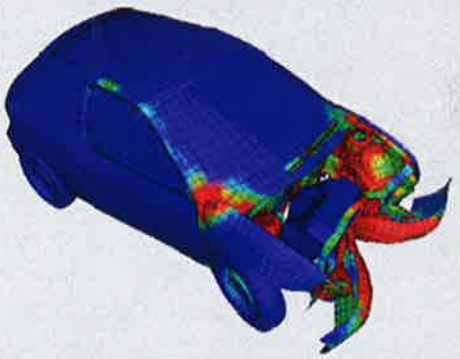


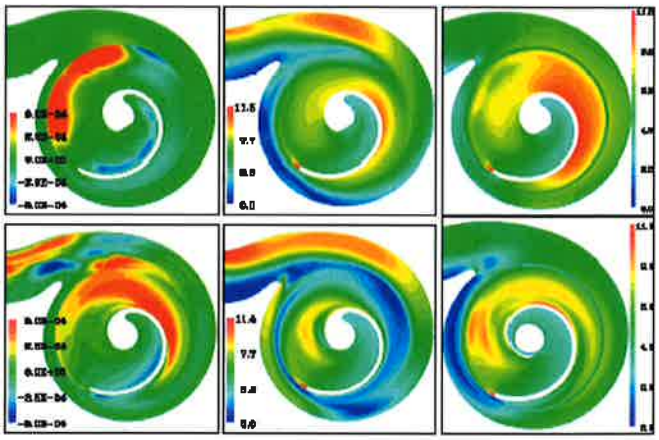
Raimo von Hertzen ja Tapani Halme (toim.)

PROCEEDINGS OF THE IX FINNISH MECHANICS DAYS

IX Suomen Mekaniikkapäivät

Lappeenrannan teknillinen yliopisto
13.-14. kesäkuuta 2006

2. Kokouspäivän esitelmät



LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

LAPPEENRANNAN TEKNILLINEN YLIOPISTO
KONETEKNIIKAN OSASTO

LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF MECHANICAL ENGINEERING

RAPORTTI 17
REPORT

Raportti 17
Report 17

Proceedings of the IX Finnish Mechanics Days
IX Suomen Mekaniikkapäivät
Lappeenrannan teknillinen yliopisto
13.-14. kesäkuuta 2006

Nide 2

2. kokouspäivän esitelmät

Toimittajat

Raimo von Hertzen

Tapani Halme

ISBN 952-214-227-1
ISSN 1459-2924

Lappeenrannan teknillinen yliopisto
Konetekniikan osasto/Koneensuunnittelun laitos
Rakenne- ja lujuustekniikan laboratorio
PL 20, FIN-53851 Lappeenranta
puh. 05-62 111

LTY digipaino 2006

Tieteellinen neuvosto

Tapani Halme	Lappeenrannan teknillinen yliopisto
Heikki Handroos	Lappeenrannan teknillinen yliopisto
Raimo von Hertzen	Lappeenrannan teknillinen yliopisto
Juhani Koski	Tampereen teknillinen yliopisto
Gary Marquis	Lappeenrannan teknillinen yliopisto
Aki Mikkola	Lappeenrannan teknillinen yliopisto
Timo Nykänen	Lappeenrannan teknillinen yliopisto
Rolf Stenberg	Teknillinen korkeakoulu
Stig-Göran Sjölind	Oulun yliopisto

Järjestelytoimikunta

Tapani Halme
Sami Heinilä
Raimo von Hertzen
Lauri Huhtala
Matti Koskimäki
Janne Lepistö
Ilkka Poutiainen
Riitta Ruokonen
Kaija Tammelin
Pasi Tanskanen

Lappeenrannan teknillinen yliopisto
Konetekniikan osasto
Rakenne- ja lujuustekniikan laboratorio
PL 20
53851 LAPPEENRANTA

Puh. (05) 621 2436
Fax. (05) 621 2499

ISBN 952-214-227-1
ISSN 1459-2924

SISÄLLYSLUETTELO

Rakenteiden mekaniikka I

Virtausmekaniikka I

Kokeellinen mekaniikka

Rakenteiden mekaniikka II

Fluidin ja rakenteen vuorovaikutus

Optimointi

Rakenteiden mekaniikka III

Väsyminen ja murtuminen I

NIDE 2

Kutsutut esitelmät

Numeeriset menetelmät II

Virtausmekaniikka II

Dynamiikka ja värähtelyt II

Mekaniikan erityisongelmia

Rakenteiden mekaniikka IV

Materiaalien käyttäytyminen II

Väsyminen ja murtuminen II

SMART SENSORS FOR RAILCAR MONITORING

Darrell Socie
Mechanical Science and Engineering Department
University of Illinois at Urbana-Champaign

ABSTRACT

Ubiquitous computing and smart sensor technology has seen tremendous growth over the last five years. The latest generation of very low-power microprocessors, low power spread spectrum radio transceivers, and networking software has driven a new round of smart sensor research, development, and commercialization. A smart sensor is distinguished by the ability to not only sense and collect data, but to process, reduce and make decisions based on its own data and data from the surrounding sensor network. The multi-sensor network is a key aspect of ubiquitous computing and promises to significantly lower the barriers that afflict current long-term, in-field monitoring systems. Those barriers include: 1) high installation and maintenance costs due to placement of sensors and routing of power/data lines 2) limited time duration due to limited battery life and 3) penalties associated with the physical size and weight of the data collection system. To be practical the whole system must be designed to totally require only a few milliwatts of power and be able to be produced in large quantities for less than €100 each.

The barriers described above establish the need for energy harvesting solutions. Smart sensors require about 10 mW of power. A visible lack of energy harvesting solutions for smart sensor systems exists today. Until recently, there were very few applications where 10 mW of power was useful. There are many different sources of energy available from the ambient environment. These sources include solar, vibration, thermal, and wind. Each of these sources has the potential to provide the 10 mW of energy needed to power a smart sensor indefinitely. In addition, the relatively small footprint of smart sensors, less than 10 cm^3, adds an additional constraint to the energy generation mechanism.

This paper reviews available power sources focuses on the design trade-offs associated with a linear electromagnetic energy harvester. The main constraint placed on the harvester is the overall size. Because smart sensors are themselves small, the energy harvesting device must also be correspondingly small. For this study, the size of the device is limited to a cylinder 2.5 cm in diameter and having a height of 2.5 cm for a total volume of about 10 cm^3. The goal is to maximize power output from the device by adjusting the following design parameters: 1) number of magnets 2) number of coils 3) orientation of magnets with respect to coils and 4) trading outer coils for steel casing.

TRANSITIOLÄMPÖTILAKRITEERIEN TAUSTA JA MERKITYS HAURASMURTUMISMITOITUKSEN KANNALTA

Kim Wallin
VTT
PL 1000, FIN-02044VTT
E-mail: kim.wallin@vtt.fi

Transitiolämpötilakriteerit kehittyivät toisen maailmansodan aikana ja pian sen jälkeen. Tällöin tajuttiin laajemmin, että rakenneterästen sitkeys riippuu lämpötilasta. Ilmiö oli toki jo aiemmin tunnettu, mutta siihen mennessä ei ollut mahdollista kehittää sitkeydelle kvantitatiivista arviota. Toisen maailmansodan aikana yleistyneet hitsatut laivat kokivat kuitenkin useita murtumia. Tämä tuotti laajan vertailuaineiston, jonka pohjalta kvantitatiivisia sitkeyden arviointikriteerejä lähdettiin kehittämään. Viidentoista seuraavan vuoden aikana kehitettiin tätä varten lukuisia koemenetelmiä, joiden avulla laivaterästen murtumiskäyttäytymistä pyrittiin selittämään. Eri kokeita yhdistämään kehittyi lämpötilaan perustuva vertailu. Tällöin kehittyivät transitiolämpötilakriteerit, jotka suurelta osin ovat käytössä tänäkin päivänä. Tänään, kuitenkin, on suurelta osin unohdettu kriteerien empiirinen tausta ja niiden rajoittuminen alun perin vanhanaikaisiin matalalujuuksisiin laivateräksiin. Tänä päivänä käytetyt teräkset poikkeavat ominaisuuksiltaan suuresti niistä teräksistä, joille transitiolämpötilakriteerit alun perin kehitettiin. Tämän takia on pakko arvioida uudella tavalla kriteerien merkitys haurasmurtumismitoituksen kannalta. Tämä voidaan tehdä murtumismekaniikan avulla.

Murtumismekaniikka, tieteenala jossa yhdistetään rakenteen jännitystila, mahdollinen vika ja materiaalin murtumisvastus, on kehittynyt valtavasti viimeisten kahdenkymmenen vuoden aikana. Tänä päivänä on siten mahdollista tutkia eri transitiolämpötilakokeiden merkitystä uudella tavalla ja näin ollen myös kriteerien merkittävyyden arviointi on mahdollista myös nykyisin käytössä oleville teräksille. Tänä päivänä yleisimmin käytetty transitiokäyttäytymistä mittaava koe, Charpy-V iskukoe, on mahdollista korrelaation avulla kytkeä materiaalin murtumissitkeyteen, jolloin rakenteen/materiaalin kvantitatiivinen murtumismekaaninen arviointi on mahdollista. Murtumissitkeyden ja Charpy-V iskukokeen välisiä korrelaatioita on toki ollut olemassa jo pidemmän aikaa, mutta vasta viime vuosina tapahtunut murtumismekaniikan kehitys on mahdollistanut luotettavien korrelaatioiden kehittämisen. Näitä uusimpia korrelaatioita hyödynnetään tänä päivänä vikojen kriittisyyttä arvioivissa menetelmissä, kuten BS7910. Uusimpana menetelmänä on kehitetty FITNET arviointimenetelmä, josta on valmistumassa "CEN workshop agreement" -dokumentti.

Tässä esitelmässä käydään läpi alkuperäisten transitiokriteerien kehitys ja esitetään niiden merkitys nykyisille teräksille, haurasmurtumismitoituksen kannalta, uusimman murtumismekaanisen tietämyksen pohjalta.

DESIGN OF ECONOMIC STEEL STRUCTURES

K. JÁRMAI
Professor at the Faculty of Mechanical Engineering, University of Miskolc,
H-3515 Miskolc-Egyetemváros, Hungary, altjar@uni-miskolc.hu

ABSTRACT

Design of economic steel structures is an important issue to be competitive. Optimum design, structural analysis, fabrication technologies and economy are in close connection is this process. We have developed a cost calculation system to calculate both material and different fabrication costs, like welding, cutting, grinding, painting, etc. The used optimization techniques can solve highly nonlinear problems. The newly developed evolutionary techniques are very robust. On a simple example we show the application of this kind of approach. The structure is a stiffened plate and we optimize it with different welding technologies.

1 INTRODUCTION

People in their everyday life always make optimization on a conscious or a subconscious way „to reach the best, which is possible with the resources available". The consciousness makes the act more efficient. They have always targets to reach and constraints to control them. The birth of optimization methods as mathematical techniques can be dated back to the days of Newton, Lagrange and Cauchy. The further development in optimization was possible by the developments of differential calculus by Newton, Leibnitz, the variational calculus by Bernoulli, Euler, Lagrange and Weierstrass, the introduction of unknown multipliers by Lagrange. The concept of multiobjective optimization was formulated one hundred years ago by Pareto in 1896.

The first written analytical work published on structural optimization was made by Maxwell in 1890, followed by the well-known work of Michell in 1904. These works provided theoretical weight minima of trusses, using highly idealised models, but the analytical way of solution of the structural optimization problem is still usable.

During the Second World War and in the late 1940's and the early 1950's the development of optimization concerned to the minimum weight design of aircraft structural components: columns, stiffened panels, subject to compressive loads and to buckling. Digital computers appear in the early 1950's and gave a strong impulse to the application of linear programming techniques. The applications were focused primarily on steel frame structures.

In the late 1950's and 1960's the applications of structural optimization on lightweight structures concentrated to the aircraft and space industries. This time a some new optimization techniques have been developed by works of Rosenbrock, Box, Powell. The great development of this period is that the finite element method, which is a powerful tool for analysis of complex structures, has been invented by Zienkiewich and applied by many others for structural analysis.

Modern structural optimization can be dated from the paper of Schmit in 1960, who drew up the role of structural optimization, the hierarchy of analysis and synthesis, the use of mathematical programming techniques to solve the nonlinear inequality constrained problems. The importance of this work is that it proposed a new philosophy of engineering design, the structural synthesis, which clarifies the methodology of optimization.

There are several international organizations which deal with optimisation, probably the largest of them is the International Society of Structural and Multidisiplinary Optimization (ISSMO) which has its own journal Structural and Multidisciplinary Optimization (Springer Verlag) and regularly organises conferences.
The International Institute of Welding (IIW) also deals with cost calculations (Journal Welding in the World), they have annual assembly each year. Also there are a great number of other conferences and courses which have been organized connected to these two main fields. At the University of Miskolc we have organised several conferences (Farkas & Jármai (Eds.) 1996, Jármai (Ed.) 1997, Jármai & Farkas (Eds.) 2003), also courses in CISM, Italy (Farkas & Jármai 1996, Jármai & Farkas (Eds.) 1999).

2 DESIGN VARIABLES, OBJECTIVE FUNCTIONS, CONSTRAINTS AND PREASSIGNED PARAMETERS

The objective function (more functions at multiobjective optimization), the design variables, the preassigned parameters and the constraints describe an optimization problem.

