat source (in the moving coordinate system) would reach quasi-stationary state conditions, i.e., $\partial\theta/\partial t = 0$. This means an observer stationed at the point heat source fails to notice any change in the temperature around him as the source moves on. Rosenthal also gave an alternate analogy for this wherein the temperature distribution around the heat source is represented by a hill which moves as a rigid body on the surface of the plane without undergoing any modification either in size or shape.

The PDE of heat conduction in a moving coordinate system for quasi-steady state condition is thus given by:

$$\frac{\partial^2\theta}{\partial X^2}+\frac{\partial^2\theta}{\partial y^2}+\frac{\partial^2\theta}{\partial z^2}=-\frac{v}{a}\frac{\partial\theta}{\partial X} \qquad (3)$$

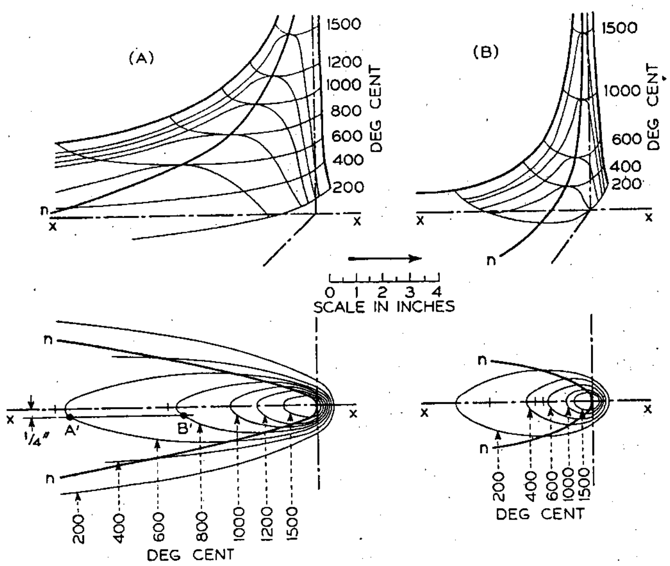
Based on the similarities between this problem and the well known problems encountered in electrical waves, Rosenthal proposed the following solution (Eqn.4) by analogy. By observing the tear drop-like shape of the weld pool on the surface experimentally, Rosenthal [22] considered the final solution of Eqn. 3 as a product of two separate functions given by:

$$\theta = e^{-vX/2a}\varphi(X,y,z) \qquad (4)$$

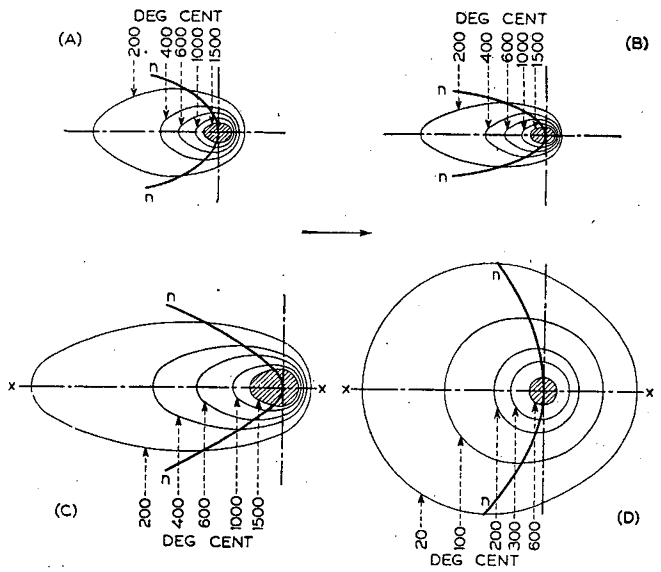
The first part of Eqn.4, namely $e^{-vX/2a}$ is an asymmetric function along the X-axis. $e^{-vX/2a} < e^{-v(-X)/2a}$, i.e. larger the value of the term $|vX/2a|$, higher the asymmetry. The second part, namely, $\varphi(X,y,z)$ is considered as a symmetric function. As a whole, θ is an asymmetric function along the X-axis. This consideration is acceptable for it is closer to the real situation as will be shown in the following.

In practical cases of moving heat source problems, rise of temperature in front of the heat source (where X is positive) is steeper than the fall of temperature behind the heat source (where X is negative), i.e. the temperature distribution is asymmetric along the X-axis relative to the heat source. The larger the value of the term $v/2a$, the higher the asymmetry. When $v/2a = 0$ ($v = 0$), it becomes a symmetrical (only the second part exists) function which is the case for a stationary heat source problem. With this substitution, one needs to solve only the symmetric function, $\varphi(X,y,z)$. Thus, the moving heat source problems are greatly simplified. For solving moving infinitely long line heat source and moving point heat source problems, Rosenthal changed the coordinate system into cylindrical and spherical, respectively to reduce the number of variables and developed solutions for a number of quasi-stationary moving heat source problems. Rosenthal justified the need for both experimental and analytical approaches. He pointed out that while experiment reveals the particular features of every process, the theory permits the establishment of the general laws and thus contributes to the fundamental knowledge of the process.

Figures 1 (a) and (b) show a comparison of the temperature distribution around the heat source for welding a thin and a thick plate under quasi stationary conditions, after Rosenthal [22, 24]. The top views are the temperature distributions along the mid-plane and the bottom figures are the temperature distributions on the surface. It can be seen that the rise of temperature in front of the heat source is steeper than the fall of temperature behind the source. Also, a wider heat affected zone (HAZ) and less steeper temperature gradient can be seen with the thinner plate. Figures 2 (a) to (d) show the temperature distribution as a function of current density, welding speed, and nature of the base material (aluminum versus steel) under quasi stationary conditions, after Rosenthal [22,24]. It can be seen that increasing the current density [Figures 2 (a) and (c)] widens the HAZ without much change in the shape of the isotherms. Also, the welding speed [Figures 2 (a) and (b)] affects mostly the shape of the isotherms. Further, greater heat diffusion in aluminum Figure 2 (d) results in a more circular shape of isotherms than in iron Figure 2 (c). Figure 3 shows the temperature distribution in the cross section of an edge welded plate under quasi stationary conditions, after Rosenthal [22,24]. The dotted lines are experimental and the continuous lines are analytical results. Rosenthal reported satisfactory agreement between the analytical and experimental results except at the immediate neighborhood of the arc. He explained for the discrepancy as due to the fact that the heat source is considered as a point or a line in the analysis while in practice the heat source has a finite size. Because of this, Rosenthal cautioned the use of analytical results a few mm from the heat source. By considering a plane heat source, this problem can be overcome and temperatures close to the heat source and even at the heat source can be determined [1,2].



Figures 1 (a) and (b) - Comparison of the temperature distribution around the heat source for welding a thin and a thick plate under quasi stationary conditions.



Figures 2 (a) to (d) - Temperature distribution as a function of current density, welding speed, and nature of the base material (aluminum versus steel) under quasi stationary conditions.

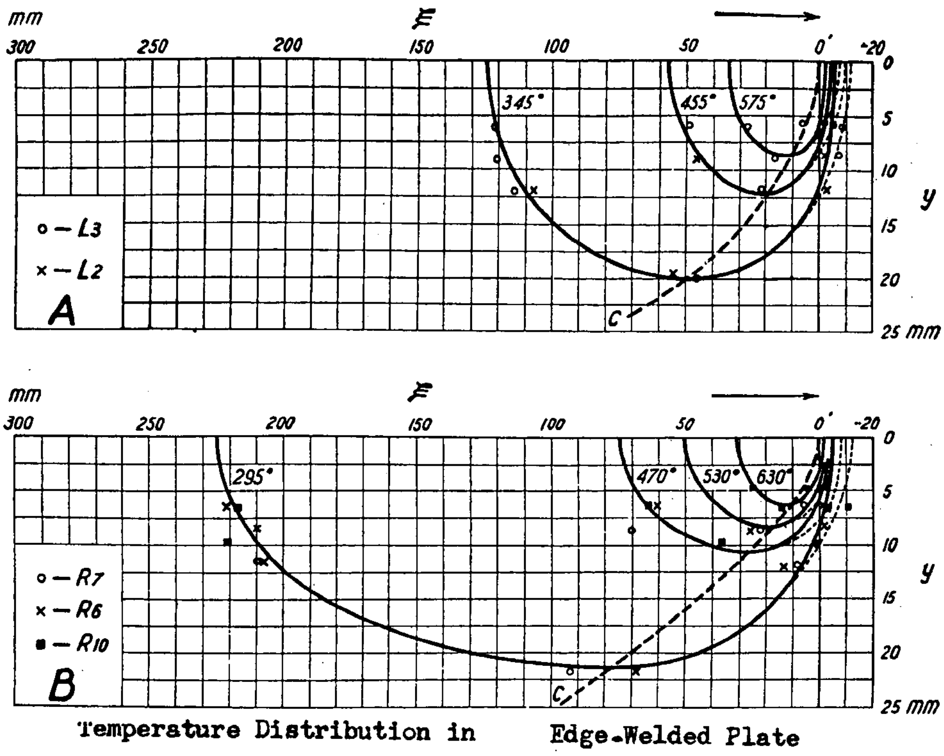


Figure 3 - Temperature distribution in the cross section of an edge welded plate under quasi stationary conditions.

3.2. Blok's Heat Partition Method

There are many applications where heat is not supplied externally but is generated at the interface between two contacting bodies in relative motion due to the inherent phenomenon involved, such as frictional heat. Examples of such applications in manufacturing processes include the frictional heat generated at the chip-tool interface in cutting, the contact area heat source between the abrasives and the workmaterial during grinding or polishing, and sliding contact heat source, such as in bearings or meshing of gears in tribological problems. In such cases, there is a partition of heat between the two sliding bodies and it is necessary to establish the heat partition ratio along the length of the interface to determine the temperature rise distribution in both moving and stationary bodies.

Blok of the Delft University in Netherlands was interested in the flash temperatures generated in gears in the late 1930's to mid 1940's [14-17]. He recognized the generation of extremely high temperatures, known as the flash temperature, instantaneously at the mating surfaces between the teeth of meshing gears. He showed that the gears would reach certain average temperature after running for some time but the flash temperatures will be superimposed on the average temperatures of the gears. He postulated that some lubricants can cease to provide effective protection of the teeth of the gears resulting in galling, due to film breakdown when the combined temperature reaches a critical value. Of course, the normal load also plays an important role. This lead to the development of new extreme pressure lubricants for gears. He thus laid the foundation for the flash temperature theory of film breakdown in lubrication. Blok commented that flash temperatures stand at the cradle of engineering, certainly in so far as man's mastery of heat is concerned. In fact, when man invented the art of making fire by rubbing two wooden sticks together, he made an intelligent and far-reaching use of flash temperature [17].

But even more significant contribution of Blok, as far as the thermal analysis of various manufacturing processes and tribology is concerned, is the development of an ingenious principle of heat partitioning between a stationary heat source and a moving heat source. He considered the case of heat generated due to friction between two bodies in relative motion. Figure 4 (a) is a schematic of a plane surface of a moving body 2 pressed against a round surface of a stationary body 1, after Blok [16]. Blok considered the case of Body 2 moving at a high velocity v over the stationary body 1 (actually, Peclet number, $L > 5$). He assumed the heat to be generated at a uniform rate per unit time per unit area over the contacting interface. The heat partition fraction into the moving body 2 is designated as A□ and that into the stationary body 2 is designated as A_2 But $A_1 + A_2 = 1$. The total heat generated at the interface will be divided between the two bodies in such a way that body 2 tends to absorb greater part A_2Q of the heat Q, as during a certain duration the moving body 2 represents more cooling material than the stationary body 1. The remaining smaller portion of heat A_1Q (or $1-A_2Q$) will penetrate into the stationary body 1.

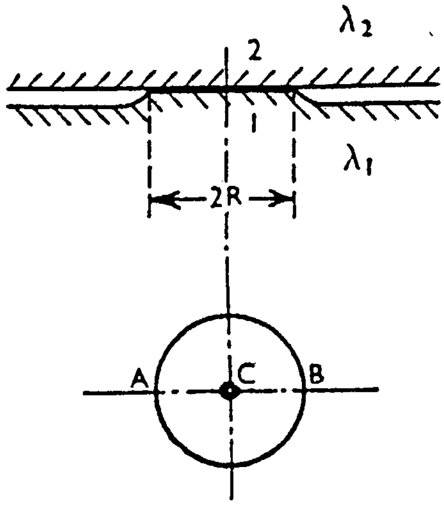
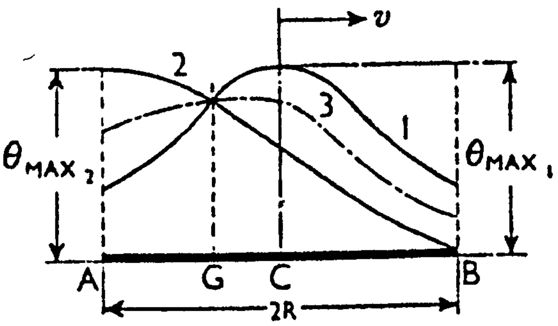


Figure 4 (a) - Schematic of a plane surface of a moving body 2 pressed against a round surface of a stationary body 1.



Curve 1 : Temperature field on stationary body 1 for evenly distributed heat supply A_1Q.
Curve 2 : Temperature field on moving body 2 for evenly distributed heat supply A_2Q.
Curve 3 : Average of curves 1 and 2.

Figure 4 (b) – Temperature distribution at the surface of actual contact as well as the location of the maximum temperature for a stationary body 1 and a moving body 2 relative to the heat source.

Blok argued that nowhere on the surface of actual contact should there be a temperature jump between the contacting bodies leading to the conclusion that the heat supplied to the stationary and moving bodies will not be evenly distributed. To verify it, he considered separately the respective temperature fields on body 1 and body 2 which are established, as if A_1Q is uniformly distributed over the surface of actual contact on the stationary body 1, and A_2Q (or $1\text{-}A_1Q$) is uniformly distributed over the surface of actual contact on the moving body 2. The heat partition fractions A_1 and A_2 are so chosen that the respective temperature fields show the same maximum temperatures, θ_{max1} and θ_{max2} when considered separately as when they are considered together. However, the temperature distribution at the surface of actual contact as well as the location of the maximum temperatures will be different for a stationary heat source compared to a moving heat source [Figure 4 (b)] [16]. He approximated the common surface temperature field by the average of the two temperature fields represented by curve3 which is an average of curves 1 and 2 in Figure 4 (b). Since the exact calculation of the distribution of A_{1Q} and A_{2Q} and of the actual common surface temperature field present considerable difficulty, Blok used an approximate method to arrive at the heat partitions A_1 and A_2 as well as the maximum temperature. Jaeger, subsequently modified the heat partition method by considering matching of the average temperatures than the maximum temperatures as will be shown in the following.

3.3. Jaeger's Heat Source Method

In 1942, Jaeger [13] introduced the mathematical analysis of moving sources of heat and the temperature at sliding contacts using the heat source method that became the basis for much of the analytical work that followed for various manufacturing processes and tribology. It is a very powerful technique for solving a wide range of complex problems encountered in various manufacturing processes, including stationary and moving heat sources of various shapes, sizes and heat intensity distributions for both transient or quasi-steady conditions [25, 38].

The basis for the heat source method is the solution of an instantaneous point heat source, i.e., $\theta = \left(\frac{Q_{pt}}{c\rho(4\pi a\tau)^{3/2}}\right) \cdot e^{-R^2/4a\tau}$. It may be noted that this equation can be derived directly from the Fourier partial differential equation of heat conduction [38]. In the heat source method, different heat sources are considered as a superposition of a point heat source. For example, by integrating the equation for the point heat source with respect to appropriate spatial and time variables, the solutions for an instantaneous line, plane, ring, circular disc, cylindrical surface, and spherical surface heat sources as well as for continuous stationary point, line, plane, ring, and other heat sources are obtained. Table 1 gives a summary of equations for various heat source problems. Depending on the type of manufacturing process or tribological situation under consideration, an appropriate heat source (shape and distribution of heat intensity) and resulting equation can be chosen from Table 1.

Table 1 – Solutions for various moving heat sources [38].

Heat source	Solution
Instantaneous point	$\theta(x,y,z,\tau)=\frac{Q_p}{c\rho(4\pi a\tau)^{3/2}}e^{-(x^2+y^2+z^2)/4a\tau}$ $\theta(R,\tau)=\frac{Q_p}{c\rho(4\pi a\tau)^{3/2}}e^{-R^2/4a\tau}$
Instantaneous line (infinitely long)	$\theta(x,y,\tau)=\frac{Q_l}{c\rho(4\pi a\tau)}e^{-(x^2+y^2)/4a\tau}$ $\theta(r,\tau)=\frac{Q_l}{c\rho(4\pi a\tau)}e^{-r^2/4a\tau}$
Instantaneous plane (infinitely large)	$\theta(x,\tau)=\frac{Q_{pl}}{c\rho\sqrt{4\pi a\tau}}e^{-x^2/4a\tau}$
Continuous line (infinitely long)	$\theta(r.\tau)=\frac{q_l}{2\pi\lambda}\Omega\left(\frac{r}{\sqrt{4a\tau}}\right)$
Continuous plane (infinitely large)	$\theta(x,\tau)=\frac{q_{pl}}{2\lambda\sqrt{\pi}}x\left\{\frac{\sqrt{4a\tau}}{x}e^{-x^2/4a\tau}-\sqrt{\pi}\,\mathrm{erfc}\left(\frac{x}{\sqrt{4a\tau}}\right)\right\}$

Jaeger [13] and Carslaw and Jaeger [25] developed a theory for uniform plane heat sources of various shapes (band heat source and rectangular heat source) moving with constant velocity in the surface of a semi-infinite medium (with no loss of heat from the surface) for addressing problems associated with plane sliding. Figure 5 shows the variation of non-dimensional temperature with the contact length (x/l) for different values of L or Peclet Numbers (N_{Pe}) in the plane of the moving heat source, after Jaeger [13]. The direction of sliding is from left to right. It can be seen that the maximum temperature is towards the trailing edge of the slider with the temperatures being more symmetric for lower Peclet numbers. With increase in Peclet number, the maximum temperature moves progressively towards the rear edge. He also developed equations for the steady-state average and maximum temperatures over the area of the heat source for low (L<0.1) and high (L>5) Peclet numbers. The former, to estimate the flash temperatures and the later, to enable comparison with the temperatures measured experimentally using thermocouples, both for a wide range of Peclet numbers. Even though Jaeger was a mathematician by training, he was very much interested on the practical implications of his analysis for addressing various thermal problems. This can be noted from his paper, to quote Jaeger, "in this paper an attempt is made, in connection with the problem of plane sliding, to set out fully the assumptions made and the numerical consequences of the mathematical theory in a form in which it is hoped they can easily be used by experimenters to discuss particular models of sliding surfaces."

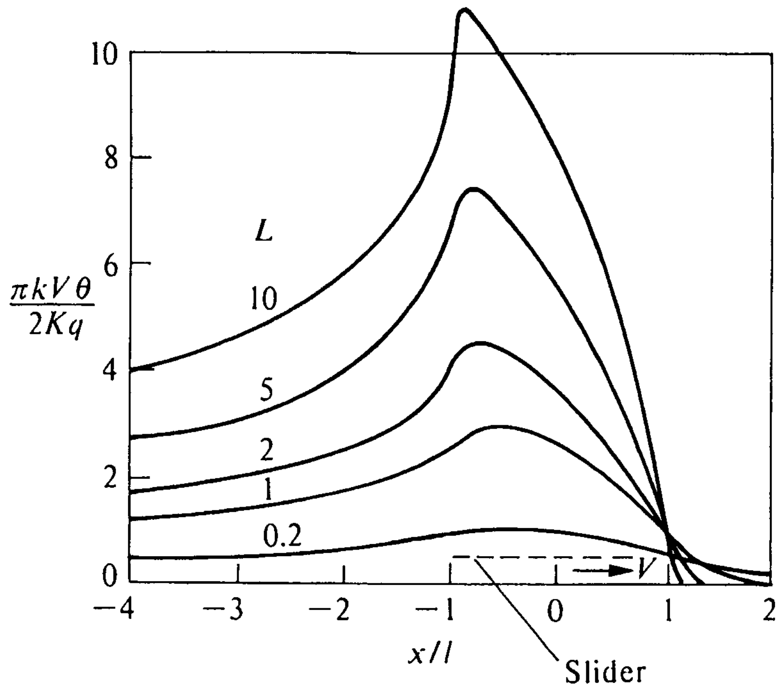


Figure 5 Variation of non-dimensional temperature with the contact length (x/l) for different values of L or Peclet Numbers (N_{Pe}) in the plane of the moving heat source.

3.4. Hahn's Inclined Moving Heat Source for the Shear Plane Heat Source

Hahn [27] considered an oblique moving band heat source model based on the true nature of the chip formation process. During cutting, the depth of the layer to be removed from the workmaterial passes continuously through the shear plane thereby undergoing extensive plastic deformation to form the chip. Hence, the shear plane is considered as a band heat source infinitely long of width $2l$, moving obliquely at an angle ϕ in an infinite medium (the workmaterial) with a heat liberation intensity of q_{pl} (J/cm²sec) at the velocity of cutting, V (see Figure 9 for details).

For calculating the temperature rise at any point M, i. e., Hahn used Jaeger's solution for an infinitely long moving line heat source in an infinite medium

$$\theta_M = \frac{q_l}{2\pi\lambda} e^{-XV/2a} K_o(RV/2a) \quad (5)$$

It may be noted that using Hahn's approach, it is not necessary to assume any heat partition between the chip and the workpiece a priori. For a better representation of the cutting process, it is, however, necessary to modify Hahn's model for a semi-infinite medium taking into account relevant image heat sources [4]. Limited computational resources in the early 1950's almost certainly seems to be the main reason for this simplification in Hahn's analysis. It appears that the significance of Hahn's ingenious idea and his general solution have not been fully appreciated; instead, an approximate approach involving heat partition between the chip and the workmaterial is frequently used. It is hoped that future researchers will take note of this and adopt Hahn's model instead.

3.5. Chao and Trigger's Functional Analysis Approach

Tool-chip interface temperature in metal cutting has long been of considerable interest as it affects the wear and the life of a cutting tool as well as limits the maximum cutting speed that can be used economically [20, 26-29, 31]. Chao and Trigger [18] applied Blok's principle of heat partition [14] by considering the frictional heat source moving relative to the chip and stationary relative to the tool. They calculated the average heat partition fractions for the chip and the tool, and from them the resulting average temperature at the tool-chip interface. Figure 6 shows the interface temperature distribution for the case of uniform distribution of heat flux, after Chao and Trigger [18]. It can be seen that while the calculated and measured average temperatures are in good agreement, the anomaly of having two completely different temperature distributions at the interface is evident. Chao and Trigger [18] pointed out the difficulties that arise from the use of Blok's partition principle. This is due to the rather unrealistic assumption used in the procedure, namely, the distribution of heat flux over the contact region for both the tool and the chip sides is uniform, for computational ease. When the temperature distribution is calculated by this method, two completely different temperature distributions will result. The maximum temperature of the moving heat source will be near the trailing edge while that for the stationary heat source will be close to the center of the heat source. In order to bring the two temperature distribution curves near to coincidence, the heat flux distribution must necessarily be non-uniform.

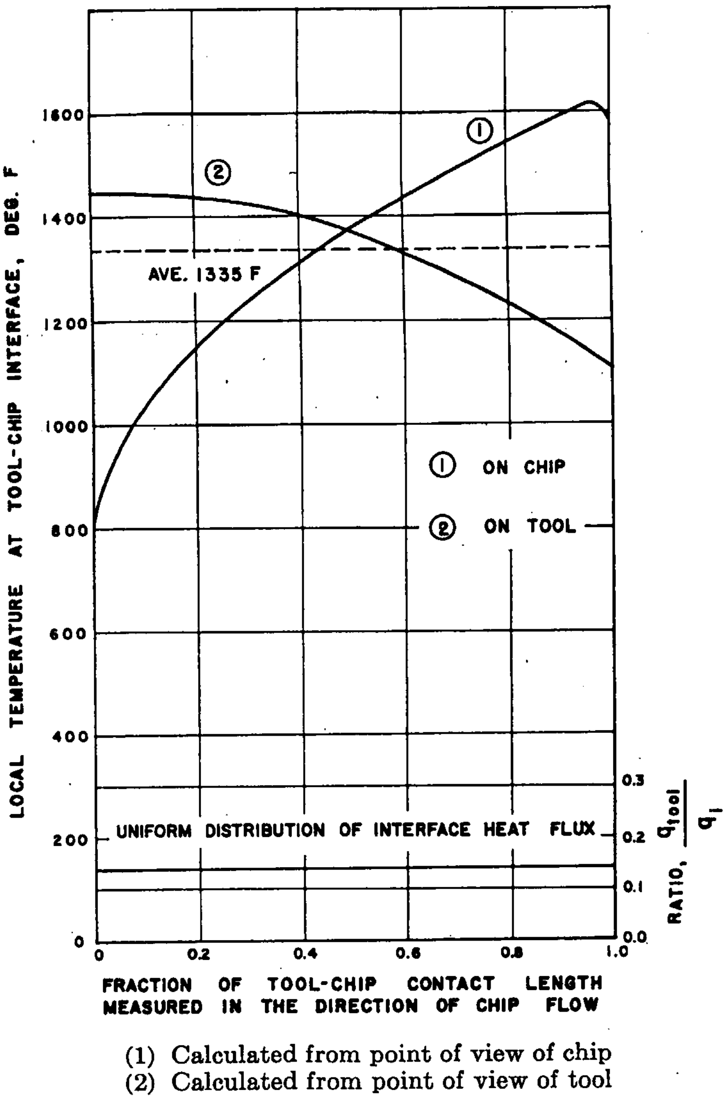


Figure 6 - Interface temperature distribution for the case of uniform distribution of heat flux.

To elucidate the temperature field at the chip-tool interface, Chao and Trigger [18] proposed in 1955 an approximate analytical procedure by assuming an appropriate functional relationship [Figure 7], namely, exponential [Eqn. (6)], for the non-uniform distribution of the heat-flux partition fraction at the tool-chip interface both for the tool side and the chip side, with the objective of coinciding the temperature distribution curves along the tool-chip interface due to moving (chip) and stationary (tool) heat sources.

$$\frac{q_{ic}}{q_i} = C_1 - C_2 e^{-C_3(1-\frac{\xi}{l})} \tag{6}$$

where q_{ic} is the rate of heat transfer into the chip per unit area at location ξ

q_i is the rate of total interface heat per unit area of contact (assumed uniform)

$2l$ is the length of the tool-chip interface contact

ξ is the distance measured from mid-length of contact, positive in the direction of chip flow, and

C_1, C_2, and C_3 are constants

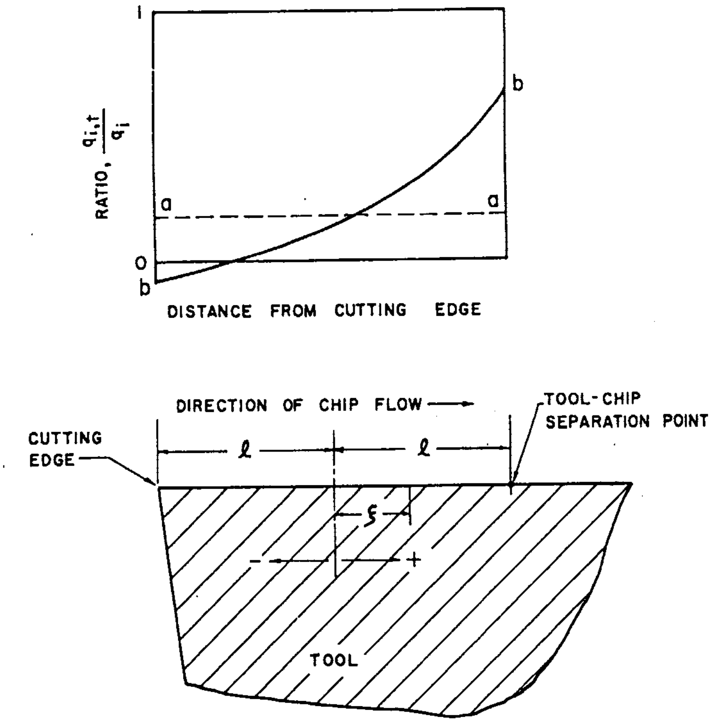


Figure 7 - Non-uniform interface heat flux distribution.