2.1 *Design variables and preassigned parameters*

The quantities, which describe a structural system can be divided into two groups: preassigned parameters and design variables. The difference between them is that the members of the first group are fixed during the design, the second group is the design variables, which are varied by the optimization algorithm. These parameters can control the geometry of the structures. It is the designer choice, which quantities will be fixed or varied. They can be cross-sectional areas, member sizes, thicknesses, length of structural elements, mechanical or physical properties of the material, number of elements in a structure (topology), shape of the structure, etc.

2.2 *Constraints*

Behaviour means those quantities that are the results of an analysis, such as forces, stresses, displacements, eigenfrequencies, loss factors etc. These behaviour quantities form usually the constraints. A set of values for the design variables represents a design of the structure. If a design meets all the requirements, it will be called feasible design. The restrictions that must be satisfied in order to produce a feasible design are called constraints. There are two kinds of constraints, explicit and implicit ones.

Explicit constraints
Explicit constraints which restrict the range of design variables may be called size constraints or technological constraints. These constraints may be derived from various considerations such as functionality, fabrication, or aesthetics. Thus, a size constraint is a specified limitation, upper or lower bounds on a design variable. Examples of such constraints include minimum slope of a portal frame structure, minimum thickness of a plate, minimum or maximum ratio of a box section height and width, etc.

Implicit constraints
Constraints derived from behaviour requirements are called behavioural constraints. Limitations on the maximum stresses, displacements, or local and overall buckling strength, eigenfrequency, damping are typical examples of behavioural constraints. The behaviour constraints can be regarded as implicit variables. The behavioural constraints are often given by formulae presented in design codes or specifications. Other part of the behavioural constraints are computed by numerical technique such as FEM. In any case the constraints can be evaluated by analytical

technique. From a mathematical point of view, all behavioural constraints may usually be expressed as a set of inequalities.

2.3 *Objective function*

In most practical cases an infinite number of feasible designs exists. In order to find the best one, it is necessary to form a function of the variables to use it for comparison of design alternatives. The objective function (also termed the cost, or merit function) is the function whose least, or greatest value is sought in an optimization procedure. It is usually a nonlinear function of the variables x, and it may represent the mass, the cost of the structure, or any other function, which extremum can give a possible and useful solution of the problem. The minimization of $f(x)$ is equivalent with the maximization of $-f(x)$.

3 DIVISIONS IN OPTIMIZATION TECHNIQUES

The different single-objective optimization techniques make the designer able to determine the optimum sizes of structures, to get the best solution among several alternatives. The efficiency of these mathematical programming techniques is different. A large number of algorithms has been proposed for the nonlinear programming solution Himmelblau (1972), Vanderplaats (1984), Schittkowski et al (1994), Snyman (2005). Each technique has its own advantages and disadvantages, no one algorithm is suitable for all purposes. The choice of a particular algorithm for any situation depends on the problem formulation and the user.

The general formulation of a single-criterion nonlinear programming problem is the following:

$$\text{minimize} \quad f(x) \quad x = \{x_1, x_2, \ldots, x_N\} \tag{1}$$

$$\text{subject to} \quad g_j(x) \le 0, \quad j = 1,2,\ldots,P \tag{2}$$

$$h_i(x) = 0 \quad i = P+1,\ldots,P+M \tag{3}$$

$f(x)$ is a multivariable nonlinear function, $g_j(x)$ and $h_i(x)$ are nonlinear inequality and equality constraints respectively.

The optimization models can be very different from each other.

- *Analytical and numerical*
- *Unconstrained and constrained*
- *Single- and multivariable*
- *Single- and multiobjective*
- *Discrete and nondiscrete*
- *Structure free and structure dependent techniques*
- *Single- and multilevel optimization*

Detailed description is given in (Farkas & Jármai 1997, Jármai & Iványi 2001, Farkas & Jármai 2003).

4 COST ELEMENTS

The cost of the structure can be calculated from the material and the various fabrication costs, where one should consider the cost differences between different technologies and also their effects on the structure. The fabrication cost at welded steel structures can be welding, flattening, cleaning, painting, cutting, grinding, etc. (Peurifoy 1975, Volkov 1978, Yeo 1983, Winkle 1986, Ramirez & Touran 1991). If one does not consider the costs elements directly, but the fabrication times, which are available for a given technology, it is easier to calculate later the real cost in a given country (Jármai & Farkas 1999).

4.1 *Material cost*

The material cost can be calculated as

$$K_m = k_m \rho V, \tag{4}$$

where K_m [in \$ or in any other currency] is the material cost, k_m [\$/kg] is the corresponding material cost factor, ρ [kg/mm^3] is material density, V [mm^3] is the volume of the structure. The range of k_m is between 0.5 – 1 [\$/kg] according to the producers' pricelists.

4.2 *Fabrication cost*

The fabrication cost can be expressed as (Pahl & Beelich 1992)

$$K_f = k_f \sum_i T_i , \tag{5}$$

where K_f [\$] is the fabrication cost, k_f [\$/min] is the corresponding fabrication cost factor, T_i [min] are production times. It is assumed that the value of k_f is constant for a given manufacturer. If not, it is possible to apply different fabrication cost factors simultaneously in Eq. (5).

4.2.1 *Fabrication times for welding*

The most important times related to welding are as follows: preparation, assembly, tacking, time of welding, changing the electrode, deslagging and chipping.

Calculation of the times of preparation, assembly and tacking

The times of preparation, assembly and tacking can be calculated with an approximation formula as follows

$$T_{w1} = C_1 \Theta_{dw} \sqrt{\kappa \rho V} , \tag{6}$$

where C_1 is a parameter depending on the welding technology (usually equal to 1), Θ_{dw} is a difficulty factor, κ is the number of structural elements to be assembled. Formula (6) can be approximately derived from Lihtarnikov (1968).

The difficulty factor expresses the complexity of the structure. Difficulty factor values depend on the kind of structure (planar, spatial), the kind of members (flat, tubular). The range of values proposed is between 1-4 (Farkas & Jármai 1997).

Calculation of real welding time

The welding technologies applied are given in Table 1. Real welding time can be calculated on the following way

$$T_{w2} = \sum_i C_{2i} a_{wi}^2 L_{wi} , \tag{7}$$

where a_{wi} is weld size, L_{wi} is weld length, C_{2i} and n are constants for different welding technologies. C_2 contains not only the differences between welding technologies but the time differences between positional (vertical, overhead) and normal welding as well (see Farkas&Jármai 2003).

Calculation of additional fabrication actions time

There are some additional fabrication actions to be considered such as changing the electrode, deslagging and chipping. The time of these is as follows

$$T_{w3} = \sqrt{\Theta_{dw}} \sum_i C_{3i} a_{wi}^2 L_{wi} \,. \tag{8}$$

Formulae (6,7,8) was proposed by Pahl & Beelich (1982) and used in (Farkas 1992, Farkas & Jármai 1993, Jármai & Farkas 1999).

Ott & Hubka (1985) proposed that C_3 = (0.2-0.4) C_2 on average $C_3 = 0.3C_2$. Thus, the modified formula for T_{w2}+T_{w3} neglecting $\sqrt{\Theta_d}$, is

$$T_{w2} + T_{w3} = 1.3 \sum C_{2i} a_{wi}^2 L_{wi} \,. \tag{9}$$

In the negligence of $\sqrt{\Theta_{dw}}$ it is assumed that the difficulty factor should be considered only for T_{w1}.

The software COSTCOMP (1990) was developed by the Netherlands Institute of Welding. It gives welding times and costs for different welding technologies (Bodt 1990) on the basis of theoretical and experimental investigations. Considering the times given by companies all over the world and the times calculated by COSTCOMP here Eq. (6) is used for T_{w1} and the other times are calculated with a generalized formula, where the power of a_w is n, which is some cases equal to 2, or close to it.

$$T_{w2} + T_{w3} = 1.3 \sum C_{2i} a_{wi}^n L_{wi} \,. \tag{10}$$

The different welding technologies are shown in Table 1. The weld types are given are as follows: fillet welds, V-, X-, T-, 1/2 V-, T-, U-, Y- butt welds and the double version of T-, U-, Y- butt welds and the K-butt weld.

Table 1. Welding technologies applied

SMAW	Shielded Metal Arc Welding
SMAW HR	Shielded Metal Arc Welding High Recovery
GMAW-C	Gas Metal Arc Welding with CO_2
GMAW-M	Gas Metal Arc Welding with Mixed Gas
FCAW	Flux Cored Arc Welding
FCAW-MC	Metal Cored Arc Welding
SSFCAW (ISW)	Self Shielded Flux Cored Arc Welding
SAW	Submerged Arc Welding
GTAW	Gas Tungsten Arc Welding

Using COSTCOMP the welding times T_{w2} (min) were calculated versus weld size a_W (mm) for all kind of welding types in downhand position. The values of power n in Eq. (7) come form curve fitting calculations.

The welding times T_{w2} (min/mm) versus weld size a_W (mm) also were calculated for longitudinal fillet welds and for longitudinal V butt welds in positional welding, which means not downhand, but vertical or overhead positions.

Figure 1 shows the welding times for longitudinal V butt welds in decreasing order SMAW, SMAW-HR, GMAW-C, GMAW-M, FCAW, FCAW-MC, ISW and SAW. The highest being SMAW and the lowest is SAW.

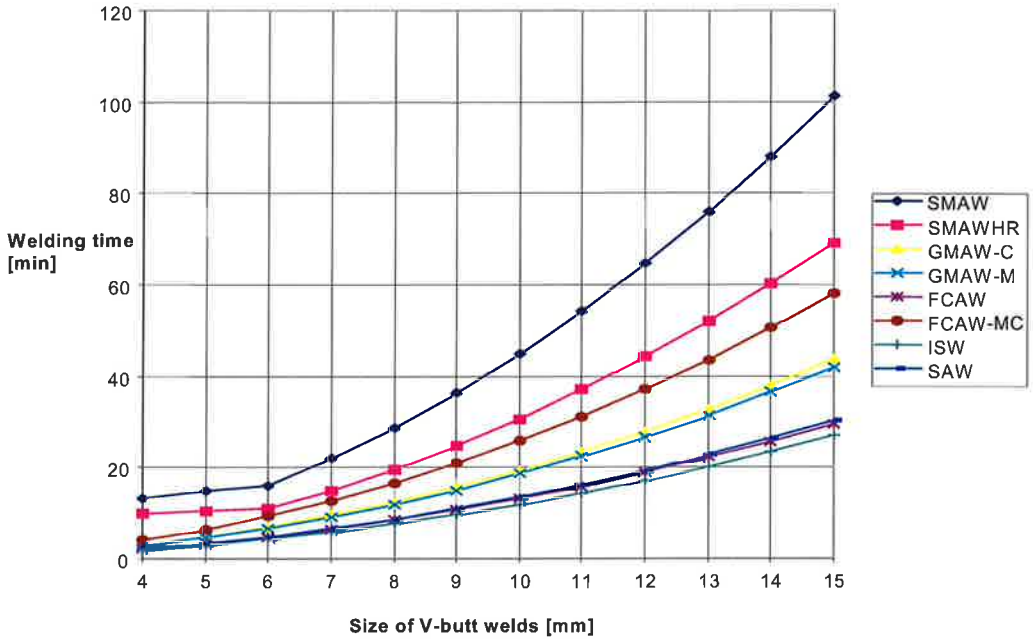


Figure 1. Welding times T_{w2} (min/mm) versus weld size a_W (mm) for downhand position

Calculation of the times of special welding technologies

The arc-spot welding is used for many structures, where only one side welding is possible. The production time is given by

$$T_{w4} = n_S T_S , \tag{11}$$

where n_S is the number of spots, T_S is the time of welding one spot weld and of transferring the electrode to the next spot. T_S depends on the welding equipment and the degree of automation (Jármai et al. 1999).

4.2.2 *Fabrication times of post-welding treatments*

To increase the dynamic behaviour of welded structures, often post weld treatments (PWT) are used (Jármai et al. 2000). These treatments are grinding, TIG dressing, hammer- and shot peening, ultrasonic impact treatment (UIT). Time for PWT is

$$T_{PWT} = T_0 L_t , \tag{12}$$

where T_0 is the specific time (min/mm), L_t is the treated weld length (mm). Table 2 shows the specific times for the given PWT.

Table 2. Time needed for different PWT techniques

Method	T_0 (min/m)
Grinding	60
TIG dressing	18
Hammer peening	4
UIT	15

4.2.3 *Time for flattening plates*

The smallest possible initial imperfection is important at the stability behaviour of plated structures that is why flattening plates can be necessary depending on the producer. In the catalogue of different companies one can find the times for flattening plates (T_{FP} [min]) versus a plate thickness (t [mm]) and the area of the plate (A_p [mm^2]). Based on curve fitting calculations the time function can be written in the form:

$$T_{FP} = \Theta_{df}\left(a_e + b_e t^3 + \frac{1}{a_e t^4}\right)A_p, \tag{13}$$

where a_e=9.2*10^{-4} [min/mm^2], b_e= 4.15*10^{-7} [min/mm^5], Θ_{df} is the difficulty parameter (Θ_{df} = 1,2 or 3). The difficulty parameter depends on the form of the plate.

4.2.4 *Surface preparation time*

The surface preparation means the surface cleaning, sand spraying, etc. The surface cleaning time can be defined versus the surface area (A_s [mm^2]) as follows:

$$T_{SP} = \Theta_{ds} a_{sp} A_s, \tag{14}$$

where a_{sp} = 3*10^{-6} [min/mm^2], Θ_{ds} is a difficulty parameter.