Figure 8 shows the interface heat flux and temperature distribution calculated using the relation given in Eqn.(6) where C_1 = 1.08, C_2 = 0.55, and C_3 = 1.00 [18]. Chao and Trigger [18] pointed out that while a more realistic interface temperature distribution has been achieved (compare Figures 6 and 8), the method has the inherent drawback of using the cut-and-try procedure in finding the constants in the exponential equation used. Besides, they pointed out that it was tedious and time consuming. Since they found an alternative method to solve this problem, namely, iterative technique, they focussed on that method and gave only the results of the functional analysis method briefly. With the availability of fast, inexpensive personal computers, it is relatively easy to use the functional analysis method. The authors have used this approach for addressing a wide range of moving heat source problems uncounted in manufacturing processes and tribological situations. A power function was used instead of an exponential function as it is convenient and fast to match the temperatures. Appendix A gives some details of the various steps to be taken in arriving at the functional relationship for the case of machining steel.

4. Thermal Analysis of Metal Cutting

4.1. Shear Plane Heat Source

The authors have adopted a new approach (built on Hahn's original model) for the thermal analysis in cutting due to the shear plane heat source alone in that the analyses is divided into two separate parts, namely, one on the workmaterial side and the other on the chip side and then combined to develop the temperature distribution in the workmaterial and the chip. It may be noted that it is not necessary to consider heat partition between the chip and the workmaterial a priori but it results as a consequence of the analysis.

To determine the temperature distribution in the workpiece, chip, and the tool due to shear plane heat source alone, Hahn's model has been modified for a semi-infinite medium taking into account the relevant image heat sources. The workmaterial is extended past the shear plane as an imaginary region for continuity [see Figure 9 (a)]. An image heat source with the same intensity as the shear plane heat source is also considered. Thus, the temperature distribution in the workmaterial is obtained. Similarly, for the temperature distribution in the chip due to the shear plane heat source, the shear plane is again considered as an infinitely long oblique heat source moving in the chip at the velocity of the chip in a semi-infinite medium [Figure 9 (b)]. The chip material is extended past the shear plane as an imaginary region for continuity. Again, an image heat source with the same intensity as the shear plane heat source is considered. This way, the distribution of temperature in the chip can be obtained. By combining the temperature distribution in the workmaterial and in the chip, the temperature distribution near the shear zone in metal cutting can be obtained.

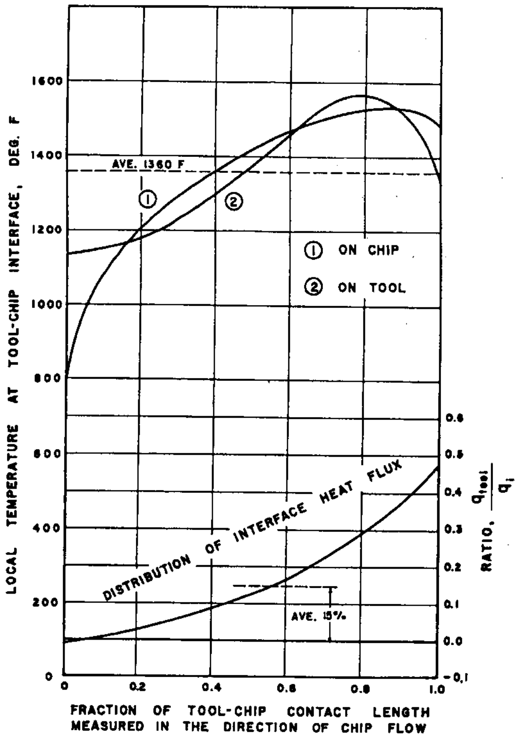
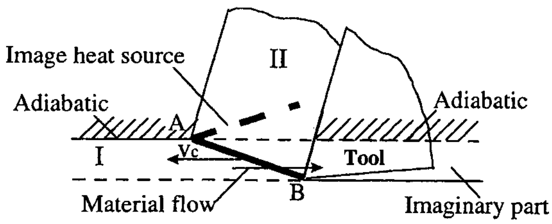
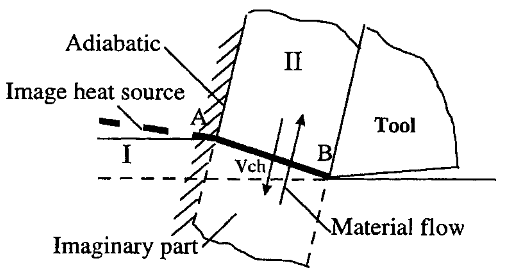


Figure 8 - Interface heat flux and temperature distribution calculated using the relation given in Eqn.(6) where C_1 = 1.08, C_2 = 0.55, and C_3 = 1.00 [18].



(a) model for thermal analysis of work



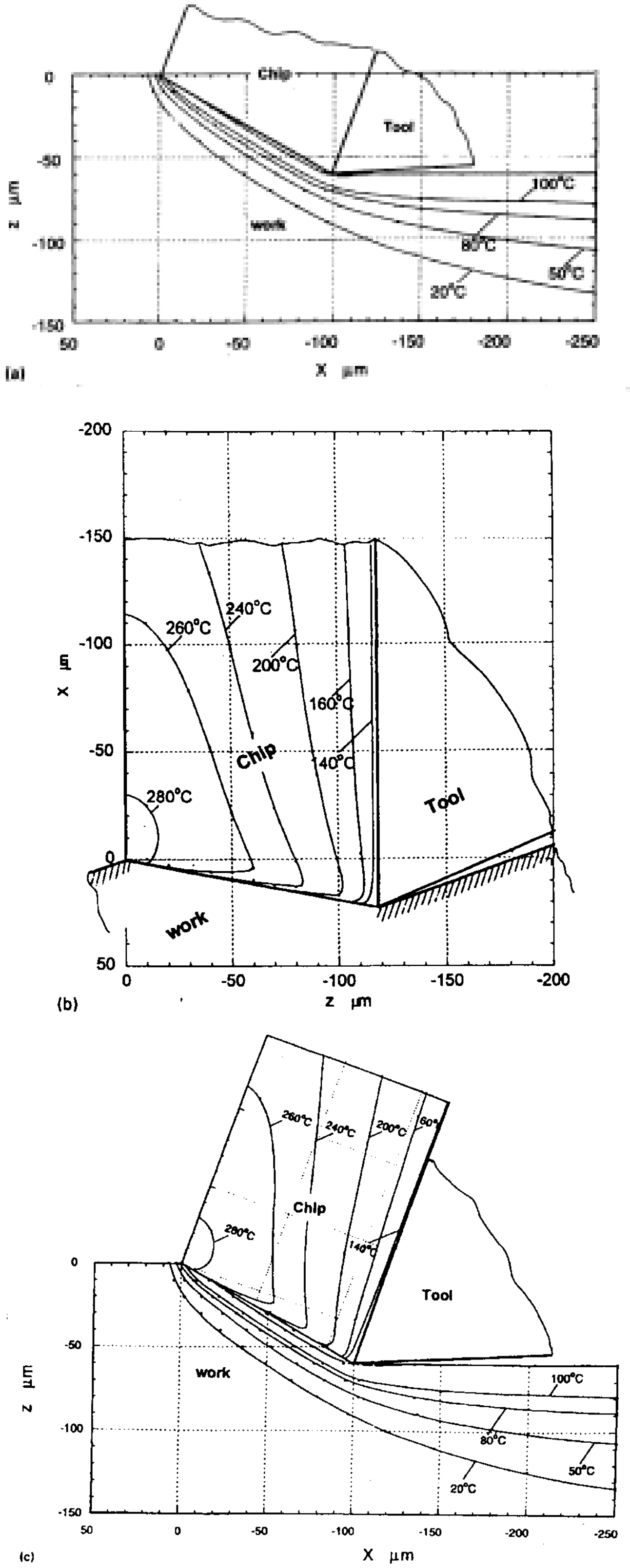
(b) model for thermal analysis of chip

Figures 9 (a) and (b) – Schematics of the moving oblique band heat source model for the continuous chip formation process in the metal cutting using modified Hahn's model for the workmaterial and chip, respectively.

The total temperature rise at any point M caused by the complete oblique moving band heat source including its image heat source of equal intensity is given by [4]

$$\theta_M = \frac{q_{pl}}{2\pi\lambda} \int_{l_i=0}^{L} e^{-(X-l_i \sin\varphi)V/2a} \cdot \left\{ \begin{array}{l} K_o\left[\frac{V}{2a}\sqrt{(X-l_i\sin\varphi)^2+(z-l_i\cos\varphi)^2}\right] + \\ \\ K_o\left[\frac{V}{2a}\sqrt{(X-l_i\sin\varphi)^2+(z+l_i\cos\varphi)^2}\right] \end{array} \right\} dl_i \qquad (7)$$

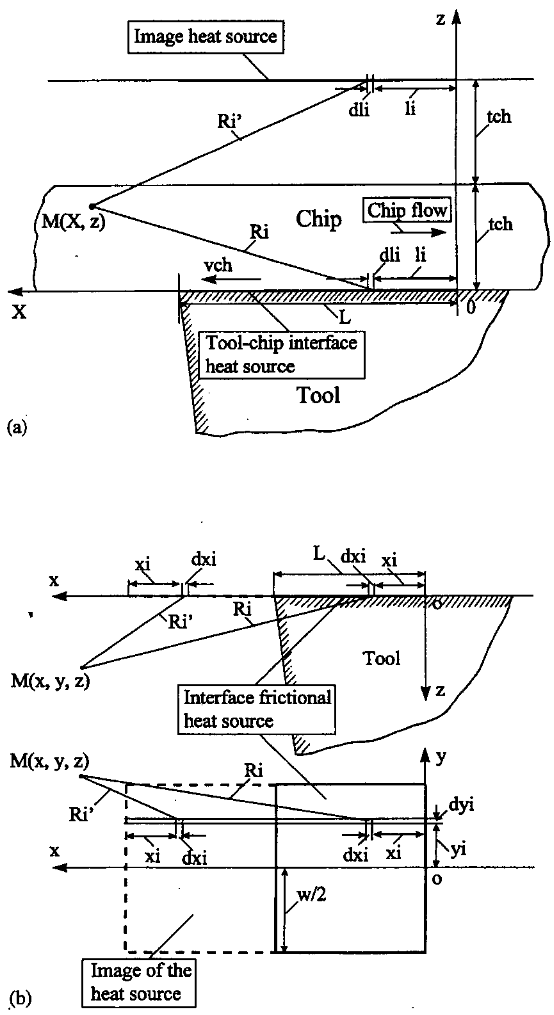
For details see Ref.4. Figures 10 (a) to (c) are the temperature rise distributions in the workmaterial, chip, and combined workmaterial and chip, respectively, in the machining of AISI B1113 steel due to shear plane heat source only, using the data from Shaw [39].



Figures 10 (a) to (c) - Temperature rise distributions in the workmaterial, chip, and combined workmaterial and chip, respectively, in the machining of AISI B1113 steel due to shear plane heat source only, using the data from Shaw [39]

4.2. Frictional Heat Source at the Chip-Tool Interface.

Figures 11 (a) and (b) are schematics of the heat transfer models of the frictional heat source at the tool-chip interface on the chip side and the tool side, respectively [5]. The model considered for the chip-tool interface frictional heat source incorporates Jaeger's [13] classical solutions for a moving band heat source for the chip and a stationary square heat source for the tool taking into account the effect of additional boundaries along with a superimposed non-uniform distribution of heat intensity using functional analysis [5].



Figures 11 (a) and (b) - Schematics of the heat transfer models of the frictional heat source at the tool-chip interface on the chip side and the tool side, respectively [5].

Figure 11 (a) is a schematic of the heat transfer model of the frictional heat source at the tool-chip interface on the chip side [5]. The interface frictional heat source relative to the chip is a band heat source moving with a velocity, v_{ch}. Considering the heat partition fraction for the chip to be B, the heat liberation rate Bq of the moving band heat source is considered totally transferred into the chip. Thus, the interface boundary is considered as adiabatic and the solution used should be for a semi-infinite medium. Since the heat source is entirely on the boundary surface, the solution for a semi-infinite medium will be twice that for an infinite medium.

Solutions for a moving band and a stationary square heat sources were developed by Jaeger [13] for the case of sliding contact of two semi-infinite bodies. But for the case of metal cutting neither sides of the tool-chip interface can be considered as semi-infinite. It may be noted that in addition to the contacting interface between the two bodies, i. e., the chip and the tool, two other boundaries very near the contacting interface, namely, the top surface of the chip and the clearance face of the tool offers a very significant effect on the temperature rise at the interface. Thus, Jaeger's solutions should be modified considering the effect of these additional boundaries for the case of metal cutting. Also, for the reasons discussed earlier, Jaeger's solutions should be modified for non-uniform heat intensity caused by the non-uniform heat partition.

The chip thickness, t_{ch} in metal cutting is considerably small. Hence, the boundary effect of the upper surface of the chip can not be neglected. For this, an image heat source which is a mirror image of the tool-chip interface heat source with respect to the upper boundary surface of the chip and located at a distance of $2t_{ch}$ from the primary interface frictional heat source as shown in Figure 1 should be considered. The heat liberation rate of the image heat source will be the same as the primary heat source when the upper surface is considered adiabatic.
Figure 11 (b) is a schematic of the heat transfer model of the frictional heat source at the tool-chip interface on the tool side [5]. The interface frictional heat source relative to the tool is a stationary rectangular heat source of length, L and width, w. Considering the heat partition fraction for the tool to be (1-B), the heat liberation rate (1-B)q of the stationary rectangular heat source is considered totally transferred into the tool. Thus, the interface boundary is adiabatic and the solution used should be for a semi-infinite medium. Also, as the heat source is on the boundary surface, the semi-infinite solution is obtained by taking twice the value for an infinite medium solution.

Figure 11 (b) is a schematic of the heat transfer model of the frictional heat source at the tool-chip interface on the tool side [5]. The interface frictional heat source relative to the tool is a stationary rectangular heat source of length, L and width, w. Considering the heat partition fraction for the tool to be $(1-B)$, the heat liberation rate $(1-B)q$ of the stationary rectangular heat source is considered totally transferred into the tool. Thus, the interface boundary is adiabatic and the solution used should be for a semi-infinite medium. Also, as the heat source is on the boundary surface, the semi-infinite solution is obtained by taking twice the value for an infinite medium solution.

The clearance face of the tool should also be considered as an adiabatic boundary. Thus, a mirror image heat source of the primary stationary rectangular heat source with respect to the clearance surface should also be considered [see Figure 11 (b)]. The heat liberation rate of the imaginary heat source is the same as that of the primary heat source.
Figure 1. Geometrical configuration.

The total temperature rise at any point $M(X,z)$ in the chip caused by the entire moving interface frictional heat source, including its image source, is given by [5]

$$\theta_M = \frac{q_{pl}}{\pi\lambda}\int_{l_i=0}^{L} e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \tag{8}$$

The total temperature rise at any point $M(X,y,z)$ in the tool caused by the entire stationary rectangular interface frictional heat source, including its image source, is given by [5]

$$\theta_M = \frac{q_{pl}}{2\pi\lambda}\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L}\left(\frac{1}{R_i} + \frac{1}{R_i'}\right)dx_i \tag{9}$$

Using Eqns.(8) and (9) and the assumed heat partition fraction for the chip and the tool, and considering $z = 0$, the local temperature rise and the average temperature rise at the contacting interface on both sides can be calculated. Considering Blok's heat partition principle [14], the average temperature rise on both sides at the contacting interface should be the same. The heat partition fraction for the chip and the tool can be determined and the average temperature rise at the interface can be calculated. Figure 12 shows the temperature rise distribution along the tool-chip interface on the tool side and chip side showing a reasonably good match using the functional analysis approach [5]. Figure 13 shows the temperature field at the chip and the tool due to the frictional heat source at the chip-tool interface [5], using the data from Chao and Trigger [18].

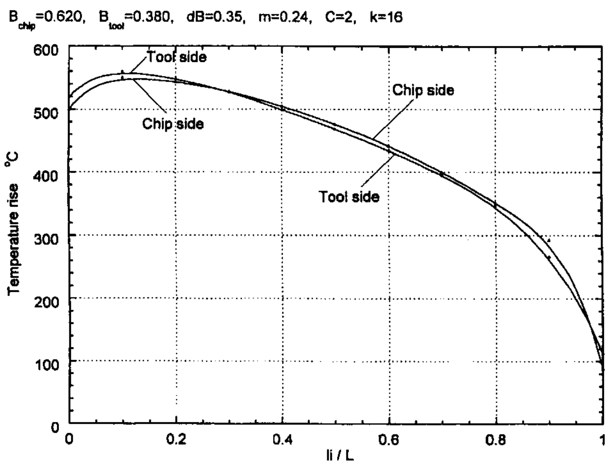


Figure 12 - Temperature rise distribution along the tool-chip interface on the tool side and chip side showing a reasonably good match using the functional analysis approach [5].

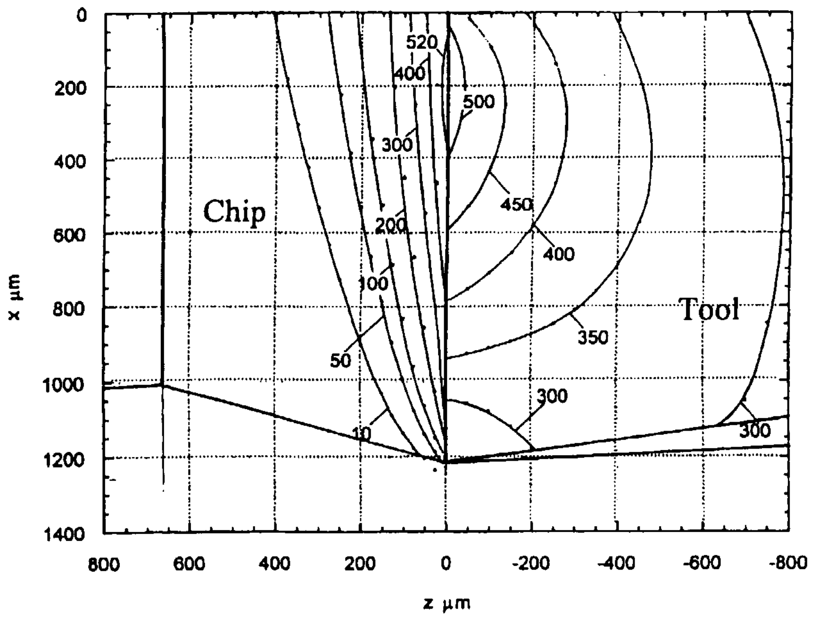


Figure 13 - shows the temperature field at the chip and the tool due to the frictional heat source at the chip-tool interface [5], using the data from Chao and Trigger [18].

4.3. Combined Effect of Shear Plane Heat Source and Frictional Heat Source at the Chip-Tool Interface

Figure 14 is a schematic of the heat transfer model of the two principal heat sources in metal cutting - the shear plane heat source AB and the tool-chip interface frictional heat source OA operating simultaneously in a common coordinate system [6]. Referring to Figure 14, the shear plane heat source together with the tool-chip interface frictional heat source moves relative to the chip at the chip velocity v_{ch} but in a direction opposite to the chip flow. As the entire shear plane heat source is under the upper boundary surface of the chip except point B (which is all the time at the boundary surface), an image heat source *A'B* need to be considered. To determine the temperature rise at any point near the heat source, the solution of the oblique moving band heat source for an infinite medium is used for both the primary shear plane and its image heat sources.

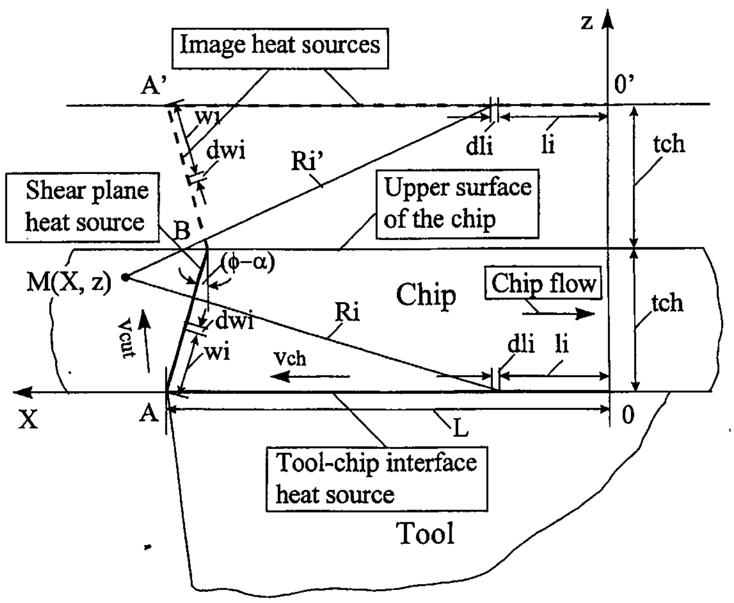


Figure 14 - Schematic of the heat transfer model of the two principal heat sources in metal cutting - the shear plane heat source AB and the tool-chip interface frictional heat source OA operating simultaneously in a common coordinate system [6].

The temperature rise at any point in the chip including all the points at the tool-chip interface caused by the entire shear plane heat source is given by [6]:

$$\theta_M = \frac{q_{pls}}{2\pi\lambda}\int_{w_i=0}^{t_{ch}/\cos(\phi-\alpha)} e^{-(X-X_i)v/2a}\left\{\begin{array}{c} K_0[\frac{v}{2a}\sqrt{(X-X_i)^2+(z-z_i)}]+ \\ \\ K_0[\frac{v}{2a}\sqrt{(X-X_i)^2+(2t_{ch}-z-z_i)^2}] \end{array}\right\}dw_i \tag{10}$$

Thus the total temperature rise at any point in the chip $M(X,z)$ including the points at the tool-chip interface caused by the two principal heat sources is [using Eqns. (8) and (9)] given by [6] :

$$\theta_M = \frac{q_{pl}}{\pi\lambda}\left\{\begin{array}{c} (B_{chip}-\Delta B)\int_{l_i=0}^{L} e^{-(X-l_i)v/2a}[K_0(R_i v/2a)+K_0(R_i^{'} v/2a)]dl_i \\ \\ +2\Delta B\int_{l_i=0}^{L}(\frac{l_i}{L})^m e^{-(X-l_i)v/2a}[K_0(R_i v/2a)+K_0(R_i^{'} v/2a)]dl_i \\ \\ +C\Delta B\int_{l_i=0}^{L}(\frac{l_i}{L})^k e^{-(X-l_i)v/2a}[K_0(R_i v/2a)+K_0(R_i^{'} v/2a)]dl_i \end{array}\right\}$$

$$+\frac{q_{pls}}{2\pi\lambda}\int_{w_i=0}^{t_{ch}/\cos(\phi-\alpha)} e^{-(X-X_i)v/2a}\left\{\begin{array}{c} K_0[\frac{v}{2a}\sqrt{(X-X_i)^2+(z-z_i)}]+ \\ \\ K_0[\frac{v}{2a}\sqrt{(X-X_i)^2+(2t_{ch}-z-z_i)^2}] \end{array}\right\}dw_i \tag{11}$$

Similarly, the total temperature rise at any point in the tool, $M(X,z)$, caused by the two principal heat sources also consists of two parts. One is due to the effect of the frictional heat source at the tool-chip interface from the tool side. This part of the temperature rise can be calculated using Eqn.(8). Another is the effect from shear plane heat source. Eqn.(9) can not be used for calculating this part of temperature rise because the shear plane heat source relative to the tool is stationary and Eqn.(9) is the solution for a moving heat source. It can not also be calculated simply by considering the shear plane heat source as stationary in a stationary body including the chip and the tool, as the chip is moving relative to the shear plane heat source. Hence, an alternate heat transfer model needs to be developed.

A reasonable heat transfer model for the calculation of the temperature rise in the tool caused by the shear plane heat source is to consider that part of heat which is coming from the shear plane heat source and flowing through the chip and the tool-chip interface into the tool, acting as a stationary heat source located at the tool-chip interface. This heat source may be considered as an induced stationary rectangular heat source caused by the shear plane heat source. For this induced heat source, Eqn.(9) can also be used. The equation for the temperature rise distribution in the tool caused by the two principal plane heat sources is given by Eqn.(12) which consists of two terms all using Eqn.(9).

To apply Eqn.(9) for the induced heat source by the shear plane heat source, initially the heat intensity of the induced heat source is unknown but the average temperature rise caused by it at the tool-chip interface is known. The latter is obtained during the calculation of the temperature rise distribution in the chip including at the tool-chip interface on the chip side caused by the shear plane heat source [Eqn.(10)]. Due to the continuity of heat flow, the temperature rise distribution at the tool-chip interface on the tool side is the same as that on the chip side. Using the computer program for Eq.(9) and knowing the average temperature rise at the tool-chip interface ($z = 0$), the average heat liberation intensity of the induced heat source q_{pli} can be calculated. Also, using the functional analysis method, the relevant coefficients and the exponents can be determined.

$$\theta_M = \frac{q_{pl}}{2\pi\lambda_{tool}} \left\{ \begin{array}{l} (B_{tool}+\Delta B)\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{1}{R_i}+\frac{n}{R_i'})dx_i \\ \\ -2\Delta B\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})^m (\frac{1}{R_i}+\frac{n}{R_i'})dx_i \\ \\ -C\Delta B\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})^k (\frac{1}{R_i}+\frac{n}{R_i'})dx_i \end{array} \right\}$$

$$+ \frac{q_{pli}}{2\pi\lambda_{tool}} \left\{ \begin{array}{l} (B_{ind}+\Delta B_i)\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{1}{R_i}+\frac{n}{R_i'})dx_i \\ \\ -2\Delta B_i\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})^{m_i} (\frac{1}{R_i}+\frac{n}{R_i'})dx \\ \\ -C_i\Delta B_i\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})^{k_i} (\frac{1}{R_i}+\frac{n}{R_i'})dx \end{array} \right\} \tag{12}$$

where $R_i = \sqrt{(x-x_i)^2+(y-y_i)^2+z^2}$; $R_i' = \sqrt{(X-2L+x_i)^2+(y-y_i)^2+z^2}$; $x_i = l_i$; q_{pli} is the average heat liberation intensity of the induced heat source in J/cm²sec, m_i, k_i, C_i and ΔB are relevant exponents and constants for the induced heat source.