4.2.5 *Painting time*

The painting means making the ground- and the topcoat. The painting time can be given versus the surface area (A_s [mm^2]) as follows:

$$T_P = \Theta_{dp}(a_{gc} + a_{tc})A_s, \tag{15}$$

where a_{gc} = 3*10^{-6} [min/mm^2] , a_{tc} = 4.15*10^{-6} [min/mm^2], Θ_{dp} is a difficulty factor, Θ_{dp}=1,2 or 3 for horizontal, vertical or overhead painting. Tizani et al. (1996) proposed a value for painting 14.4 [\$/m^2].

4.2.6 *Plate cutting and edge grinding times*

The cutting and edge grinding can be made by different technologies, like Acetylene, Stabilized gasmix and Propane with normal and high speed.

The cutting time can be calculated also by COSTCOMP. The normal speed acetylene has the highest time and the high-speed propane has the smallest cutting time (Fig. 2).

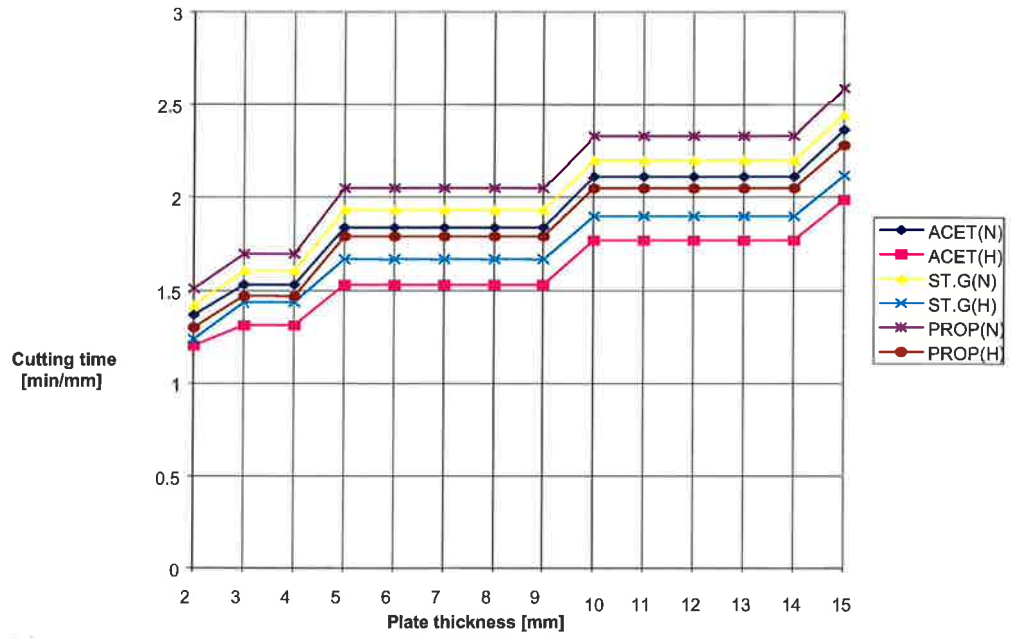


Figure 2. Cutting time for 1 mm length of plates, T_{CP} (min/mm) versus thickness for fillet, T-, V-, 1/2 V butt welds

The cutting cost function can be formulated versus the thickness (t [mm]) and cutting length (L_c [mm]):

$$T_{CP} = \sum_i C_{CPi} t_i^n L_{ci} \,, \tag{15}$$

where t_i the thickness in [mm], L_{ci} is the cutting length in [mm]. The value of n comes from curve fitting calculations.

4.2.7 *Times of hand cutting and machine grinding of strut ends*

At tubular structures a main part of the total cost is the cost of hand cutting and machine grinding of strut ends. We use the following formula (Farkas & Jármai 1997)

$$T_{CG} = \Theta_{dc} \sum_i \frac{2\pi d_i}{\sin\varphi} \left(4.54 + 0.4229 t_i^2\right), \tag{16}$$

where the fabrication cost factor is taken on the basis of Tizani et al. (1996) as Ł25/h = 40$/h = 0.6667 [$/min], and the difficulty factor is considered as $\Theta_{dc} = 3$. The diameter of the brace is d_i in m, thickness is t_i in mm. φ is the angle between the two members (chord and brace) connected.

Note that Glijnis (1999) proposed a formula for one strut end in the case of oxyfuel cutting on CNC machine as follows:

$$K_{CG}(\$) = \frac{2.5\pi d_i}{(350 - 2t_i)0.3\sin\varphi_i} \tag{17}$$

where 350 mm/min is the cutting speed, 0.3 is the efficiency factor, d_i and t_i are in mm.

4.2.8 *Total cost function*

The total cost function can be formulated by adding the previous cost functions together (depending on the structure some can be zero).

$$\frac{K}{k_m} = \rho V + \frac{k_f}{k_m}\left(T_{w1} + T_{w2} + T_{w3} + T_{w4} + T_{PWT} + T_{FP} + T_{SP} + T_P + T_{CP} + T_{CG}\right) \tag{18}$$

Taking k_m = 0.5-1 \$/kg, k_f =0 -1 \$/min. The k_f/k_m ratio varies between 0 - 2 kg/min. If k_f/k_m = 0, then we get the mass minimum. If k_f/k_m = 2.0 it means a very high labour cost (Japan, USA), k_f/k_m = 1.5 and 1.0 means a West European labour cost, k_f/k_m = 0.5 means the labour cost of developing countries. Even if the production rate is similar for these cases, the difference between costs due to the different labour costs is significant.

5 NUMERICAL EXAMPLE

We show the cost calculations at stiffened plates, where the welding cost is larger.

5.1 *Welded stiffened plate*

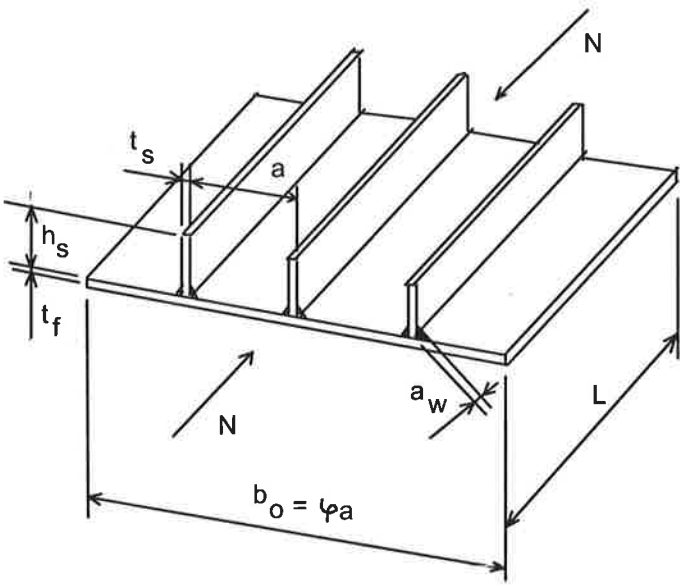


Figure 3. Stiffened plate

The problem is to minimise the cost of a stiffened plate considering the stress, overall and local buckling constraints in the case of different welding technologies.

The stiffened plates are widely applied in bridge and ship structures. Since the welding cost is a great part of the total cost, it is economic to optimize these structural components for minimum cost (Farkas & Jármai 1993, Jármai 2000a, 2000b).

The cost function is calculated according to Eq (18), where

$A=b_0t_f+\varphi h_s t_s$, $\Theta_{dw}=3$, $\kappa=\varphi+1$, $L_w=2L\varphi$ and φ is the number of stiffeners.

The stiffeners are welded to the plate by double fillet welds. The welding costs can be calculated for different welding technologies (Fig. 1).

The main data for the optimization are as follows:

Young modulus of the steel is $E = 2.1*10^5$ MPa, material density is $\rho = 7.85*10^{-6}$ kg/mm^3, Poisson parameter is $\nu = 0.3$, yield stress is $f_y = 235$ MPa, width of the plate is b_0=4200 mm and the plate length is L=4000 mm.
The compression force is

$$N = f_y b_0 t_{f\,max} = 235*4200*20 = 1.974*10^7 \ \text{(N)} \tag{19}$$

The independent design variables are as follows (Fig. 3):
Thickness of the plate (t_f), height and thickness of the stiffeners (h_s, t_s) and the number of stiffeners ($\varphi=b_0/a$).

5.2 *Design constraints*

a) Overall buckling design rules, according to API (1987) for the compressed plate with uniform distance stiffeners (Fig. 3).

$$N \le \chi f_y A \tag{20}$$

where, χ is the buckling constraints, versus the reduced slenderness factor:

$$\chi = 1 \quad \text{when} \quad \bar{\lambda} \le 0.5,$$
$$\chi = 1.5 - \bar{\lambda} \quad \text{when} \quad 0.5 \le \bar{\lambda} \le 1,$$
$$\chi = \frac{0.5}{\bar{\lambda}} \quad \text{when} \quad \bar{\lambda} \ge 1, \tag{21}$$

and

$$\bar{\lambda} = \frac{b_0}{t_f}\sqrt{\frac{12(1-\nu^2)f_y}{E\pi^2 k}}, \tag{22}$$

$$k_{min} = min(k_F, k_R), \tag{23}$$

$$k_R = 4\varphi^2. \tag{24}$$

$$k_F = \frac{(1+\alpha^2)^2 + \varphi\gamma}{\alpha^2(1+\varphi\delta_P)} \quad \text{when } \alpha = \frac{L}{b_0} \le 4\sqrt{1+\varphi\gamma} \tag{25}$$

and

$$k_F = \frac{2\left(1+\sqrt{1+\varphi\gamma}\right)}{1+\varphi\gamma} \text{ when } \alpha \ge 4\sqrt{1+\varphi\gamma} \tag{26}$$

where

$$\delta_P = \frac{h_s t_s}{b_0 t_f}, \tag{27}$$

$$\gamma = \frac{EI_s}{b_0 D}, \tag{28}$$

$$I_s = \frac{h_s^3 t_s}{3}, \tag{29}$$

$$D = \frac{Et_f^3}{12(1-\nu^2)}. \tag{30}$$

Eq (28) can be rewritten as

$$\gamma = 4(1-\nu^2)\frac{h_s^3 t_s}{b_0 t_f^3} = 3.64\frac{h_s^3 t_s}{b_0 t_f^3}. \tag{31}$$

where I_s is the moment of inertia of one stiffener about an axis parallel to the plate surface at the base of the stiffener, D is the torsional stiffness of the main plate.

b) Buckling constraint of the stiffener is (Eurocode 3 1992):

$$\frac{h_s}{t_s} \le \frac{1}{\beta_s} = 14\sqrt{\frac{235}{f_y}} \tag{32}$$

The size constraints for the variables are as follows:

- t_f= 6 - 20 mm,
- h_s= 84 - 280 mm,
- t_s= 6 - 25 mm,
- φ= 4 - 15 mm.

The elements of cost function for the welded stiffened plate are as follows

Size of welded joint $a_w = t_s$

Cross section area $A = b_0 t_f + \varphi h_s t_s$

Material cost $\rho V = \rho LA$

Fabrication costs $k_f/k_m \sum_i T_i$

$T_{1w} = C_1 \Theta_{dw} \sqrt{\kappa \rho V}$, where $\rho = 7.85*10^{-6}$, $C_1 = 1$, $\kappa = \varphi + 1$, $\Theta_{dw} = 2$

$T_{2w} + T_{3w} = 1.3 \sum C_{2i} a_{wi}^n L_{wi}$ where $C_{2i} = 0.7889$, $n = 2$ for SMAW, $L_{wi} = 2\,L\,\varphi$

$T_{FP} = \Theta_{de}\left(a_e + b_e t^3 + \frac{1}{a_e t^4}\right) A_p$ where $a_e = 9.2*10^{-4}$, $b_e = 4.15*10^{-7}$, $t = t_s$, or t_f, $A_p = \varphi h_s L$ or $b_0 L$, $\Theta_{de} = 1$,

$T_{SP} = \Theta_{ds} a_{sp} A_s = 5*10^{-7}$ where $a_{sp} = 3*10^{-6}$, $A_s = \varphi h_s L + b_0 L$, $\Theta_{ds} = 1$,

$T_P = \Theta_{dp}(a_{gc} + a_{tc}) A_s$ where $a_{gc} = 3*10^{-6}$, $a_{tc} = 4.15*10^{-6}$, $A_s = \varphi h_s L + b_0 L$, $\Theta_{dp} = 2$,

$T_{CG} = \sum_i C_{CGi} t_i^n L_{ci}$ where $C_7 = 1.1388$, $t = t_s$ or t_f, $n = 0.25$, $L_{ci} = (h_s + L)$ or $(b_0 + L)$.

Table 3 shows the optimum discrete sizes of the stiffened plate with different welding technologies.

5.3 *Conclusions*

Figures 4 shows the distribution of the total cost. The diagrams illustrate that this distribution depends on the welding technologies, the type of welding, the ratio of material and fabrication specific costs and the structure type too.

The welding technologies in Figure 4 are given in decreasing order relating to the welding time and cost. The differences are great among them. The welding time and cost is the greatest for SMAW, the quickest and cheapest are the SAW, FCAW and ISW. For stiffened plates using SMAW 46% of the total cost is the welding cost, using SAW, this is only 20%.

The mass of stiffened plate is ρLA = 3258 kg (Table 3), the fabrication cost is 100 (15559-3258) / 15559 = 79 % of the total cost.