Figure 15 shows the temperature rise distribution at the tool-chip interface due to combined heat sources for the case of conventional machining of steel [6] using data from Chao and Trigger [18]. Figure 16 shows the isotherms of the temperature rise in the chip, the tool, and the workmaterial under the combined effect of the shear plane heat source and the chip-tool interface frictional heat source for the case of conventional machining of steel with a carbide tool [6] using again the data from Chao and Trigger [18]. Good agreement between the analytical and the experimental results was found.

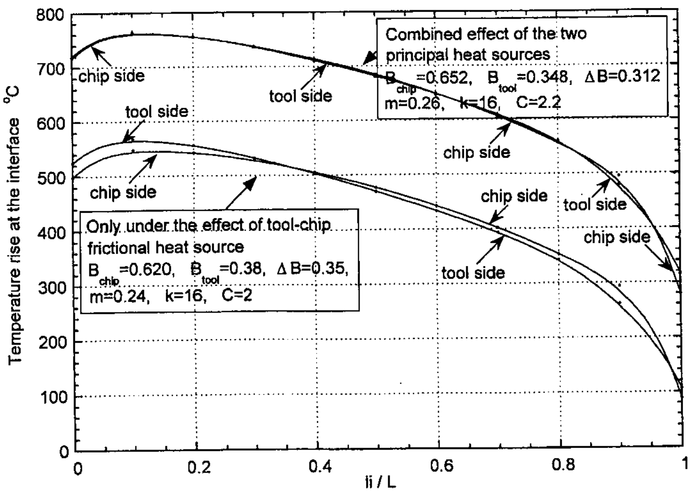


Figure 15 - Temperature rise distribution at the tool-chip interface due to combined heat sources for the case of conventional machining of steel [6] using data from Chao and Trigger [18].

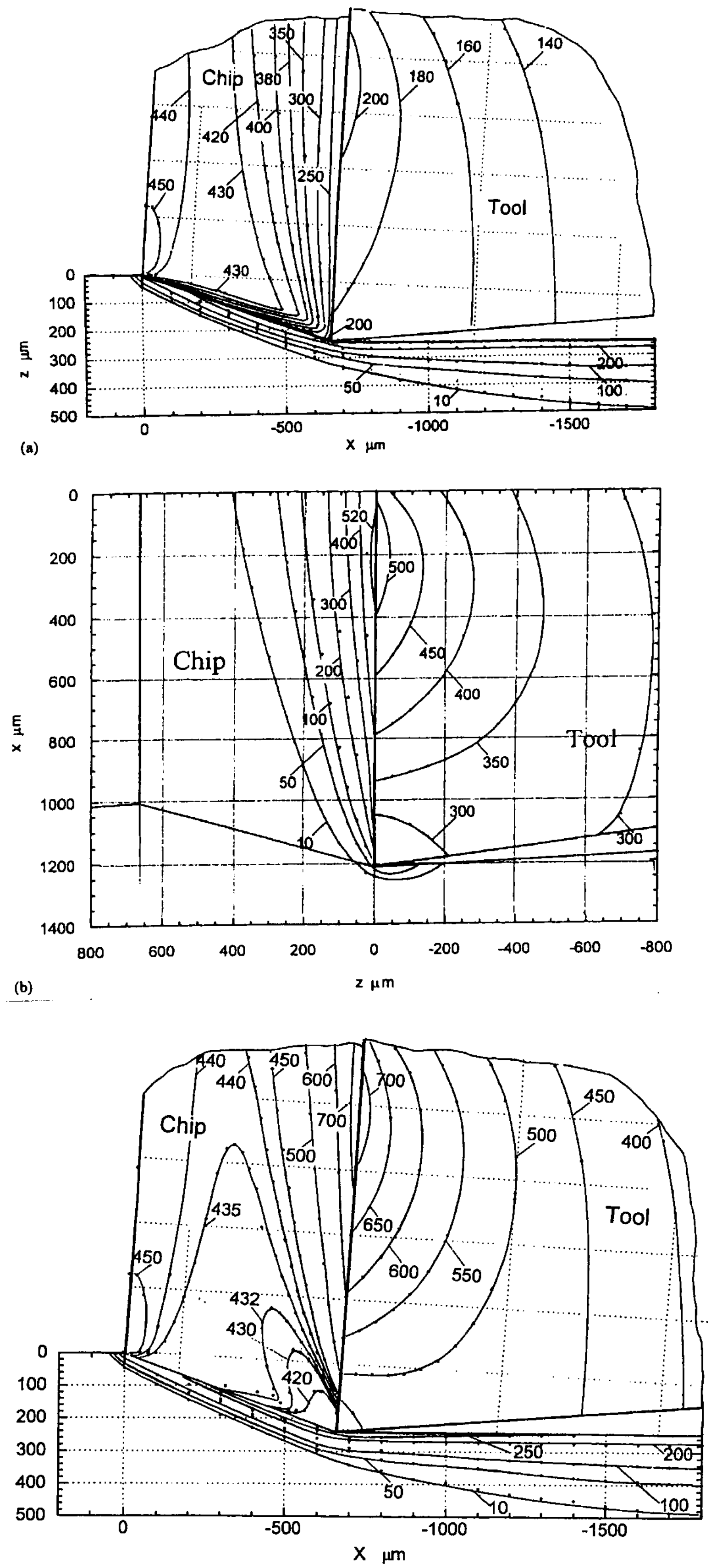


Figure 16 - Isotherms of the temperature rise in the chip, the tool, and the workmaterial under the combined effect of the shear plane heat source and the chip-tool interface frictional heat source for the case of conventional machining of steel with a carbide tool [6] using data from Chao and Trigger [18]

5. Concluding Remarks

It may be pointed out that the thermal analysis presented here considers only pure conduction. While this may not be a drawback in some manufacturing processes, convection may play an important role in other processes. In the present analysis, convective boundary conditions can be considered in a semi-quantitative manner via the boundary conditions (image heat sources). Even though only one example, namely, metal cutting, is used in this paper to illustrate the thermal aspects of various manufacturing processes, the principles outlined are equally applicable for many of the manufacturing processes as well as tribological applications as demonstrated in References 1-12, and 19.

It may be pointed out that analytical methods are very powerful in analyzing complex manufacturing processes and should be used with equal zeal as the numerical methods, such as FEM and FDM. In fact, the analytical methods are much simpler, more powerful, and can be extended to wide range of conditions without much difficulty. Also, limitations on the number of nodes as well as size and shape of the finite elements would not be an issue in the analytical method. Jaeger's heat source method, Blok's heat partition, and Chao and Trigger's functional analysis approach are powerful tools in addressing a variety of manufacturing problems as illustrated in this paper.

One major factor that contributed to the recent success of the analytical approach is the availability of fast, inexpensive personal computers (PC's). One is not limited by the availability of look-up tables for some of the special functions as was the case in the early 1950's. They can now be computed directly using PC's.

If there is any contribution that can be attributed to the authors, it is the attempt made to revive the analytical approach long forgotten by most researchers perhaps due to the availability of numerical techniques, such as FEM. Some even question the very need for analytical approach in view of the availability of commercial software for FEM analysis. To them FEM is the solution rather than a technique for the analysis of a problem.

In this connection, it is appropriate to quote a comment made by one of the reviewer of the authors paper on welding, namely, "contrary to the view of many academicians, the topic of weld thermal condition prediction is still of great importance to the manufacturing and fabrication industries. Despite the fancy mathematical and computational tools available today, the ability to describe weld pool thermal conditions has not improved much lately. Refining the computational technics alone, without recognizing the fundamentals of material properties, did not result in fantastic enhancements in modeling."

In closure, the authors also would like to sincerely apologize for not quoting the extensive literature on the thermal aspects of various manufacturing processes and tribology due to a variety of reasons, namely, (a) limited space available, (b) much of it is covered in some of the recent publications of the authors (see Refs. 1-12, 19), and (c) the emphasis of this paper is to illustrate the power of Jaeger's heat source method combined with Blok's heat partition technique and Chao and Trigger's functional analysis approach.

6. Acknowledgments

This project was initiated by a grant from the NSF U. S. - China Co-operative research project on the Thermal Aspects of Manufacturing. One of the authors (R.K.) thanks Dr. Alice Hogen of NSF for facilitating this activity and for her interest in this project. The authors are indebted to NSF for their continuing support to one of the authors (R.K.) at OSU on the various aspects of the manufacturing processes. Thanks are due, in particular, to Drs.Kesh Narayanan, K. Rajurkar, Delci Durham of the Division of Design, Manufacturing, and Industrial Innovation (DMII) and to Dr. Jorn Larsen Basse of Tribology and Surface Engineering program, and Dr. B. M. Kramer of the Engineering Centers Division of NSF. The author also thanks the A. H. Nelson, Jr. Endowed Chair in Engineering for the financial support in the preparation of this paper.This paper is respectfully dedicated to the pioneers in the field, especially, Danial Rosenthal, Hermon Blok, J. C. Jaeger, Bob Hahn, Milton Shaw, Ken Trigger, and B. T. Chao.

7. References

[1] Komanduri, R., Hou, Z. B., Thermal Analysis of the Arc Welding Process - Part I: General Solutions, Metallurgical and Materials Transactions B,.31 (2000) pp.1353-1370,

[2] Komanduri, R., Hou, Z. B., Thermal Analysis of the Arc Welding Process - Part II: Effect of variation of thermo-physical properties with temperature, accepted for publication in the Metallurgical and Materials Transactions B, 32 (2001) pp.483-499.

[3] Komanduri, R., Hou, Z. B., Thermal Analysis of the Laser Surface Hardening Process, Int. J of Heat & Mass Transfer, 44 (2001) pp. 2845-2862

[4] Komanduri, R., Hou, Z. B., Thermal Modeling of the Metal Cutting Process, Part I - the temperature rise distribution due to shear plane heat source, Int. J. Mech. Sci., 42 (2000) pp.1715-1752

[5] Komanduri, R., Hou, Z. B., Thermal Modeling of the Metal Cutting Process, Part II - the temperature rise distribution due to frictional heat source at the chip tool interface, Int. J. Mech. Sci., 43 (2001) pp.57-88

[6] Komanduri, R., Hou, Z. B., Thermal Modeling of the Metal Cutting Process, Part III - the temperature rise distribution due to the combined effect of shear plane heat source and frictional heat source at the chip tool interface, Int. J. Mech. Sci., 43 (2001) pp.89-107

[7] Hou, Z. B., Komanduri, R., Magnetic Field Assisted Finishing of Ceramics, Part I: Thermal Model, Trans. ASME, J. of Tribology, 120 (1998) pp.645-651

[8] Hou, Z. B., Komanduri, R., Magnetic Field Assisted Finishing of Ceramics, Part II: On the Thermal Aspects of Magnetic Float Polishing (MFP) of Ceramic Balls, Trans. ASME, J. of Tribology, 120 (1998) pp.652-659

[9] Hou, Z. B., Komanduri, R., Magnetic Field Assisted Finishing of Ceramics, Part III: On the Thermal Aspects of Magnetic Abrasive Finishing (MAF) of Ceramic Rollers, Trans. ASME, J. of Tribology, 120 (1998) pp. 660-667

[10] Hou, Z. B., Komanduri, R., Modeling of Thermo-Mechanical Shear Instability in Machining, Int. J. Mech. Sci., 39,(1997) pp.1273-1314

[11] Komanduri, R, Hou, Z. B., Thermal Analysis of Dry Sleeve Bearing - A Comparison between analytical, numerical (FEM) and experimental results, Tribology International, 34 (2001) pp.145-160

[12] Komanduri, R, Hou, Z. B., Analysis of Heat Partition and Temperature Distribution in Sliding Systems, (in print) Wear (2001)

[13] Jaeger, J. C., Moving Sources of Heat and the Temperature at Sliding Contacts, Proc. of the Royal Society of NSW, 76 (1942) pp.203-224,

[14] Blok, H., Theoretical Study of Temperature Rise at Surfaces of Actual Contact Under Oiliness Lubricating Conditions, Proc. of the General Discussion on Lubrication and Lubricants, Inst. of Mech. Engrs., London, 2 (1937) pp.222-235

[15] Blok, H., The Surface Temperatures Under Extreme Pressure Lubricating Conditions, Second World Petroleum Congress, Paris, France (1937) pp.151-182

[16] Blok, H., The Dissipation of Frictional Heat, App. Sci. Res., Sec A, 5 (1949) 151-181

[17] Blok, H.,The Flash Temperature Concept, Wear 6 (1963) pp.483-494

[18] Chao, B. T., and K.J. Trigger,Temperature Distribution at the Chip-Tool Interfacae in Metal Cutting, Trans ASME 77 (1955) pp.1107-1121

[19] Hou, Z. B., Komanduri, R, General Solutions for Stationary/Moving Plane Heat Source Problems in Manufacturing and Tribology, Int. J of Heat and Mass Transfer, 43 (2000) pp. 1679-1698

[20] Komanduri, R., Machining and Grinding -A Historical Review of the Classical Papers, Applied Mechanics Reviews 46 (1993) pp.80-132

[21] Rosenthal, D., Theoretical Study of the Heat Cycle During Arc Welding, (in French) 2-eme Congress. Nat. Sci. Brussels (1935) pp.1277-92

[22] Rosenthal, D., The Theory of Moving Sources of Heat and its Application to Metal Treatments, Trans. ASME, 80 (1946) pp.849-866

[23] Rosenthal, D., Mathematical Theory of Heat Distribution During Welding and Cutting, Welding J, Welding Research Supplement (May 1941) pp.220s-233s,