Table 3. Optimum rounded sizes of welded stiffened plates in mm with fillet welds using different welding technologies for k_f/k_m=2.0

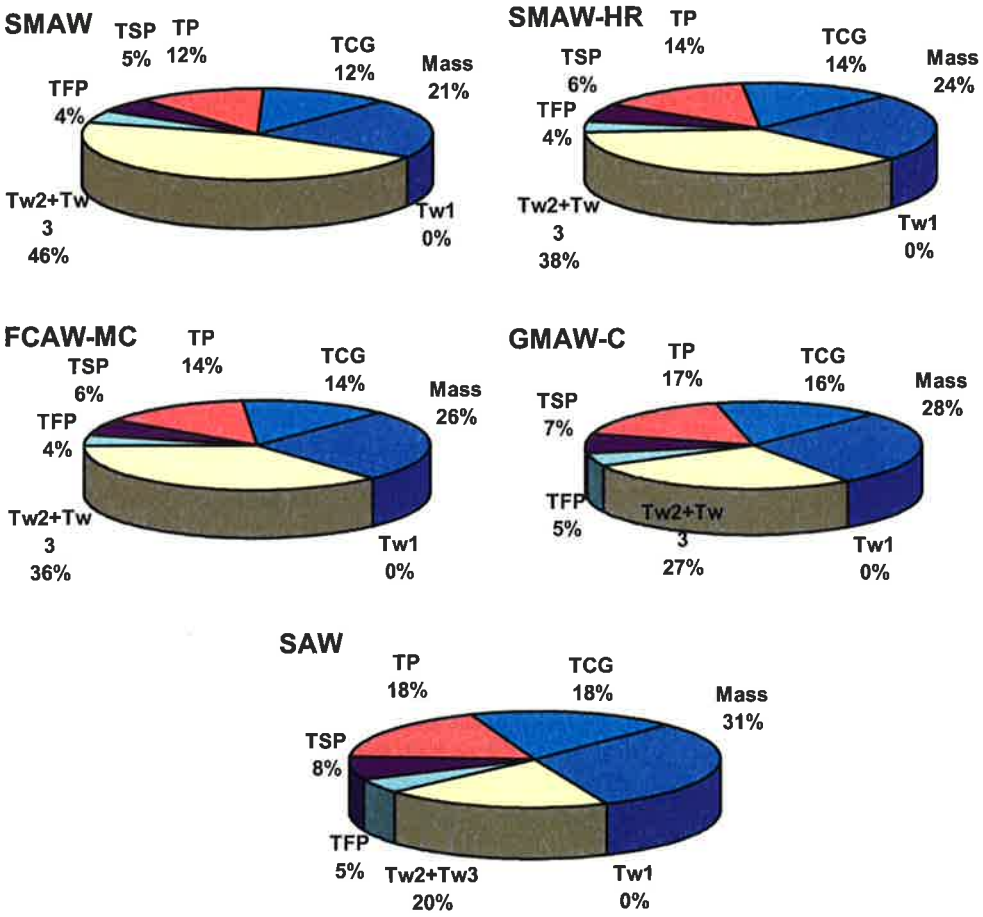
Welding technology	k_f/k_m	h_s	t_f	φ	t_s	ρV (kg)	K/k_m (kg)
SMAW	0.0	210	17	13	11	2737	2737
	0.5	230	17	6	19	3242	6313
	1.0	235	17	6	19	3258	9409
	1.5	235	17	6	19	3258	12484
	2.0	235	17	6	19	3258	**15559**
SMAW HR	0.0	210	17	13	11	2737	2737
	0.5	230	17	6	19	3242	5749
	1.0	230	17	6	19	3242	8257
	1.5	230	17	6	19	3242	10764
	2.0	235	17	6	19	3258	**13306**
FCAW-MC	0.0	210	17	13	11	2737	2737
	0.5	230	17	6	19	3242	5553
	1.0	230	17	6	19	3242	7864
	1.5	230	17	6	19	3242	10175
	2.0	235	17	6	19	3258	**12521**
GMAW-C GMAW-M	0.0	210	17	13	11	2737	2737
	0.5	230	17	6	19	3242	5299
	1.0	230	17	6	19	3242	7357
	1.5	235	17	6	19	3258	9444
	2.0	230	17	6	19	3242	**11471**
SAW ISW FCAW	0.0	210	17	13	11	2737	2737
	0.5	230	17	6	19	3242	5064
	1.0	230	17	6	19	3242	6886
	1.5	230	17	6	19	3242	8707
	2.0	235	17	6	19	3258	**10564**

Cost savings can be achieved using a cheaper welding technology, like SAW instead of SMAW or GMAW, if it is possible. Table 4 shows the cost savings for the two different structures and for the five different groups of welding. For stiffened plates the cost savings can be 32 % of the total cost. All compared results are optimized.

Table 4 Cost savings using different welding technologies

Welding technology k_f/k_m=2.0	Stiffened plate Total cost	Cost savings in %
SMAW	15559	0
SMAW-HR	13305	14
FCAW-MC	12521	20
GMAW-C	11471	27
SAW	10560	32

Figure 4. The total cost distribution of the welded stiffened plate with fillet welds using different welding technologies for k_f/k_m=2.0



There are a great number of examples published on different fields to demonstrate the usability of optimisation forming a structure or a system to be more reliably and more economic. For example for bridgedecks (Jármai et al. 1997), for pipelines (Jármai & Lukács 1999), for sugar drying (Szabó & Jármai 2000), for furnace wall structures (Szűcs et al. 1997), etc.

ACKNOWLEDGEMENTS

The research work was supported by the Hungarian Scientific Research Foundation grants OTKA T-38058 and T-37941. Also the long term Socrates agreement between the universities of Miskolc and Lappeenranta is acknowledged.

REFERENCES

American Petroleum Institute, 1987. *API Bulletin on Design of flat plate structures.* Bul. 2V, 1st edn.

Bodt,H.J.M. 1990. *The Global Approach to Welding Costs*. The Netherlands Institute of Welding, The Hague.

COSTCOMP, 1990. *Programm zur Berechnung der Schweisskosten.* Deutscher Verlag für Schweisstechnik, Düsseldorf.

Eurocode 3, 1992. *Design of steel structures*, Part 1.1, CEN. European Committee for Standardization, Brussels.

Farkas,J. 1992. Minimum cost design of welded structures. In *Engineering Design in Welded Constructions*, Proceedings of the Int. Conference IIW, Madrid, Spain, 135-142. Pergamon Press, Oxford.

Farkas,J., Jármai,K. 1992. Minimum cost design of laterally loaded welded rectangular cellular plates. In *Structural Optimization '93 World Congress*, Rio de Janeiro. Proc. **1**, 205-212.

Farkas,J.,Jármai, K. 1996. Backtrack method with applications to DSO, June 17-21. 1996. *Discrete Structural Optimization, Advanced School* Coordinated by W. Gutkowski, *International Centre for Mechanical Science,* CISM; Udine, Italy. *Discrete Structural Optimization,* Springer Verlag, Edited by W. Gutkowski, 1997. *Chapter 4.* pp. 167-232. ISBN 3-211-82901-6

Farkas & Jármai (Eds.) 1996. *7th International Symposium on Tubular Structures,* Proceedings, Balkema Publishers, Rotterdam, 490 p. ISBN 90 5410 828 2

Farkas,J.,Jármai,K. 1997. *Analysis and Optimum Design of Metal Structures*. Balkema Publishers, Rotterdam, Brookfield.

Farkas,J.,Jármai,K. 2003. *Economic design of metal structures*. Millpress Science Publishers, Rotterdam.

Glijnis,P.C. 1999. Private communication.

Himmelblau D.M. 1972. Applied nonlinear programming. McGraw-Hill, New York.

Jármai,K. (Ed.) 1997. *International Symposium on Design of Metal Structures,* December 12, 1997. University of Miskolc, Hungary. Publications of the University of Miskolc, Series C, Mechanical Engineering, **47**: 217 HU ISSN 0237-6016

Jármai,K.,Horikawa,K.,Farkas,J. 1997. Economic design of steel bridge decks with open ribs, *Transactions of JWRI*, Joining and Welding Research Institute, Osaka University, Japan, **26** (1): 147-161.

Jármai,K.,Farkas,J.(Eds.) 1998. *Mechanics and Design of Tubular Structures, Advanced Professional School,* Coordinated by J. Farkas & K. Jármai, International Centre for Mechanical Science, CISM; Udine, Italy, Springer Verlag, 337 p. ISBN 3-211-83145-2

Jármai,K.,Lukács,J. 1999. Optimum design of pipelines considering fatigue crack propagation. *MicroCAD'99, University of Miskolc*, Febr. 24-25, 1999. Section K. Machine and Structure Design pp. 77-83. ISBN 963 661 361 3

Jármai,K.,Farkas,J. 1999. Cost calculation and optimization of welded steel structures, *Journal of Constructional Steel Research*, Elsevier, **50** (2): 115-135, ISSN 0143-974X

Jármai,K.,Farkas,J.,Petershagen,H-J. 1999. Optimum design of welded cellular plates for ship deck panels, *Welding in the World*, Pergamon Press, **43** (1): 50-54. ISSN 0043-2288

Jármai,K.,Farkas,J.,Haagensen,P.J. 2000. Effect of post-welding treatments on the optimum fatigue design of welded I-beams, *Welding in the World*, Pergamon Press, **44** (2): 56-59. ISSN 0043-2288

Jármai,K. 2000a. Optimum design of stiffened plates, *Journal of Computational and Applied Mechanics*, **1** (1): 91-98, Miskolc University Press, HU ISSN 1586-2070

Jármai,K. 2000b. Optimum of welded structures for cost, *Journal of Computational and Applied Mechanics*, **1** (2): 149-166, Miskolc University Press, HU ISSN 1586-2070

Jármai,K.,Iványi,M. 2001. *Analysis and design of economic steel structures*, Műegyetemi Kiadó, Budapest, 226 p. ISBN 963 420 674 3 (in Hungarian)

Jármai,K.,Farks,J. (Eds.) 2003. International Conference on Metal Structures, ICMS 2003, April 3-5. Miskolc, Hungary, Proceedings 400 p. Millpress Science Publisher, Rotterdam. (under publication)

Likhtarnikov,Y.M. 1968, *Metal Structures*. (in Russian), Stroyizdat, Moscow.

Mallik Ntuen, N.C.A., Mallik,A.K. 1987. Applying artificial intelligence to project cost estimating. *Cost Engineering* **29** (5) 8-13.

Michell, A.G.M. 1904. The limits of economy of material in framed structures. Phil. Mag. Series 6. Vol. 8. pp.589-597.

Ott,H.H., Hubka,V. 1985. Vorausberechnung der Herstellkosten von Schweisskonstruktionen (Fabrication cost calculation of welded structures). *Proc. Int. Conference on Engineering Design ICED*, 1985, Hamburg, 478-487. Heurista, Zürich.

Pahl,G., Beelich,K.H. 1992. Kostenwachstumsgesetze nach Ähnlichkeitsbeziehungen für Schweiss-verbindungen. *VDI-Bericht*, Nr. 457, 129-141, Düsseldorf.

Pareto,V. 1896. Cours d'economie politique. Vols. I and II. Lausanne: F. Rouge

Peurifoy,R.L. 1975, *Estimating construction costs*. 3rd ed. McGraw Hill, New York.

Ramirez,J.C.,Touran,A. 1991. An integrated computer system for estimating welding cost. *Cost Engineering*, Vol. 33, No. 8. pp. 7-14.

Schittkowski, K., Zillober, C.,Zotemantel, R.1994: Numerical comparison of nonlinear programming algorithm for structural optimization. Journal of Structural Optimization, Vol.7, No. 1/2, 1-19.

Snyman J.A. 2005. Practical Mathematical Optimization, An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, Springer Verlag, ISBN-10: 0-387-29824-X

Szabó,Sz.,Jármai,K. 2000. Heat engineering model and optimization of the operation of drier cylinder, *Sugar Industry,* **52** (127): 265-269. ISSN 0344-8657

Szűcs,I.,Jármai,K.,Réz,I.,Szemmelveiszné,K. 1997. Optimization of refractory wall structures of furnaces in the rolling mills, *Hutní Keramika,* 1-2. Rijna 1997. Ostrava, Czeck Republic, Proceedings pp. 25-30.

Tizani,W.M.K., Yusuf,K.O.et al. 1996. A knowledge based system to support joint fabrication decision making at the design stage – Case studies for CHS trusses. *Tubular Structures* VII. Eds Farkas,J. & Jármai,K. Rotterdam-Brookfield, Balkema, 483-489.

Vanderplaats, G.N. 1984: Numerical optimization techniques for engineering design. New York: McGraw-Hill

Volkov,V.V. 1978. Determining fabrication times for structural parts of industrial buildings. (in Russian). *Trudy TsNII Proektstalkonstruktsiya*, Vyp. 23. Moskva, pp.34-45.

Winkle,I.E.,Baird,D. 1986, Towards more effective structural design through synthesis and optimisation of relative fabrication costs. *Transactions Royal Inst. Naval Archit.* RINA No. 128. pp. 313-336.

Yco,R.B.G. 1983. Cost effective steel fabrication. Part 2. Design for welding. *Metal Construction*, Vol. 15. No. 3. pp. 151-156, 158.

ERROR-CONTROLLED ADAPTIVE MIXED FINITE ELEMENT METHODS FOR SECOND-ORDER ELLIPTIC EQUATIONS

C. LOVADINA[1], R. STENBERG[2], M. RÜTER[2]

[1] Dipartimento di Matematica, Università di Pavia and IMATI-CNR
Via Ferrata 1, Pavia 27100, Italy
e-mail: lovadina@dimat.unipv.it

[2] Institute of Mathematics, Helsinki University of Technology
P.O. Box 1100, 02015 TKK, Finland
e-mail: Rolf.Stenberg/Marcus.Ruter@tkk.fi

ABSTRACT

In this contribution, we deal with a posteriori error estimates and adaptivity for mixed finite element discretizations of second-order elliptic equations. Here, the method proposed is applied to the mixed model problem of the Poisson equation which is discretized by mixed finite elements of the so-called Raviart-Thomas type. The residual-type a posteriori error estimator presented relies on a postprocessed and therefore improved solution for the displacement field that can be carried out locally, i.e. on the element level, see also [10]. With this postprocessed solution at hand, upper and lower bounds on the finite element discretization error can be obtained. Finally, a numerical example is presented that illustrates our theoretical results.