[24] Rosenthal, D., and R. Schmerber, Thermal Study of Arc Welding: Experimental Verification of Theoretical Formulas, Welding Research Supplement, (April 1938) pp.2-8

[25] Carlsaw, H.S. and J. C. Jaeger, Conduction of Heat in Solids, Oxford University Press, Oxford, U.K. (1959)

[26] Chao, B. T., and K.J. Trigger, Cutting Temperature and Metal Cutting Phenomena, Trans ASME 73 (1951) pp.777-793

[27] Hahn, R. S., On the Temperature Developed at the Shear Plane in the Metal Cutting Process, Proc. First U.S. National Congress of Applied Mechanics, (1951) pp.661-666

[28] Trigger, K. J., and B. T. Chao, An Analytical Evaluation of Metal Cutting Temperature, Trans. ASME, 73 (1951) pp.57-68

[29] Loewen, E. G. and M. C. Shaw, On the Analysis of Cutting Tool Temperatures, Trans. ASME 71 (1954) pp.217-231

[30] Leone, W.C., Distribution of Shear-Zone Heat in Metal Cutting, Trans. ASME, 76 (1954) pp.121-125

[31] Nakayama, K.,Temperature Rise of Workpiece During Metal Cutting, Bull. Fac. Engng. Nat. Univ. Yokohama, 5 (1956) 1-

[32] Boothroyd, G.,Temperatures in Orthogonal Metal cutting, Proc. of the Instn. Mech. Engrs. 177, No.29 (1963) pp.789-810

[33] Weiner, J. H., Shear Plane Temperature Distribution in Orthogonal Machining, Trans. ASME 77 (1955) pp.1331-1341

[34] Rapier, A.C., Theoretical Investigationof the Temperature Distribution on Metal Cutting Temperature, British J Appl. Phys.,5 (1954) pp.400-405

[35] Dutt, R.P., and R. C. Brewer, On the Theoretical Determination of the Temperature Field in Orthogonal Machining, Int. J of Prod. Research, 4 (1964) pp.91-114

[36] Dawson, P. R., and S. Malkin, Inclined Moving Heat Source Model for Calculating Metal Cutting Temperatures, Trans. ASME, 106 (1984) pp.179-186

[37] Barrow, G. A., Review of Experimental and Theoretical Techniques for Assessing Cutting Temperature, Annals of CIRP, 22/2 (1973) pp.203-211

[38] Komanduri, R, Hou, Z. B., Thermal Analysis of Manufacturng Processes, (a monograph under preparation) Oxford University Press (2001)

[39] Shaw, M.C.,Metal Cutting Principles, Oxford University Press, Oxford, (U.K.) 1984

APPENDIX A - FUNCTIONAL ANALYSIS

To develop functional relationships between the heat intensity and the ratio l_i/L on the chip side (of a moving band heat source) and on the tool side (of a stationary rectangular heat source) at the tool-chip interface for conventional machining of steel at high Peclet numbers ($N_{Pe} \approx 5 \sim 20$), it is firstly necessary consider the local heat partition fraction B_i and the ratio l_i/L. The simplest equation over uniform distribution is a superimposed simple linear function as given by Eq.(A1) (for the chip side).

$$B_{i,chip} = (B_{chip} - \Delta B) + 2\Delta B \frac{l_i}{L} \qquad (A1)$$

Figure A1 is a plot of this equation. Here B_{chip} is the average heat partition fraction for the chip obtained by matching the average temperature rise at the contacting interface on both sides. ΔB is the maximum compensation of B_{chip} at the two ends of the interface heat source. ΔB enables shifting of the location of the maximum temperature rise point and to changing the form of the temperature distribution curve. Thus, by appropriate selection of the value of ΔB, the location of the maximum temperature rise point of both sides can be coincided.

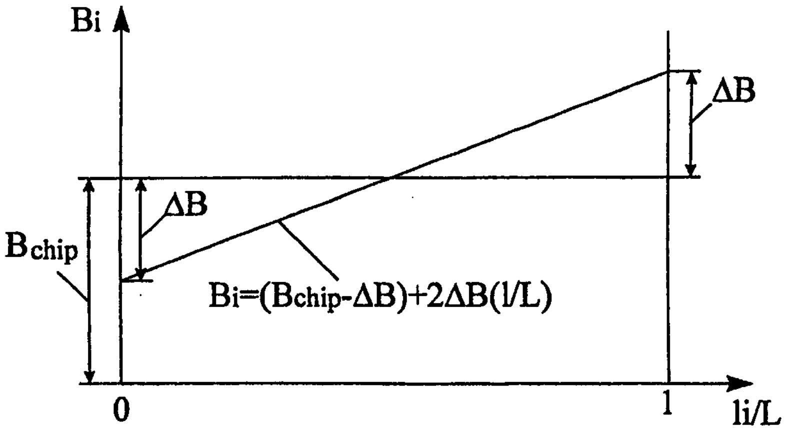


Figure A1 - Variation of the heat partition B_i with l_i/L using a simple linear function of the local partition fraction B_i, chip in terms of l_i/L.

For the tool side, a similar linear function can be considered. Thus

$$B_{i,tool} = (B_{tool} + \Delta B) - 2\Delta B \frac{l_i}{L} \qquad (A2)$$

Here, B_{tool} is the average partition fraction of the heat for the tool obtained by matching the average temperature rise at the contacting interface on both sides. $B_{tool} + B_{chip} = 1$.

Substituting the variable heat intensity for the chip side heat source, $B_{chip} * q_{pl}$, into Eq.(8) of the text, an equation for the temperature rise θ_M for a moving band heat source with variable heat intensity can be obtained as:

$$\theta_M = \frac{q_{pl}}{\pi\lambda} \left\{ \begin{array}{l} (B_{chip} - \Delta B) \int_{l_i=0}^{L} e^{-(X-l_i)v/2a} [K_0(R_i v/2a) + K_0(R_i' v/2a)] dl_i \\ +2\Delta B \int_{l_i=0}^{L} (\frac{l_i}{L}) e^{-(X-l_i)v/2a} [K_0(R_i v/2a) + K_0(R_i' v/2a)] dl_i \end{array} \right\} \qquad (A3)$$

Similarly, substituting the variable heat intensity for the tool side heat source, $B_{i,tool} * q_{pl}$, into Eq.(9) of the text, an equation for the temperature rise θ_M for a stationary rectangular heat source with variable heat intensity can be obtained as:

$$\theta_M = \frac{q_{pl}}{2\pi\lambda} \left\{ \begin{array}{l} (B_{tool} + \Delta B) \int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{1}{R_i} + \frac{n}{R_i'}) dx_i \\ +2\Delta B \int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})(\frac{1}{R_i} + \frac{n}{R_i'}) dx \end{array} \right\} \qquad (A4)$$

The results of the computation, using Eqns.(*A3*) and (*A4*) for the above example are shown in Table A1 and Figure A2.

Table A1 – Temperature rise distribution at the chip-tool interface considering variable heat intensity (using Eqns. A1 and A2) [5].

No. of point	Tool side	Chip side
0	496.1	508.1
1	578.5	539.9
2	573.2	535.9
3	545.3	526.9
4	503.9	512.1
5	454.6	490.4
6	401.5	460.2
7	348.6	419.0
8	299.5	361.9
9	259.1	277.1
10	235.3	64.1
Average temperature rise $\bar{\theta}$	426.9	426.9
Heat partition fraction	0.1245	0.8755

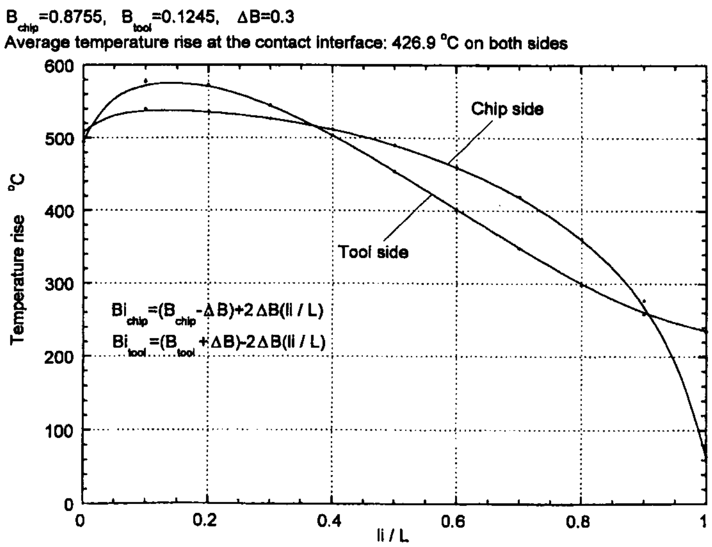


Figure A2 - Temperature rise distribution at the tool-chip interface considering a simple linear function for the non-uniform distribution of heat partition fraction.

Comparing the results shown in Table 1 and Figure 12 of the main text with the results in Table A1 and Figure A2, it can be seen that the temperature distribution at the interface on both sides are closer when a linearly variable heat intensity is considered but the difference is still quite significant.

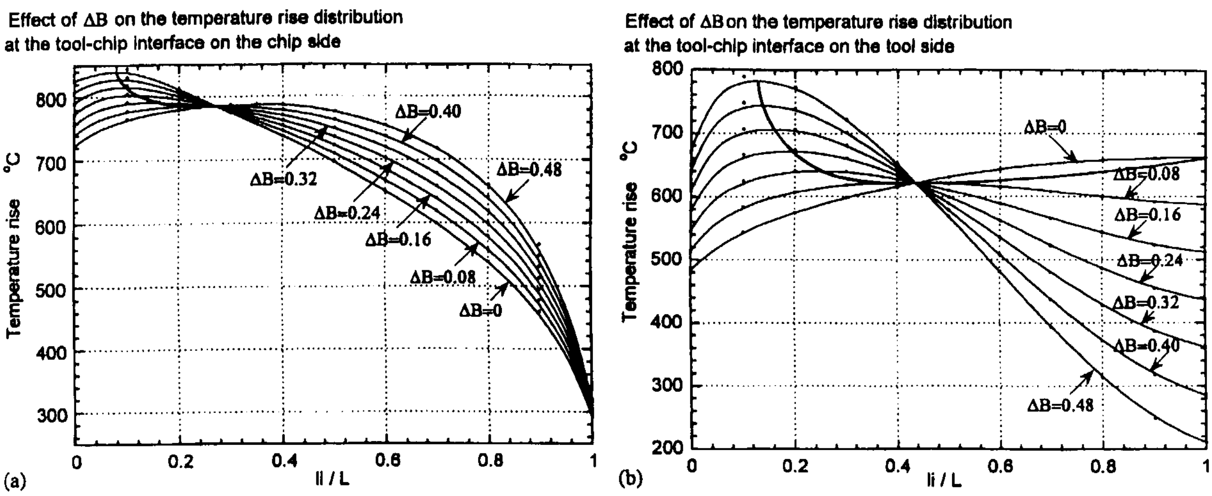
Figures A3 (a) and (b) are the plots of temperature distribution at the interface on the chip and the tool sides for different values of ΔB, calculated using the equations for the heat sources of linearly variable heat intensity, Eqns.(*A3*) and (*A4*). The effect of ΔB on the form of the temperature distribution curve and the location of the maximum temperature rise point can be seen clearly on these graphs. The dark line shows the location of the maximum temperature points for different values of ΔB. It can also be seen that the location of the maxima of the temperature distribution curve for a moving heat source on the chip side and that for a stationary heat source on the tool side can be shifted to coincide when ΔB is $\sim$0.3.

It can be seen from Figure A2 that for the temperature rise distribution curve at the interface on the chip side, the rate of increasing of the temperature rise seems too low when the value of the ratio l_i / L is small (in the range of 0 to 0.2). Consequently, the equation for the variable partition fraction of heat for the chip, Eqn.(*A1*), has to be modified as follows.

$$B_{i,chip} = (B_{chip} - \Delta B) + 2\Delta B(\frac{l_i}{L})^m \qquad (A5)$$

where m is a number < 1. It is found, when m is ~0.22-0.26, a satisfactory matching of the two temperature rise distribution curves at the interface on both sides (chip side and tool side) for machining steel under conventional cutting conditions (at high Peclet numbers) can be obtained. Based on Eq.(*A5*), the equation for the variable partition fraction of heat for the tool can be obtained using the relationship, $B_{i,tool} = 1 - B_{i,chip}$. Thus

$$B_{i,tool} = (B_{tool} + \Delta B) - 2\Delta B(\frac{l_i}{L})^m \qquad (A6)$$



Figures A3 (a) and (b) Plots of temperature rise distribution at the tool-chip interface considering a simple linear function for the non-uniform distribution of heat partition fraction for various values of on the chip and the tool sides for different values of ΔB.

Using Eqns.(*A5*) and (*A6*), the equations for the moving band heat source (for the chip) and the stationary rectangular heat source (for the tool) with variable heat intensity, Eqns.(*A3*) and (*A4*) can be further modified as follows.

For a moving band heat source with a modified function of variable heat intensity [see Eq.(*A5*)], the solution for the temperature rise θ_M is given by

$$\theta_M = \frac{q_{pl}}{\pi\lambda}\left\{ \begin{array}{l} (B_{chip} - \Delta B)\int_{l_i=0}^{L} e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \\ \\ +2\Delta B\int_{l_i=0}^{L} (\frac{l_i}{L})^m e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \end{array} \right\} \qquad (A7)$$

For a stationary rectangular heat source with a modified function of variable heat intensity [as Eq.(*A6*) shows], the solution for the temperature rise θ_M is given by

$$\theta_M = \frac{q_{pl}}{2\pi\lambda}\left\{ \begin{array}{l} (B_{tool} + \Delta B)\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{1}{R_i} + \frac{n}{R_i'})dx_i \\ \\ -2\Delta B\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L} (\frac{x_i}{L})^m (\frac{1}{R_i} + \frac{n}{R_i'})dx \end{array} \right\} \qquad (A8)$$

The results of computation using Eqns.(*A7*) and (*A8*) with m = 0.26 for above example are shown in Figure A4. It shows that by incorporating power functions [Eqns.(*A5*) and (*A6*)] with the above modifications instead of a simple linear function [Eqns.(*A1*) and (*A2*)] of variable intensity makes the temperature rise distribution curves at the contact interface on both sides much closer in the range of $l_i / L = 0 \sim 0.4$. However, in the range of $l_i / L = 0.4 \sim 1.0$, the difference of the two temperature distribution curves is increased. This can, however, be compensated by an additional term $C(l_i / L)^k$ with higher power. When k is very large (k>>1), this additional term will only influence those points where $l_i / L \to 1$, the smaller the value of l_i / L the smaller the effect of the third term. As for the points where

$l_i / L < 0.4$, the effect of the additional term would be negligibly small. The following equations are found to be capable of yielding very close results.

$$B_{i,chip} = (B_{chip} - \Delta B) + 2\Delta B(\frac{l_i}{L})^m + 2\Delta B(\frac{l_i}{L})^{16} \tag{A9}$$

$$B_{i,tool} = (B_{tool} + \Delta B) - 2\Delta B(\frac{l_i}{L})^m - 2\Delta B(\frac{l_i}{L})^{16} \tag{A10}$$

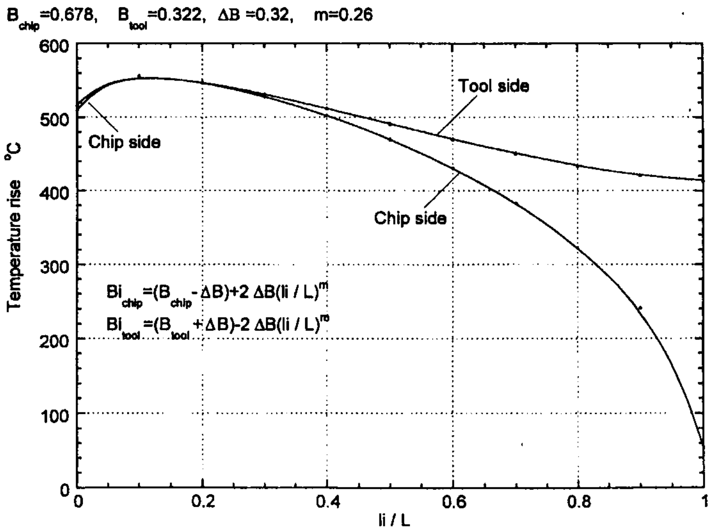


Figure A4 Temperature rise distribution at the tool-chip interface considering a power function for the non-uniform distribution of heat partition fraction.

Using Eqns.($A9$) and ($A10$) and substituting $B_{i,chip} * q_{pl}$ and $B_{i,tool} * q_{pl}$ as q_{pl} into the Eqns.(8) and (9) of the main text, a pair of well matched equations for moving band and stationary rectangular heat sources with variable heat intensity can be obtained as follows.

For a moving band heat source with a modified function of variable heat intensity [see Eqn.(A9)], the solution for the temperature rise θ_M is given by

$$\theta_M = \frac{q_{pl}}{\pi\lambda}\left\{\begin{array}{l} (B_{chip} - \Delta B)\int_{l_i=0}^{L} e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \\ \\ +2\Delta B\int_{l_i=0}^{L}(\frac{l_i}{L})^m e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \\ \\ +2\Delta B\int_{l_i=0}^{L}(\frac{l_i}{L})^{16} e^{-(X-l_i)v/2a}[K_0(R_i v/2a) + K_0(R_i' v/2a)]dl_i \end{array}\right\} \tag{A11}$$

For a stationary rectangular heat source with modified function of variable heat intensity [see Eqn.($A10$)], the solution for the temperature rise θ_M is given by

$$\theta_M = \frac{q_{pl}}{2\pi\lambda}\left\{\begin{array}{l} (B_{tool} + \Delta B)\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L}(\frac{1}{R_i} + \frac{n}{R_i'})dx_i \\ -2\Delta B\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L}(\frac{x_i}{L})^m(\frac{1}{R_i} + \frac{n}{R_i'})dx \\ -2\Delta B\int_{y_i=-b_o}^{+b_o} dy_i \int_{x_i=0}^{L}(\frac{x_i}{L})^{16}(\frac{1}{R_i} + \frac{n}{R_i'})dx \end{array}\right\} \tag{A12}$$

Figure A5 shows the results of the computation for the same example using Eqns.(A11) and (A12) with m=0.26. The difference between the two temperature distribution curves is < 5%.

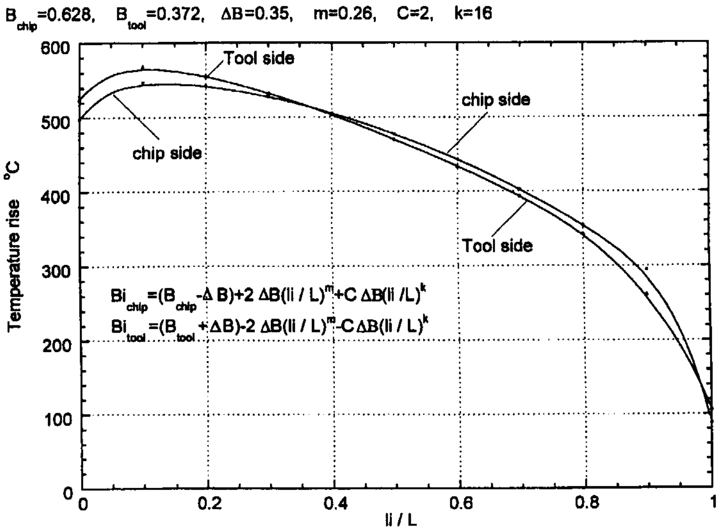


Figure A5 - Temperature rise distribution at the tool-chip interface considering a power function with a third term for the non-uniform distribution of heat partition fraction, m = 0.26, C = 2; k = 16.

BIOGRAPHY

Dr. Ranga Komanduri is a Regents Professor and holds the A.H. Nelson, Jr. Endowed Chair in Engineering in the School of Mechanical & Aerospace Engineering, Oklahoma State University, Stillwater, OK. He graduated from Osmania University, Hyderabad in India [both B.E. (Mechanical) and M.E. (Heat Power)]. He earned Ph.D. and D. Engg. from Monash University in Melbourne, Australia. He joined the Mechanical Engineering Department of Carnegie-Mellon University, Pittsburgh in 1972 interacting with Professor Milton Shaw on various aspects of grinding and tribology. In 1977, he joined the Corporate R&D of General Electric in Schenectady, NY as a member of the scientific staff. He was involved in high-speed machining, ultra-precision machining, and tool materials. He was also an Adjunct Full Professor at Rensselaer Polytechnic Institute in Troy, NY. He was on sabbatical for about 3 years at the National Science Foundation in Washington D.C. where he was Program Director for Materials Processing, Tribology, and Manufacturing Processes programs as well as Deputy and Acting Division Director of the Division of Design Manufacturing and Computer Integrated Manufacturing (DMCI). He joined OSU in 1989. His researech interests are in the areas of advanced manufacturing processes and materials. In particular, machining; grinding; high-speed machining, finishing of advanced ceramics, thermal aspects of manufacturing; hard coatings on cutting tool materials, such as diamond coatings and multiple nano-coatings; and molecular dynamics simulation of nanometric cutting and tribology; and laser assisted machining. He has published about 140 technical papers and holds 21 U.S. Patents. He is Fellow of the ASME and the SME, and an Active Member of CIRP (International Institution for Production Engineering Research). His awards include the F. W. Taylor Medal of CIRP, the ASME Blackall Machine Tool and Gage Award, and the ASME Pi Tau Sigma's Charles Russ Richards Memorial Award.

FRACTURE OF FUNCTIONALLY GRADED MATERIALS

Speaker: Glaucio H. Paulino, Associate Professor
Burton and Erma Lewis Faculty Scholar
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign

Abstract:

Functionally graded technology refers to techniques to functionally grade the mechanical, electronic, chemical, optical, and/or biological material properties. This presentation focuses on the effects of introducing compositional gradients in engineered materials from a thermo-mechanical point of view with special emphasis on fracture mechanics.

For metal/ceramic functionally graded materials (FGMs), cracks originating in the brittle side should be prevented from catastrophic failure by increasing fracture toughness as the material increases in metal content. These properties have been demonstrated in a titanium/titanium monoboride (Ti/TiB) FGM for which an average resistance curve (R-curve) has been obtained. Fracture testing of ductile/brittle FGMs single edge notched beams (SENB) with the starting notch in the ceramic-rich region requires the initiation of a sharp short crack at a specified desired location in the brittle region. Thus, a novel technique is presented to generate sharp cracks in metal/ceramic FGMs by reverse 4-point bending.

An investigation is made of FGMs that are elastically homogeneous but thermally nonhomogeneous. A transient thermal stress analysis for an edge crack in an FGM is performed. A multilayered material model is used to solve the temperature field. By using the Laplace transform and an asymptotic analysis, an analytical first order temperature solution is obtained. Thermal stress intensity factors are obtained for ceramic/ceramic FGMs such as TiC/SiC and MoSi_2/SiC.

Modeling of crack problems in ceramic/ceramic FGMs using gradient elasticity is also addressed in this presentation. The crack boundary value problem is solved by means of the Fourier transform and hypersingular integral equations of the Fredholm type. The solution naturally leads to a cusping crack. Examples comparing the results of various types of material microstructure (described by characteristic lengths) and gradation (described by moduli functions) are provided.

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

GRAIN REFINING MECHANISMS OF CAST Mg-Al-Zn ALLOYS

Tetsuichi Motegi
Department of Metallurgical Engineering
Chiba Institute of Technology
Narashino, Chiba-ken, 275-0016, Japan
TEL 81-474-78-0313
FAX 81-474-78-0329
E-mail: motegi@pf.it-chiba.ac.jp

Abstract: *The grain refining mechanisms and technology by superheat treating process, by addition of pure carbon or and niobium were proposed. Quenching technique was applied to clarify the grain refining mechanism of superheat treated samples. In each grain, one Al_4C_3 compound existed as a nucleation site. The alloy made by high-purity magnesium and aluminum refined the grain size without superheat treating because the impurity C existing in 99.99%Mg reacted with Al and formed Al_4C_3 in the molten alloy. As pure carbon powder or Nb_2O_5 were added into the molten alloy, the grain refining occurred effectively. The carbon powder reacted with the molten aluminum, and formed Al_4C_3. On the other hand, as the Nb_2O_5 was added in the molten alloy, Nb_3Al or some niobium compounds were formed and acted as nucleation sites. In this case, Al-Mn compounds became smaller and granular.*

Keywords: grain refining ,magnesium alloy, solidified structure, superheat treatment, grain refiner

1. Introduction

Grain refinement of cast structures in castings and ingots are very important to improve their mechanical properties and to reduce various casting defects. Grain refinement of cast magnesium alloys containing Al component is attained by superheat treating or by adding carbon materials. Both techniques are well established, but the grain refining mechanisms have not been clarified yet. Several theories have been proposed. One of these is heterogeneous nucleation due to Al-Mn or Al-Fe-Mn compounds that are formed during superheat treatment (Fox et al, 1940; Tiner, 1945). Another is the formation of Al_4C_3 in the molten magnesium alloy by existence of the impure carbon (Kaufmann , 1961). An important problem that has recently arisen is that C_2Cl_6 used as the grain refiner generates harmful dioxin materials. Therefore it is impossible to use the grain refiners containing chloride. In this study, the grain refining mechanism by superheat treating of Mg-Al-Zn alloys and new grain refiners, pure carbon or Nb, are introduced.

2. Experimental procedures

2.1 Quenching of superheat-treated Mg-Al-Zn alloy

An AZ91E magnesium alloy was used in this experiment. Table 1 shows chemical compositions of the AZ91E magnesium alloy. The alloy was superheat-treated at 1123 K for 900 s. It was then cooled at 2.5 K/s and cast at 873, 923, 973, and 1023 K, respectively. During pouring, the molten alloy was quenching by using two chilled copper blocks. Samples without superheat-treatment were also quenched at 1023 K to compare with the superheat treated one.

Table 1 Chemical compositions of AZ91E magnesium alloy. (mass%)

Al	Zn	Mn	Si	Cu	Ni	Fe	Mg
8.5	0.71	0.23	0.001	0.001	0.001	0.002	Bal.

2.2 Grain refinement of Mg-Al alloy made from high purity metals

Mg-Al alloys containing 5.6 to 14.1% Al were alloyed by melting 99.99% Mg and 99.999% Al in order to avoid the influence of the impurities on the microstructures. Table 2 shows impurities in the distilled 99.99% Mg. Mixed SF_6 and CO_2 gases were used to prevent burning of the molten alloy during melting and cooling. The samples were also superheat-treated at 1123K for 900 s, then cooled at 2.5K/s and cast into the copper mold. Samples without superheat treatment were also cast into the same one.

Table 2 Impurities in the distilled 99.99% Mg. (mass%)

Al	Zn	Mn	Si	Cu	Fe	Ni	C	Mg
0.0010	0.0005	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0007	Bal.