1 Introduction

Following the recent paper by Lovadina and Stenberg [10], we consider mixed finite element discretizations of second-order elliptic equations. In this paper, we choose the Poisson problem as a model problem in which we search for a solution $u \in H_0^1(\Omega)$ such that

$$-\Delta u = f \quad \text{in } \Omega \subset \mathbb{R}^n. \tag{1}$$

Upon introducing the flux $\boldsymbol{\sigma}$, we may recast (1) into the system

$$\boldsymbol{\sigma} - \nabla u = \mathbf{0} \quad \text{in } \Omega, \tag{2a}$$

$$\operatorname{div} \boldsymbol{\sigma} + f = 0 \quad \text{in } \Omega. \tag{2b}$$

The associated finite element discretization then results into the following mixed method of solving

$$(\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + (\operatorname{div} \boldsymbol{\tau}, u_h) = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{S}_h, \tag{3a}$$

$$(\operatorname{div} \boldsymbol{\sigma}_h, v) + (f, v) = 0 \quad \forall v \in V_h \tag{3b}$$

for a solution $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h \subset \boldsymbol{H}(\operatorname{div}:\Omega) \times L^2(\Omega)$.

In this method, the polynomial used for approximating the flux $\boldsymbol{\sigma}$ is of higher degree than the one used for the displacement u, which is counterintuitive with respect to (2a). Consequently, the mixed method has to be carefully designed in order to satisfy the Babuška-Brezzi conditions, cf. e.g. [7], which can be shown either by means of the $\boldsymbol{H}(\mathrm{div}:\Omega)$-norm for the flux and the $L^2(\Omega)$-norm for the displacement or by means of so called mesh dependent norms [3] which are close to the energy norm of the continuous problem. Note that both ways yield the same a priori error estimate.

The a posteriori error analysis of mixed methods has been shown in [1], [9] and [4], where, e.g. in [9], the $\boldsymbol{H}(\mathrm{div}:\Omega)$-norm has been used which is, however, trivially computable and also may dominate the error. Furthermore, an estimate for the L^2-norm of the flux, as used e.g. in [4], is not optimal, since the estimator includes the element residual in the constitutive relation (2a) which is large due to the fact that the polynomial degree of the displacement approximation is lower than that for the flux.

In this paper, we present a remedy to this in terms of a postprocessed solution for the displacements. Since the work of Arnold and Brezzi [2] it is known that the mixed finite element solution can be locally postprocessed which results in an improved displacement field. Later, other postprocessing has been suggested [5, 8, 6, 14, 13]. The cornerstone of the postprocessing presented in this paper is that the postprocessed displacements are of one degree higher than the flux, which is in accordance with (2a).

2 A priori estimates and postprocessing

In this section, we derive a priori error estimates for the model problem presented. Furthermore, we introduce the postprocessing method for the displacements as used in the a posteriori error analysis later in Section 3.

To begin with, let us introduce the finite element regular partitioning $\mathcal{C}_h$ in terms of triangular (or tetrahedral) elements K and the collection Γ_h of edges (or faces) E. In this paper, we confine ourselves to triangular finite elements of the Raviart-Thomas type [12], although also Brezzi-Douglas-Marini elements could be chosen for the FE-discretization. In this case, we define the finite-dimensional subspaces $\boldsymbol{S}_h \times V_h \subset \boldsymbol{H}(\mathrm{div}:\Omega) \times L^2(\Omega)$ as piecewise polynomial spaces on $\mathcal{C}_h$ defined as

$$\boldsymbol{S}_h = \{\, \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}:\Omega) \,|\, \boldsymbol{\tau}|_K \in [P_{k-1}(K)]^n \oplus \boldsymbol{x}\tilde{P}_{k-1}(K)\ \forall K \in \mathcal{C}_h \,\} \tag{4a}$$

$$V_h = \{\, v \in L^2(\Omega) \,|\, v|_K \in P_{k-1}(K)\ \forall K \in \mathcal{C}_h \,\} \tag{4b}$$

with integer $k \geq 1$ and $\tilde{P}_{k-1}(K)$ denoting homogeneous polynomials of degree $k-1$.

Throughout this paper, we also use the following norms for the displacements

$$\|v\|_{1,h}^2 = \sum_{K\in\mathcal{C}_h} \|\nabla v\|_{0,K}^2 + \sum_{E\in\Gamma_h} h_E^{-1} \|[\![v]\!]\|_{0,E}^2 \tag{5}$$

and for the flux

$$\|\boldsymbol{\tau}\|_{0,h}^2 = \|\boldsymbol{\tau}\|_0^2 + \sum_{E\in\Gamma_h} h_E \|\boldsymbol{\tau}\cdot\boldsymbol{n}\|_{0,E}^2, \tag{6}$$

where $\boldsymbol{n}$ is the unit normal to $E \in \Gamma_h$ and $[\![v]\!]$ is the jump in v along the interelement boundaries and the boundary $\partial\Omega$.

Upon defining the bilinear form

$$\mathcal{B}(\boldsymbol{\varphi}, w; \boldsymbol{\tau}, v) = (\boldsymbol{\varphi}, \boldsymbol{\tau}) + (\mathrm{div}\,\boldsymbol{\tau}, w) + (\mathrm{div}\,\boldsymbol{\varphi}, v), \tag{7}$$

the mixed method can be defined in a more compact fashion as:

Find $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ such that

$$\mathcal{B}(\boldsymbol{\sigma}_h, u_h; \boldsymbol{\tau}, v) + (f, v) = 0 \qquad \forall (\boldsymbol{\tau}, v) \in \boldsymbol{S}_h \times V_h. \tag{8}$$

In order to derive an a priori error estimate we employ the *interpolation operator* $\boldsymbol{R}_h$: $\boldsymbol{H}(\mathrm{div}:\Omega) \cap [L^s(\Omega)]^n \to \boldsymbol{S}_h$, with $s > 2$, such that

$$(\mathrm{div}\,(\boldsymbol{\tau} - \boldsymbol{R}_h\boldsymbol{\tau}), v) = 0 \quad \forall v \in V_h, \tag{9}$$

which can be constructed by using the degrees of freedom for $\boldsymbol{S}_h$, cf. [12, 11, 8, 6]. With the *equilibrium property*

$$\mathrm{div}\,\boldsymbol{S}_h \subset V_h \tag{10}$$

and denoting by $P_h : L^2(\Omega) \to V_h$ the L^2-projection, it then follows that

$$(\mathrm{div}\,\boldsymbol{\tau}, u - P_h u) = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{S}_h. \tag{11}$$

Moreover, the projection and interpolation operators satisfy the *commuting property*:

$$\mathrm{div}\,\boldsymbol{R}_h = P_h \mathrm{div}\,. \tag{12}$$

With these notations at hand, we arrive at the following a priori error estimate: There is a positive constant C such that

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|P_h u - u_h\|_{1,h} \leq C\|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0 \tag{13}$$

which results in

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|P_h u - u_h\|_{1,h} \leq Ch^k|\boldsymbol{\sigma}|_k. \tag{14}$$

For further details we refer to [10]. Notice that these estimates contain a *superconvergence* result for $\|P_h u - u_h\|_{1,h}$. Recalling (2a), it is furthermore obvious that $\boldsymbol{\sigma}_h$ is a good approximation of ∇u. Hence, we may construct an improved approximation for the displacements u by introducing a local postprocessing technique on the element level. Here, we use the approach as introduced in [14, 13], i.e. we seek an improved solution $u_h^* \in V_h^*$ such that

$$P_h u_h^* = u_h \tag{15}$$

and

$$(\nabla u_h^*, \nabla v)_K = (\boldsymbol{\sigma}_h, \nabla v)_K \quad \forall v \in (I - P_h)V_h^*|_K, \tag{16}$$

where the space $V_h^* \subset V_h$ is defined as

$$V_h^* = \{\, v \in L^2(\Omega) \mid v|_K \in P_k(K)\ \forall K \in \mathcal{C}_h \,\}. \tag{17}$$

The error analysis of this postprocessing has been carried out in [14, 13]. In this paper, however, we consider the method and the postprocessing as one approach. Therefore, we define the bilinear form

$$\begin{aligned}\mathcal{B}_h(\boldsymbol{\varphi}, w^*; \boldsymbol{\tau}, v^*) =& (\boldsymbol{\varphi}, \boldsymbol{\tau}) + (\mathrm{div}\,\boldsymbol{\tau}, w^*) + (\mathrm{div}\,\boldsymbol{\varphi}, v^*) \\ &+ \sum_{K \in \mathcal{C}_h} (\nabla w^* - \boldsymbol{\varphi}, \nabla(I - P_h)v^*)_K.\end{aligned} \tag{18}$$

Then, the following equivalence to the original problem can be easily verified: Let $(\boldsymbol{\sigma}_h, u_h^*) \in \boldsymbol{S}_h \times V_h^*$ be the solution of

$$\mathcal{B}_h(\boldsymbol{\sigma}_h, u_h^*; \boldsymbol{\tau}, v^*) + (P_h f, v^*) = 0 \qquad \forall (\boldsymbol{\tau}, v^*) \in \boldsymbol{S}_h \times V_h^*, \tag{19}$$

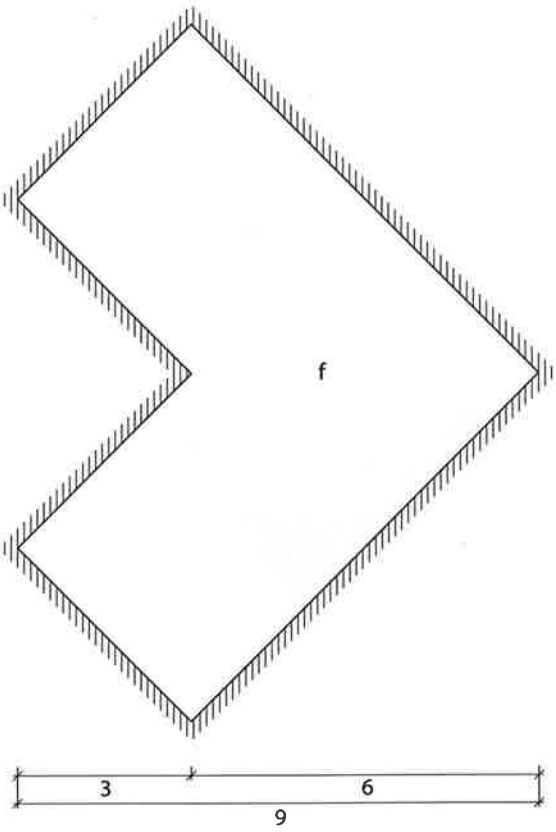


Figure 1: System, loading and measurements.

and set $u_h = P_h u_h^* \in V_h$. In this case, $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ coincides with the solution of (3a)–(3b). Conversely, let $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ be the solution of (3a)–(3b), and let $u_h^* \in V_h^*$ be the postprocessed displacement defined by (15)–(16). Then $(\boldsymbol{\sigma}_h, u_h^*) \in \boldsymbol{S}_h \times V_h^*$ is the solution to (19). It can now be shown that the following a priori error estimate holds:

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \leq C\big(\|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0 + \inf_{v^* \in V_h^*} \|u - v^*\|_{1,h}\big).$$

For our choice of space we obtain the a priori estimate (with the assumption of a sufficiently smooth solution)

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \leq Ch^k |u|_{k+1} \tag{20}$$

for a positive constant C.

3 A posteriori estimates

In what follows, we briefly derive a posteriori error estimates for the mixed problem presented which are based on the postprocessed and therefore improved displacement field u_h^*.

As a starting point, let us define local error indicators on the element level. More

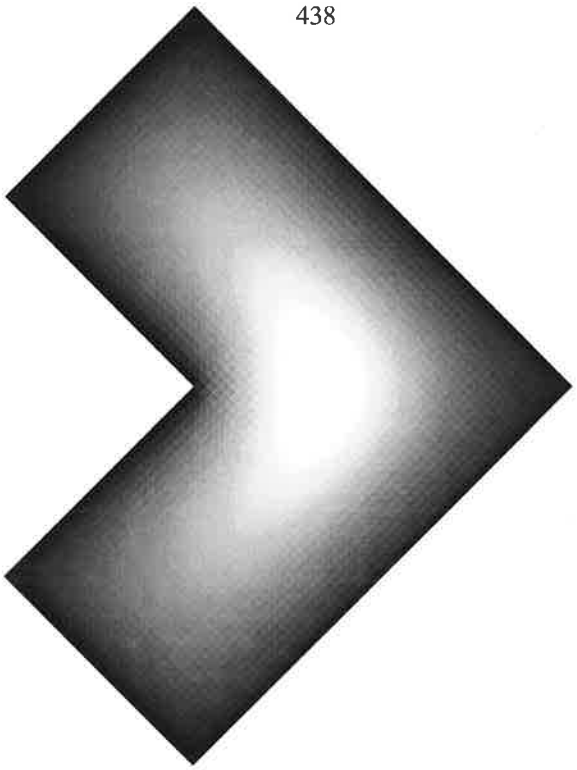


Figure 2: Approximate solution of the displacements.

precisely, we define the element domain contributions

$$\eta_{1,K} = \|\nabla u_h^* - \boldsymbol{\sigma}_h\|_{0,K}, \tag{21a}$$

$$\eta_{2,K} = h_K \|f - P_h f\|_{0,K} \tag{21b}$$

and the edge contribution

$$\eta_E = h_E^{-1/2} \|[\![u_h^*]\!]\|_{0,E}. \tag{21c}$$

With these local error indicators at hand, we may then define the global error estimator

$$\eta = \Big(\sum_{K \in \mathcal{C}_h} (\eta_{1,K}^2 + \eta_{2,K}^2) + \sum_{E \in \Gamma_h} \eta_E^2 \Big)^{1/2}. \tag{22}$$

All that remains is to show the efficiency and the reliability of the error estimator presented above.

The efficiency of the estimator is simply given by the lower bound

$$\begin{aligned} \eta_{1,K} &\leq \|\nabla(u - u_h^*)\|_{0,K} + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,K}, \\ \eta_E &= h_E^{-1/2} \|[\![u - u_h^*]\!]\|_{0,E}, \end{aligned} \tag{23}$$

which is an immediate consequence of (2a), using the triangle inequality, and of (21c) noting that $[\![u]\!] = 0$ on each edge E.

The reliability of the estimator can be shown in two different ways, namely either via a saturation assumption or via a Helmholtz decomposition. For further details we refer the interested reader to Lovadina and Stenberg [10].

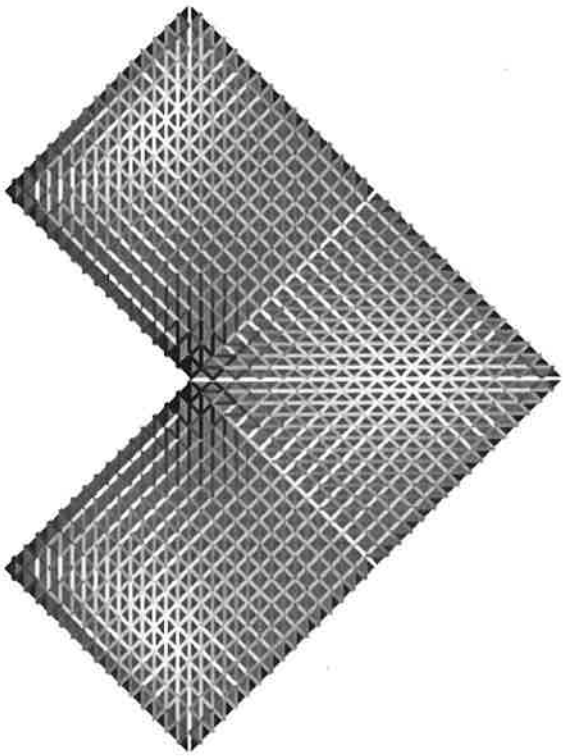


Figure 3: Distribution of the estimated error.

4 Numerical example

In the numerical example, let us consider an L-shape domain which is clamped along the entire boundary $\partial\Omega$ as depicted in Fig. 1. Here, the loading $f = 1$ is equally distributed over the whole domain Ω. As a consequence, the term $\eta_{2,K}$ vanishes in the error estimator (22).

The approximate displacement solution to the model problem, which is visualized in Fig. 2, is obtained using lowest-order Raviart-Thomas finite elements, i.e. we get a piecewise constant solution u in each triangle and a linear approximation for $\boldsymbol{\sigma}$. It should be noted that the homogeneous boundary condition is fulfilled only in a weak sense, i.e. the solution u converges towards zero with finer meshes, whereas in a standard displacement formulation of the model problem the homogeneous boundary condition is a priori fulfilled with the choice of FE-spaces.

In Fig. 3, the qualitative distribution of the estimated error on a uniformly refined mesh with 3904 degrees of freedom is plotted. More precisely, dark areas or edges indicate areas where the estimated error is large (either in the element or on its edges according to (21a) and (21c)), whereas light areas indicate quite small estimated errors. Obviously, the error estimator finds large errors around the edge singularity and small errors in some parts of the interior of the domain Ω.

This qualitative error distribution then explains the adaptive mesh refinements as can be seen in Fig. 4. In this figure, the 23rd adaptively refined mesh is plotted with 19373 degrees of freedom. As can be observed, the mesh is clearly heavily refined around the edge

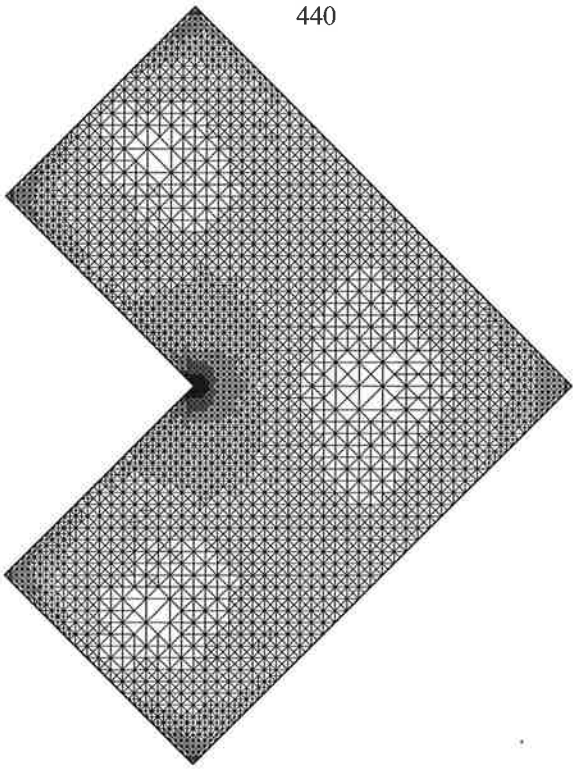


Figure 4: 23rd adaptively refined mesh with 19373 degrees of freedom.

singularity. In this paper, an element is refined if the error indicator on an element divided by the maximum error indicator is larger than a given tolerance which was set here to 0.5.

Finally, the convergence of the estimated error as obtained in terms of (22) is visualized in Fig. 5 on a double logarithmic scale and an optimal convergence can be verified.

5 Conclusions

In this paper, we derived a priori and a posteriori error estimates for the mixed model problem of the Poisson problem. The error estimator relies on a postprocessed finite element solution than can be done locally, i.e. on the element level. With this postprocessed solution at hand both upper and lower bounds on the error can be computed in terms of volume and edge terms. In the computed example we obtained good numerical evidence of the error estimator presented.

REFERENCES

[1] A. Alonso, *Error estimators for a mixed method*, Numer. Math., 74 (1994), pp. 385–395.

[2] D. N. Arnold and F. Brezzi, *Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates*, RAIRO Modél. Math. Anal. Numér., 19 (1985), pp. 7–32.

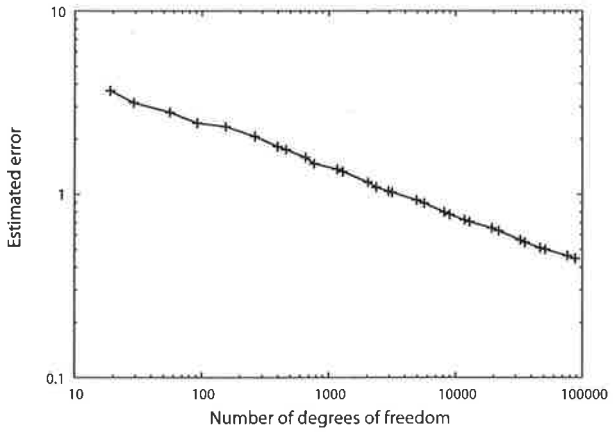


Figure 5: Estimated error.

[3] I. Babuška, J. Osborn, and J. Pitkäranta, *Analysis of mixed methods using mesh dependent norms*, Math. Comp., 35 (1980), pp. 1039–1062.

[4] D. Braess and R. Verfürth, *A posteriori error estimators for the Raviart-Thomas element*, SIAM J. Numer. Anal., 33 (1996), pp. 2431–2444.

[5] J. H. Bramble and J. Xu, *A local post-processing technique for improving the accuracy in mixed finite-element approximations*, SIAM J. Numer. Anal., 26 (1989), pp. 1267–1275.

[6] F. Brezzi, J. Douglas, Jr., R. Durán, and M. Fortin, *Mixed finite elements for second order elliptic problems in three variables*, Numer. Math., 51 (1987), pp. 237–250.

[7] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, 1991.

[8] F. Brezzi, J. Douglas Jr., and L. Marini, *Two families of mixed finite elements for second order elliptic problems.*, Numer. Math., 47 (1985), pp. 217–235.

[9] C. Carstensen, *A posteriori error estimate for the mixed finite element method*, Math. Comp., 66 (1997), pp. 465–476.

[10] C. Lovadina and R. Stenberg, *Energy norm a posteriori error estimates for mixed finite element methods*, Math.Comp. (2006), accepted for publication.

[11] J.-C. Nédélec, *A new family of mixed finite elements in R^3*, Numer. Math., 50 (1986), pp. 57–81.

[12] P. Raviart and J. Thomas, *A mixed finite element method for second order elliptic problems*, in Mathematical Aspects of the Finite Element Method. Lecture Notes in Math. 606, Springer-Verlag, 1977, pp. 292–315.

[13] R. Stenberg, *Some new families of finite elements for the Stokes equations*, Numer. Math., 56 (1990), pp. 827–838.

[14] R. Stenberg, *Postprocessing schemes for some mixed finite elements*, RAIRO Modél. Math. Anal. Numér., 25 (1991), pp. 151–167.

AN ADAPTIVE INTEGRATION SCHEME IN COMPUTATIONAL INELASTICITY

Reijo KOUHIA[1] and Marcus RÜTER[2]

[1] Helsinki University of Technology, Structural Mechanics, PO Box 2100, FIN-02015 TKK
[2] Helsinki University of Technology, Institute of Mathematics, PO Box 1100, FIN-02015 TKK

ABSTRACT

In this paper, an adaptive approach based on a time discontinuous Galerkin method (dG) is proposed. For large time steps and in a linear non-autonomous case, it reduces to the asymptotically second-order accurate Lobatto IIIC type implicit Runge-Kutta method, which is equal to the Padé-(0,2) approximation of the exponential function. In the asymptotic range it will result in the asymptotically third-order accurate scheme, the dG(1) method.

Accuracy and efficiency of the proposed method are studied in detail for common models in creep and plasticity. Comparisons are made to the commonly used backward Euler scheme which is known to give accurate results for large time steps.

1 INTRODUCTION

There are many different algorithms for the integration of inelastic constitutive models. However, the fully implicit backward Euler scheme seems to be the most popular, although it is asymptotically only first-order accurate [1, 2, 3]. In analysing practical problems, especially in creep analysis and viscoplasticity, the time steps to be used are often large, several magnitudes larger than the critical time step of some explicit methods, e.g. the forward Euler method. Therefore, the integrator should be unconditionally stable and sufficiently accurate for large time steps.

As shown in [4], the asymptotic convergence rate does not necessarily reflect high accuracy outside the asymptotic range, which is usually step sizes smaller than the critical time step of the explicit Euler method. The asymptotically first-order accurate implicit Euler method seems to be more accurate than many asymptotically higher-order schemes for large time steps. Therefore, an integrator for inelastic constitutive models should be at least [4]:

1. L-stable
2. for $\dot{\sigma} + \lambda\sigma = 0$ (λ = constant), the amplification factor should be
   (a) strictly positive, and
   (b) monotonous (convex).

It is obvious that the standard backward Euler scheme fulfills these requirements.

The discontinuous Galerkin family of methods seem rather ideal for the integration of inelastic constitutive models. However, they have some shortcomings, as explaned in [4].

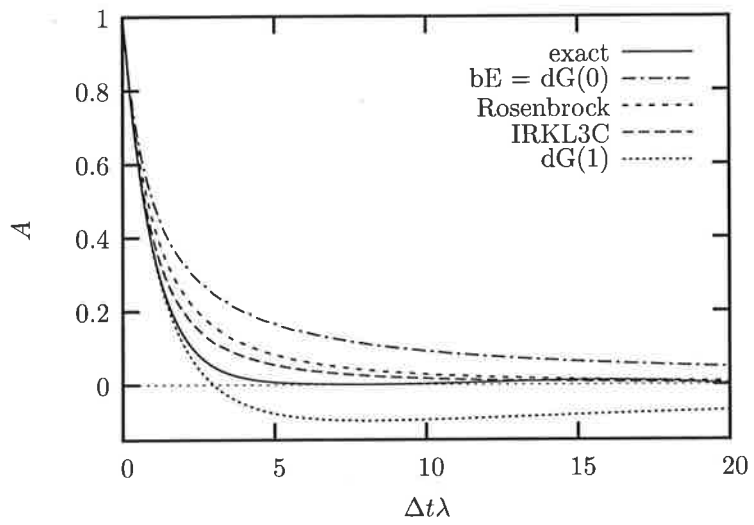


Figure 1: Amplification factors of different time integrators.

The second-order accurate two-stage Lobatto IIIC type implicit Runge-Kutta method (IRKL3C) exhibits good accuracy properties also for large time steps. When the integrals in the dG(1) scheme are underintegrated by using the two-point Gauss-Lobatto scheme (endpoint rule) the dG(1) scheme is identical to the IRKL3C-scheme [5]. The dG(0) and dG(1) schemes have been used in elastoplasticity by Alberty and Carstensen [6, 7]. In [8] the first-order accurate Rosenbrock method is recommended to inviscid plasticity computations.