2.3 Grain refinement by adding pure carbon powder

In this experiment, high-purity carbon powder was added to molten AZ91E alloy of 700g using an argon gas carrier. 5 micron meters in diameter of carbon powder was charged into a tank, and argon gas that imposed pulsating motion flowed in the tank. The pulsating motion of argon gas agitated the carbon powder grains and carried into the molten alloy. In order to examine the optimum temperature for addition, carbon powder was added at 973, 993, 1053, and 1073 K for 600s. To examine the optimum time for addition of the carbon powder, carbon was added for 300, 600, 900, 1200, 1500,and 1800 s at 1023, 1053, and 1073K. After carbon was added, the molten alloy was poured into the copper mold. In the experiments at 1053K and 1073 K for more than 120 s, the molten alloy burned and the experiments were impossible.

2. 4 Grain refinement by adding Nb

Nb_2O_5 was added into the molten magnesium alloy, because the melting point of niobium is very high. So the molten magnesium could reduce easily Nb2O5. In order to examine the optimum amounts of niobium, 0.1, 0.2, 0.5, and 1.0% of Nb_2O_5 were added in the molten alloys of 993, 1013, 1033, 1053, and 1073 K. A phosphorizer was used to add the Nb_2O_5 tablets into the molten alloy. The molten alloy was covered with a mixture gas of SF_6+CO_2 to prevent the burning. All casting temperatures after addition of Nb_2O_5 were cast at 973 K into the copper mold.

2.5 Microscopic observation and EPMA analysis

All samples obtained were observed through a microscope, and a quality analysis was performed by an EPMA. Grain sizes of the samples that were cast in a chilled mold were measured.

3. Results and discussion

3.1 Solidified structures of quenched samples at various temperatures

Figure 1 shows the microstructures of superheat treated and untreated AZ91E alloy samples that were obtained by quenching at 873 K. The superheat treated samples show that foreign substances existed in the center of dendrite crystals grown radially .

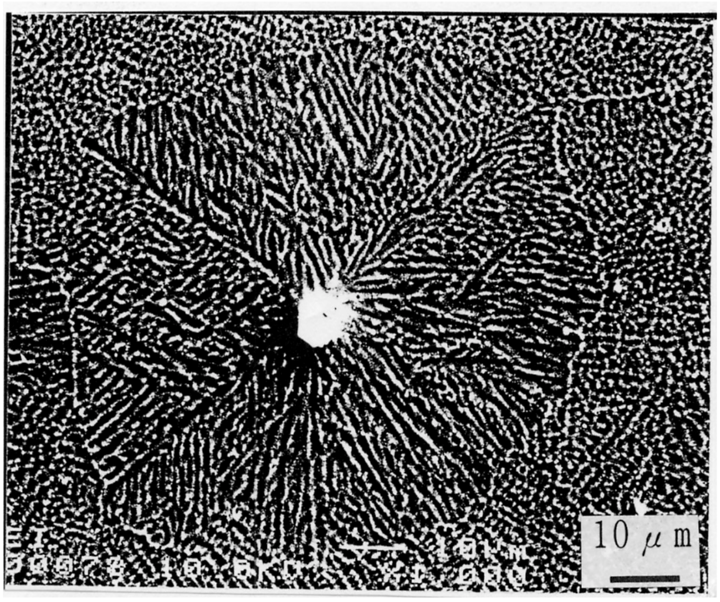


Figure 1. Al_4C_3 compounds appeared by quenching at 1023 K.

The results of analysis by the EPMA reveal that the foreign substance consisted of Al, C, and O or Al, C, Mn, Si, and O. C, Mn, and Si were impurities in the AZ91E alloy. It is estimated that an Al_4C_3 or Al_2CO compound was probably formed during superheat treatment. However, O in the Al_2CO might have been introduced by polishing the sample because water was used for polishing and the following chemical reaction occurred.

$$Al_4C_3 + 12H_2O = 3CH_4 + 4Al(OH)_3 \tag{1}$$

The samples quenched at 873, 923, and 973 K reveal that Al, Mn, C, O and Si elements coexisted. Mn and Si free Al-C-O compounds appeared only in the sample of 1073 K. Al, Mn, and Si were probably soluble in the molten magnesium during superheat treating, then Mn and Si crystallized on the Al_4C_3 compounds or only Mn compounds formed below 973 K. In contrast, the compounds consisting of Al, Mn, and Si also existed in the samples without superheat treatment, but the compounds did not nucleate the magnesium dendrite crystals.

3.2 Grain refining of Mg-Al alloy made by high-purity metals

In general, commercial magnesium alloys contain various impurities, especially Mn and Fe. Therefore, it is difficult to clarify the grain-refining mechanisms. In this study, high-purity metals of magnesium and aluminum were used to eliminate the influence of impurities on the microstructures. The high purity alloy reveals that the grain refining occurred without superheat treatment, but this was not so in the commercial alloy.

Figure 2 shows the influence of superheat treatment and the aluminum content on the grain size. In the commercial alloy, the grain refinement occurred with increasing aluminum content, and the superheat treatment decreased the grain size. The high-purity alloy appeared to produce fine grains independent of superheat treatment and aluminum content. It seems that numerous nucleation substrates of Al_4C_3 formed in the high purity alloy. The Al_4C_3 compound can nucleate magnesium crystals as described above. I conclude that the same nucleation phenomenon that occurs in the AZ91 alloy also occurs in the high-purity alloy in spite of the lack of superheat treatment.

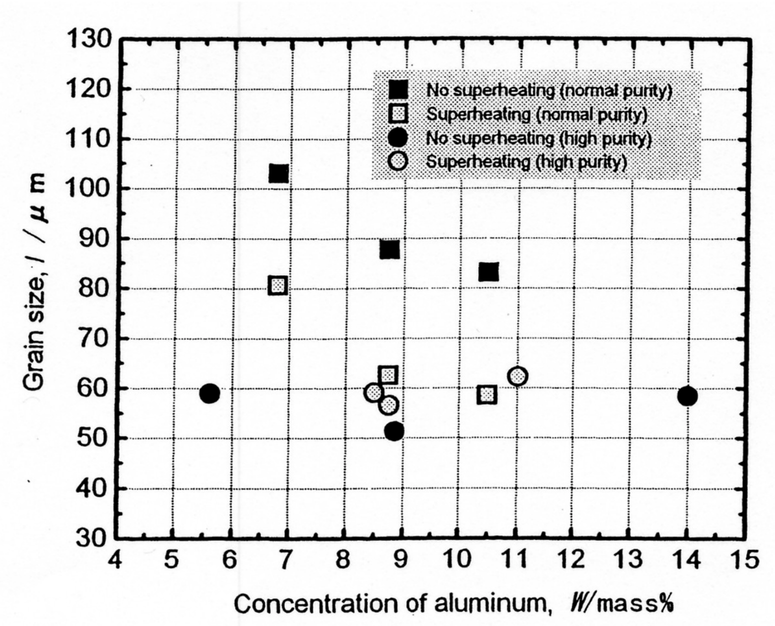


Figure 2. Influence of superheat treatment and Al content on the grain size.

3.3 Grain refinement by adding pure carbon powder

It is clear that Al_4C_3 is very effective nucleant for Mg-Al system alloys. In this study, high-purity carbon powders were used instead of C_2Cl_6. Figure 3 shows the influence of adding temperature and time on grain size. Both results show that the optimum time for adding carbon powders at 1023 K was 1800 s, and times at 1053 K and 1073 K were 600 and 300 s. Finer grain sizes were obtained for higher additive temperatures.

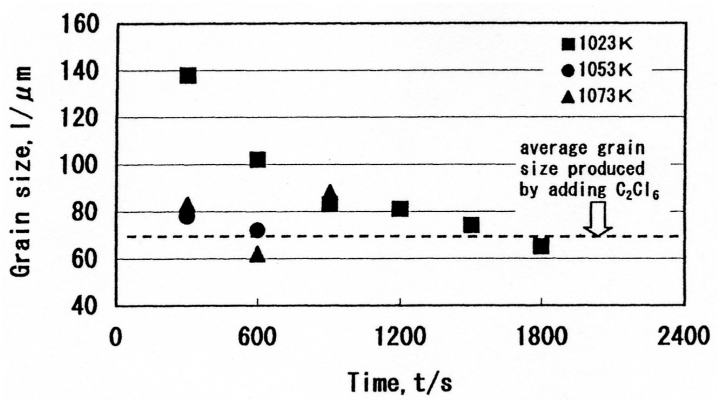


Figure 3. Influences of adding times and temperatures of carbon powder on grain size.

3.4 Grain refinement by adding Nb

The higher adding temperature of Nb_2O_5, the finer grain size was obtained because the reduction rate of Nb_2O_5 due to molten magnesium was larger. Figure 4 shows the typical microstructures of cast structures added at 1073 K. In addition, manganese compounds existing in the matrix became fine and round shape as shown in Fig.5.

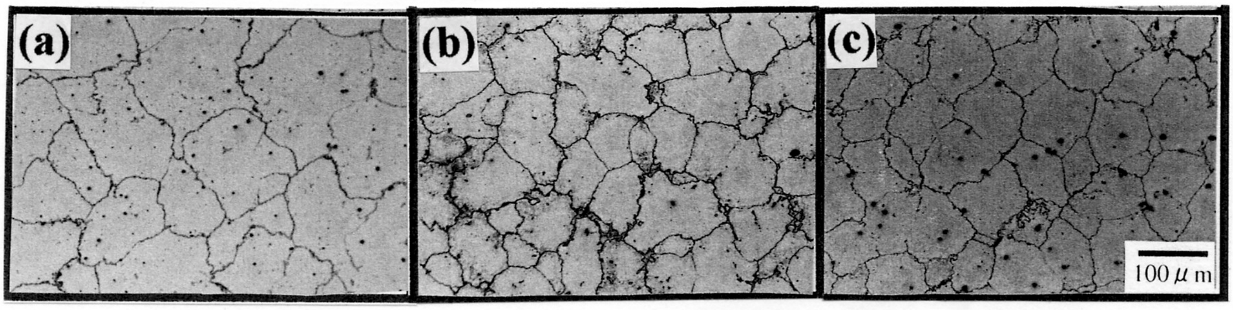


Figure 4. Influence of Nb_2O_5 contents and temperatures on microstructures of AZ91E alloy. (a) 1.0% at 1033 K, (b) 0.2% at 1073 K, (c) 1.0% at 1073 K.

In addition, manganese compounds existing in the matrix became fine and granular shapes as shown in Fig. 5.

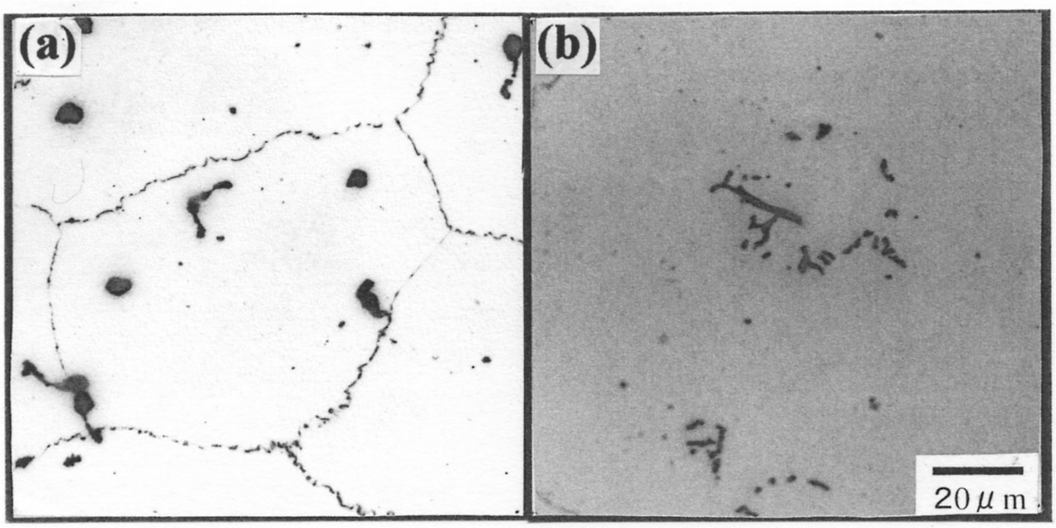


Figure 5. Shapes of manganese compounds by addition and non addition of Nb.

The grain refining mechanism by addition of Nb_2O_5 has not clarified yet. Nb_3Al or NbC compounds as nucleation sites were formed in the molten alloy and nucleation of magnesium occurs during solidification.

4. Conclusions

In order to clarify the grain refining mechanisms of Mg-Al system alloys by superheat treatment and two kinds of grain refiners, carbon and niobium were added into the molten alloy for grain refiners. The results obtained were as follows.

(1) In the commercial alloy of AZ91, Al_4C_3 heterogeneous nucleant formed as the superheat treatment was performed.
(2) In the Mg-Al alloy made of 99.99%Mg and 99.999%Al, grain refining occurred without superheat treatment.
(3) Pure carbon powder is an effective replacement for C_2Cl_6 in Mg-Al system alloys.
(4) Nb_2O_5 is an active grain refiner for Mg-Al system alloys.

5. References

Fox, F.A.; Lardner, E. Jr., 1940, J. Inst. Metals 71, p.1.
Kaufmann,V.B., 1961, Trans.AIME 22, p.540.
Tiner, N., 1945, Metals Technology 12, p.13.

MICROFABRICATION OF POLYMER-COMPONENTS AND SYSTEMS.

Saile, Volker
Institute for Microstructure technology
Karlsruhe – Germany

XVI CONGRESSO BRASILEIRO DE ENGENHARIA MECÂNICA
16th BRAZILIAN CONGRESS OF MECHANICAL ENGINEERING

NEW DEVELOPMENTS IN LASER DOPPLER VIBROMENTER OPTICAL SYSTEMS AND DEMODULATION SCHEMES FOR MEASUREMENTS ON MEMS AND OTHER MICRO STRUCTURES

Johansmann, Martin
Division Laser Measurement Systems – Polytec GmbH
Waldbronn – Germany