The amplification factors applied to the linear scalar problem $\dot{\sigma} + \lambda\sigma = 0$ are shown for certain integrators in Fig. 1.

2 NUMERICAL INTEGRATION

In this section, the basic equations of discontinuous Galerkin and implicit Runge-Kutta methods are described. More details can be found in [10, 11].

2.1 The discontinuous Galerkin method

The following evolution problem will be considered:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{f}(\boldsymbol{\sigma}), \tag{1}$$

where $\boldsymbol{f}$ is some function of the stress $\boldsymbol{\sigma}$. The discontinuous Galerkin method of degree q can be stated as: in a time interval $I_n = (t_n, t_{n+1}]$ find $\boldsymbol{\sigma}$ (polynomial of degree q) such that

$$\int_{I_n} (\dot{\boldsymbol{\sigma}} - \boldsymbol{f}(\boldsymbol{\sigma})) : \boldsymbol{\tau} \, \mathrm{d}t + [\![\boldsymbol{\sigma}_n]\!] : \boldsymbol{\tau}_n^+ = 0. \tag{2}$$

For the test functions $\boldsymbol{\tau}$, polynomials of degree q are used. The notations $\boldsymbol{\sigma}_n^+$ and $\boldsymbol{\sigma}_n^-$ are the limits $\boldsymbol{\sigma}_n^\pm = \lim_{\epsilon\to 0} \boldsymbol{\sigma}(t_n \pm |\epsilon|)$, $[\![\boldsymbol{\sigma}_n]\!] = \boldsymbol{\sigma}_n^+ - \boldsymbol{\sigma}_n^-$. These notations are illustrated in Fig. 2.

The discontinuous Galerkin method allows the use of piecewise constant trial and test functions. In this case, $\tau_{ij} = 1$ and $\boldsymbol{\sigma}$ is constant on the time step n and the dG(0) method

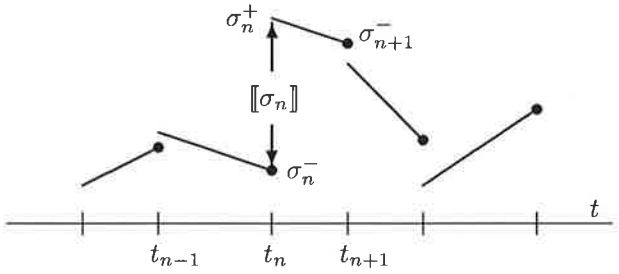


Figure 2: Discontinuous Galerkin method, dG(1); notation.

can be stated as follows

$$\boldsymbol{\sigma}_{n+1} - \int_{I_n} \boldsymbol{f}(\boldsymbol{\sigma}_{n+1})\,\mathrm{d}t = \boldsymbol{\sigma}_n, \tag{3}$$

where $\boldsymbol{\sigma}_{n+1} = \boldsymbol{\sigma}_{n+1}^- = \boldsymbol{\sigma}_n^+$, and $\boldsymbol{\sigma}_n = \boldsymbol{\sigma}_n^-$. If the function $\boldsymbol{f}$ is linear in the stress, i.e $\boldsymbol{f} = \boldsymbol{f}_0 - \mathbb{A} : \boldsymbol{\sigma}$, where $\boldsymbol{f}_0, \mathbb{A}$ are second- and fourth-order tensors independent of the stress, the dG(0) method can be written as:

$$(\mathbb{I} + \tilde{\mathbb{A}}_{n+1}) : \boldsymbol{\sigma}_{n+1} = \boldsymbol{\sigma}_n + \int_{I_n} \boldsymbol{f}_0\,\mathrm{d}t, \tag{4}$$

where

$$\tilde{\mathbb{A}}_{n+1} = \int_{I_n} \mathbb{A}\,\mathrm{d}t \tag{5}$$

and $\mathbb{I}$ is the fourth-order identity tensor. For constant $\mathbb{A}$ and without the source term ($\boldsymbol{f}_0 = \boldsymbol{0}$), the dG(0) method is identical to the implicit backward Euler scheme.

2.2 Implicit Runge-Kutta methods

There is a close connection between the implicit Runge-Kutta (IRK) and discontinuous Galerkin methods. The discontinuous Galerkin method of degree q, dG(q), is equivalent to the $q+1$-stage Radau IIA type IRK-method, which is of order $2q-1$. Their amplification factors are the $(q, q+1)$ subdiagonal Padé approximations of $\exp(-\lambda\Delta t)$ and the methods are L-stable.

An s-stage implicit Runge-Kutta method for the solution of the evolution equation (1) can be defined as

$$\boldsymbol{k}_i = \boldsymbol{\sigma}_n + \Delta t \sum_{j=1}^{s} a_{ij} \boldsymbol{f}(t_n + c_j \Delta t, \boldsymbol{k}_j) \qquad i = 1, \ldots, s \tag{6}$$

$$\boldsymbol{\sigma}_{n+1} = \boldsymbol{\sigma}_n + \Delta t \sum_{j=1}^{s} b_j \boldsymbol{f}(t_n + c_j \Delta t, \boldsymbol{k}_j). \tag{7}$$

For the third order accurate two stage Radau IIA method, the coefficients have the values $a_{11} = 5/12, a_{21} = 3/4, a_{12} = -1/12, a_{22} = 1/4, c_1 = 1/3, c_2 = 1, b_1 3/4, b_2 = 1/4$. This scheme is equal to the dG(1)-method. If the dG(1)-method is integrated by the two-point Gauss-Lobatto rule, is it equal to the second order accurate two stage Lobatto IIIC method, which have the following values for the coefficients: $a_{11} = a_{21} = a_{22} = -a_{12} = 1/2, c_1 = 0, c_2 = 1, b_1 = b_2 = 1/2$.

3 SMALL STRAIN ELASTOPLASTICITY

Next, we briefly present the small strain elastoplasticity problem. The isotropic elastoplastic body is given by the closure of a bounded open set $\Omega \subset \mathbb{R}^3$ with a piecewise smooth, polyhedral and Lipschitz continuous boundary Γ such that $\Gamma = \bar{\Gamma}_D \cup \bar{\Gamma}_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$, where Γ_D and Γ_N are the portions of the boundary Γ where Dirichlet and Neumann boundary conditions are imposed, respectively.

Due to the assumed additive decomposition of the total strain tensor $\boldsymbol{\varepsilon} = \operatorname{sym} \operatorname{grad} \boldsymbol{u}$ into an elastic part $\boldsymbol{\varepsilon}^{\mathrm{e}}$ and a plastic part $\boldsymbol{\varepsilon}^{\mathrm{p}}$, the elastic strains can be determined by

$$\boldsymbol{\varepsilon}^{\mathrm{e}} = \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}}. \tag{8}$$

From the dissipation inequality we then infer the constitutive relation

$$\boldsymbol{\sigma} = \frac{\partial W}{\partial \boldsymbol{\varepsilon}^{\mathrm{e}}} \tag{9}$$

with stress tensor $\boldsymbol{\sigma}$ and specific strain-energy function $W = \frac{1}{2}\boldsymbol{\varepsilon}^{\mathrm{e}} : \mathbb{C}^{\mathrm{e}} : \boldsymbol{\varepsilon}^{\mathrm{e}}$, where we introduced the elasticity tensor

$$\mathbb{C}^{\mathrm{e}} = \varkappa \mathbf{1} \otimes \mathbf{1} + 2\mu \mathbb{P}. \tag{10}$$

Here, $\varkappa$ denotes the bulk modulus, μ is a Lamé parameter, $\mathbf{1}$ is the second-order identity tensor and $\mathbb{P}$ is the fourth-order projection tensor defined as $\mathbb{P} = \mathbb{I} - \frac{1}{3}\mathbf{1} \otimes \mathbf{1}$.

In order to determine the elastic strains (8) and thus the stresses (9) we need to determine the plastic strains. Therefore, we next introduce the yield function $\Phi = \Phi(\boldsymbol{\sigma})$ which defines the closed and convex set

$$\bar{\mathcal{E}} = \{\boldsymbol{\sigma} : \Phi(\boldsymbol{\sigma}) \leq 0\} \tag{11}$$

in stress space. The interior $\mathcal{E}$ of this set is called the elastic domain, whereas the boundary $\partial\mathcal{E}$ is called the yield surface. In this paper, we restrict our considerations to the case of von-Mises plasticity as used to model, e.g., steel structures. In this case, the yield function is defined as

$$\Phi(\boldsymbol{\sigma}) = \| \operatorname{dev} \boldsymbol{\sigma} \| - \sqrt{\frac{2}{3}} y_0 \tag{12}$$

with yield stress y_0 and "dev" denoting the deviator.

From the principle of maximum dissipation we then obtain the associative flow rule

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{p}} = \lambda \boldsymbol{n} \tag{13}$$

and the Kuhn-Tucker conditions, also referred to as loading and unloading conditions,

$$\lambda \geq 0 \quad ; \quad \Phi(\boldsymbol{\sigma}) \leq 0 \quad ; \quad \lambda \Phi(\boldsymbol{\sigma}) = 0. \tag{14}$$

In the above, $\boldsymbol{n}$ is the normal defined as $\boldsymbol{n} = \partial\Phi(\boldsymbol{\sigma})/\partial\boldsymbol{\sigma}$ and λ is the plastic parameter. In the special case of von-Mises plasticity we obtain $\boldsymbol{n} = \operatorname{dev} \boldsymbol{\sigma} / \| \operatorname{dev} \boldsymbol{\sigma} \|$ and $\lambda = \boldsymbol{n} : \dot{\boldsymbol{\varepsilon}}$.

In order to determine the plastic flow and thus the stresses at the end of the time step I_n, we make use of the well-known operator split technique, i.e. we split the problem into an elastic predictor and a plastic corrector step.

In the elastic predictor step we assume that no plastic flow appears. Hence, we solve the initial value problem

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^{\mathrm{e,trial}} \tag{15}$$

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{p}} = \boldsymbol{0} \tag{16}$$

subjected to the initial condition

$$\boldsymbol{\varepsilon}_n^{\mathrm{e},+} = \boldsymbol{\varepsilon}_n^{\mathrm{e,trial},+} \tag{17}$$

for the trial strains $\boldsymbol{\varepsilon}_{n+1}^{\mathrm{e,trial},-}$. The notations are illustrated in Fig. 2.

Integrating (15) over the time interval I_n yields

$$\int_{I_n} \dot{\boldsymbol{\varepsilon}} \, \mathrm{d}t = \int_{I_n} \dot{\boldsymbol{\varepsilon}}^{\mathrm{e,trial}} \, \mathrm{d}t \tag{18}$$

and thus

$$\boldsymbol{\varepsilon}_{n+1}^{-} - \boldsymbol{\varepsilon}_n^{+} = \boldsymbol{\varepsilon}_{n+1}^{\mathrm{e,trial},-} - \boldsymbol{\varepsilon}_n^{\mathrm{e,trial},+}. \tag{19}$$

With the initial condition (17), the exact solution reads

$$\boldsymbol{\varepsilon}_{n+1}^{\mathrm{e,trial}-} = \boldsymbol{\varepsilon}_{n+1}^{-} - \boldsymbol{\varepsilon}_n^{\mathrm{p},+}. \tag{20}$$

Clearly, the stresses at the end of the time interval I_n can then be computed as

$$\boldsymbol{\sigma}_{n+1}^{\mathrm{trial},-} = \mathbb{C}^{\mathrm{e}} : \boldsymbol{\varepsilon}_{n+1}^{\mathrm{e,trial}-}. \tag{21}$$

If the trial stresses are element of $\bar{\mathcal{E}}$, then no plastic flow appears and thus no plastic corrector step is required. If, however, the yield condition is not satisfied by the trial stresses $\boldsymbol{\sigma}_{n+1}^{\mathrm{trial},-}$, then we have to solve the following initial value problem, also referred to as the plastic corrector step,

$$\dot{\boldsymbol{\varepsilon}} + \dot{\boldsymbol{\varepsilon}}^{\mathrm{e,trial}} = \boldsymbol{0} \tag{22}$$

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{p}} = \lambda \boldsymbol{n} \tag{23}$$

subjected to the initial condition

$$\boldsymbol{\varepsilon}_{n+1}^{\mathrm{e},+} = \boldsymbol{\varepsilon}_n^{\mathrm{e,trial},-} \tag{24}$$

and the Kuhn-Tucker conditions (14).

The associated discontinuous Galerkin method dG(q) of polynomial order q (in terms of stresses) then reads: find a solution $\boldsymbol{\sigma}_{n+1}^{-}$ such that

$$b_n(\dot{\boldsymbol{\sigma}}, \boldsymbol{\tau}) + \boldsymbol{\sigma}_n^{+} : \boldsymbol{\tau}_n^{+} = -c_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) + \boldsymbol{\sigma}_n^{-} : \boldsymbol{\tau}_n^{+} \tag{25}$$

with the bilinear form

$$b_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \int_{I_n} \boldsymbol{\sigma} : \boldsymbol{\tau} \, \mathrm{d}t \tag{26}$$

and the semi-linear form (i.e. it is linear only with respect to its second argument)

$$c_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \int_{I_n} \lambda \boldsymbol{n} : \mathbb{C}^{\mathrm{e}} : \boldsymbol{\tau} \, \mathrm{d}t \,. \tag{27}$$

In the special case where $q = 0$, i.e. for piecewise constant functions in time, we can set $\tau_{ij} = 1$ and thus the dG-method (25) reduces to

$$\boldsymbol{\sigma}_n^{+} = -\int_{I_n} \lambda \boldsymbol{n} : \mathbb{C}^{\mathrm{e}} \, \mathrm{d}t + \boldsymbol{\sigma}_n^{-}. \tag{28}$$

Consequently, the stresses at the end of the time interval I_n can be determined by

$$\boldsymbol{\sigma}_{n+1}^{-} = -\Delta t_n \lambda \boldsymbol{n} : \mathbb{C}^{\mathrm{e}} + \boldsymbol{\sigma}_{n+1}^{\mathrm{trial},-}, \tag{29}$$

where Δt_n is the length of the time interval I_n. Evaluating $\boldsymbol{n}$ at the end of the time interval, the classical backward Euler scheme

$$\boldsymbol{\sigma}_{n+1}^{-} = -\Delta t_n \lambda_{n+1} \boldsymbol{n}_{n+1} : \mathbb{C}^{\mathrm{e}} + \boldsymbol{\sigma}_{n+1}^{\mathrm{trial},-} \tag{30}$$

can be recovered.

In the case where $q \geq 1$, we first have to linearise the nonlinear variational problem (25) which yields the linearised problem of solving

$$b_n(\Delta\dot{\boldsymbol{\sigma}}, \boldsymbol{\tau}) + c_{n,\sigma}(\Delta\boldsymbol{\sigma}, \boldsymbol{\tau}) = -b_n(\dot{\boldsymbol{\sigma}}, \boldsymbol{\tau}) - c_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \boldsymbol{\sigma}_n^{+} : \boldsymbol{\tau}_n^{+} + \boldsymbol{\sigma}_{n+1}^{\mathrm{trial},-} : \boldsymbol{\tau}_n^{+} \tag{31}$$

for a stress increment $\Delta\boldsymbol{\sigma}$. In the special case of von-Mises plasticity, the bilinear form $c_{n,\sigma}$ is given as

$$c_{n,\sigma}(\Delta\boldsymbol{\sigma}, \boldsymbol{\tau}) = \int_{I_n} \Delta\boldsymbol{\sigma} : \left[\left(\frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\sigma}} : \dot{\boldsymbol{\varepsilon}} \right) \otimes (\boldsymbol{n} : \mathbb{C}^{\mathrm{e}}) \right] : \boldsymbol{\tau} + \lambda \Delta\boldsymbol{\sigma} : \frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\sigma}} : \mathbb{C}^{\mathrm{e}} : \boldsymbol{\tau} \, \mathrm{d}t, \tag{32}$$

where it was used that

$$\frac{\partial \lambda}{\partial \boldsymbol{\sigma}} = \frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\sigma}} : \dot{\boldsymbol{\varepsilon}} \tag{33}$$

with

$$\frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\sigma}} = \| \operatorname{dev} \boldsymbol{\sigma} \|^{-1} \left(\mathbb{P} - \boldsymbol{n} \otimes \boldsymbol{n} \right). \tag{34}$$

Clearly, the exact solution should be continuous in time. Hence, the jump in the time discretization at the end of each time interval I_n can be used to estimate the error in the time discretization. Consequently, it becomes possible to refine the time mesh adaptively based on the jumps at the end of I_n.

4 NUMERICAL EXAMPLES

In a uniaxial case and using the Garofalo type creep model [12], the evolution equation for the stress (1) has the form

$$\dot{\sigma} = E \left[\dot{\varepsilon} - f^{*} \exp\left(-Q/R\theta\right) \sinh^{m}\left(\sigma/\sigma_{\mathrm{r}}\right) \right], \tag{35}$$

where E is the Young's modulus, f^{*} the fluidity parameter, Q the process activation energy, R the gas constant, θ the absolute temperature and σ_{r} is the "flow stress" stress. In this example, the exponent m and the coefficient f^{*} are assumed to be constant, although they depend on the grain size [12].

Thermally softening and hardening cases have been studied, although the results are only presented from the softening case since the behaviour of the integrators is similar for both cases. The temperature is assumed to change linearly with time $\theta(t) = \theta_0 \pm \Delta\theta(t/t_{\max})$, and $\theta_0 = 293$ K, $\Delta\theta = 40$ K. Increase or decrease in temperature result in softening or hardening behaviour, respectively. The material parameters used correspond to the binary near eutectic Sn40Pb solder and are the following [13]: $E = 33$ GPa, $Q = 12$ kcal/mol, $R = 2{\cdot}10^{-3}$ kcal/mol·K, $\sigma_y = 20$ MPa, $f = 10^5$ s^{-1}, $m = 3.5$.

The behaviour of the dG(1), with two-point Gauss-Legendre integration, the underintegrated dG(1), with two-point Gauss-Lobatto integration, abbreviated as dG(1)-Lobatto, the two stage IRKL3C-scheme and the backward Euler (bE) methods are studied numerically using two loading patterns; a constant strain rate loading ($\varepsilon = \dot{\varepsilon}_0 t$, $\dot{\varepsilon}_0 = 10^{-5}$ 1/s) and a pulsatile cyclic straining, see Fig. 3. As shown in [4, 5] the behaviour of the dG(0)-scheme is inferior as compared to the backward Euler method, hence the results of the dG(0) are not shown here.

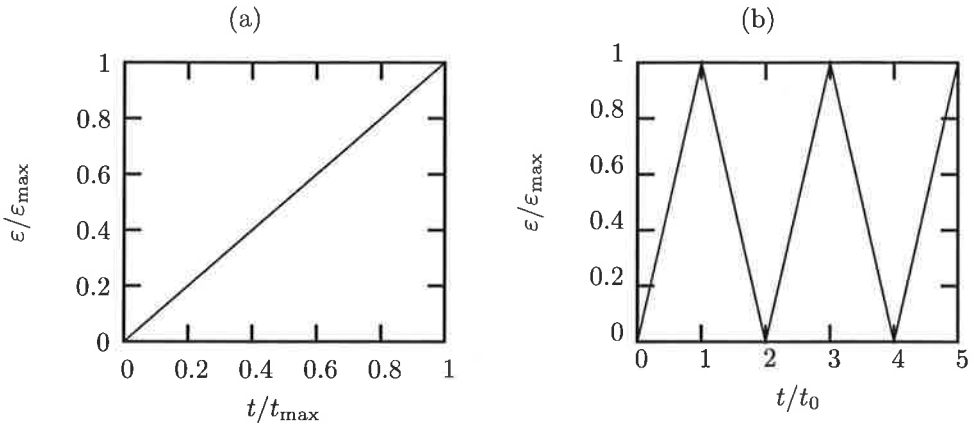


Figure 3: Loading types: (a) constant strain-rate loading $\dot{\varepsilon} = 10^{-5}$ s^{-1}, $\varepsilon_{\max} = 0.01$, $t_{\max} = 1000$ s, (b) pulsatile loading with alternating strain-rates $\dot{\varepsilon} = \pm 10^{-5}$ s^{-1}, $\varepsilon_{\max} = 0.001$, $t_0 = 100$ s.

It can be seen from Fig. 4 that the IRKL3C method is equivalent to the dG(1) method with two point Lobatto integration at the endpoints of the steps. The dG(1) largely overshoots the stress at the end of the first step. However, the linear dG(1) solution is clearly a good approximation in an average sense over the time step. At the strain 0.01 (after the fifth step) the relative error in stress is 0.16 % for the dG(1) method and only 0.02 % for the IRK3LC and dG(1)-Lobatto methods.

It can also be seen from the figure that the backward Euler-scheme perform well in the second, fully inelastic step. Therefore, the most disasterous case for the bE-scheme seems to be a cyclic loading with repeated changes from elastic to inelastic states. In Fig. 5 the results of a pulsatile uniaxial straining case are shown. A large time-step, $\Delta t = t_0 = 100$ s is used. Convergence of the error at the end of loading, $t = t_{\max} = 5t_0$, is shown in Fig. 6; linear, quadratic and cubic rate of convergence in the asymptotic range is clearly visible for the bE, dG(1)-Lobatto and dG(1) methods, respectively. As expected for large step sizes, the underintegrated dG(1) is the most accurate.

5 CONCLUDING REMARKS

Formulation of the symptotically third order accurate dG(1)-scheme for small strain plasticity and creep-type inelastic models is presented. It is also shown that underintegrating the coefficients in the dG(1)-method by the two-point Gauss-Lobatto quadrature results in a quadratically convergent scheme, which has improved accuracy properties for large step sizes, even better than the popular first order accurate backward Euler scheme. The underintegrated dG(1)-method is shown to be equal to the two-stage Lobatto IIIC type implicit Runge-Kutta method.

Further investigations for the dG(1)-type schemes are needed to relate the jump in the stress to a proper error indicator. Also design of a switching strategy to change the quadrature from Gauss-Lobatto to Gauss-Legendre to obtain maximal accuracy for all step sizes is under development.

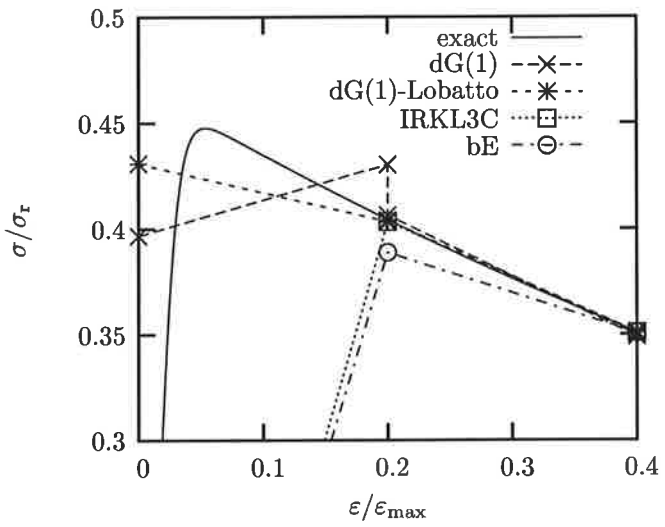


Figure 4: Uniaxial straining, strain rate 10^{-5} s^{-1}.

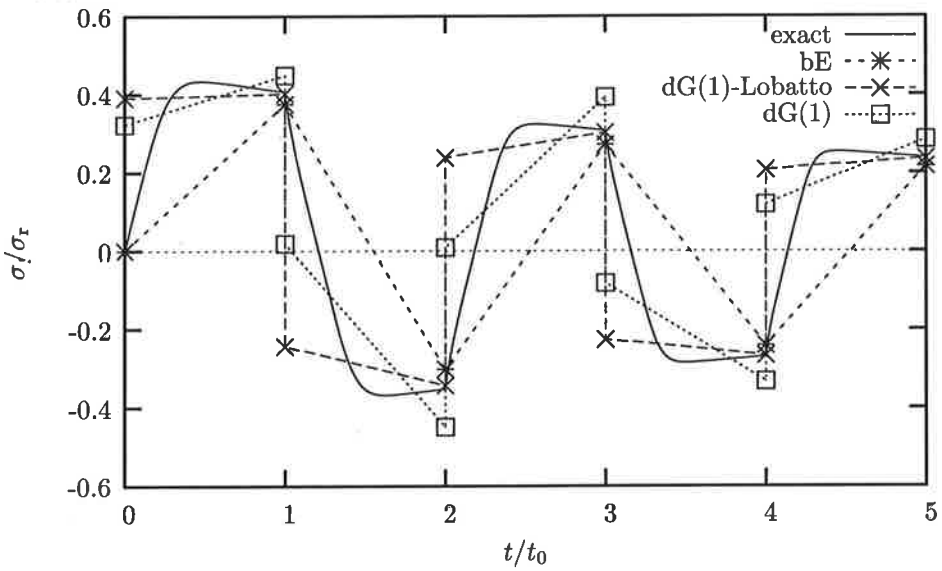


Figure 5: Pulsatile uniaxial straining, strain rate $\pm 10^{-5}$ s^{-1}.

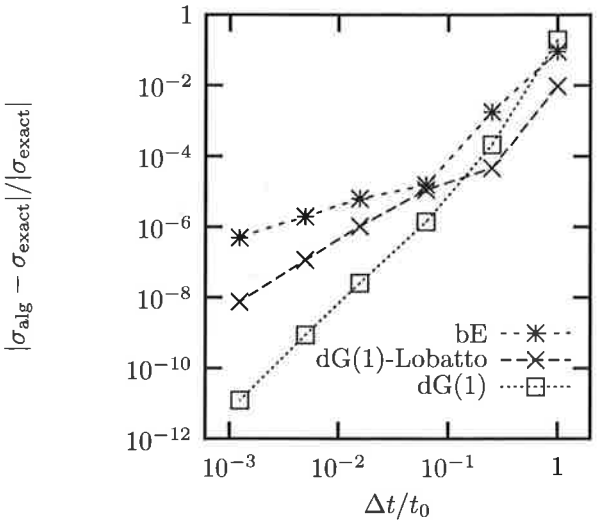


Figure 6: Relative error after five pulses as a function of time step.

ACKNOWLEDGEMENTS

This research has been supported in part by Tekes - the National Technology Agency of Finland (project KOMASI, decision number 40288/05) and by the DFG (German Research Foundation) under grant number Ru1213/1-1.

REFERENCES

[1] J.C. Simo and T.J.R. Hughes. *Computational Inelasticity.* Springer: New York, 1998.

[2] M. Wallin and M. Ristinmaa. *Accurate stress updating algorithm based on constant strain rate assumption.* Comp. Meth. Appl. Mech. Eng. **190** (42), 5583–5601 (2001)

[3] K. Runesson, S. Sture, K. Willam. *Integration in computational plasticity.* Comp. Struct., **30** (1/2), 119–130 (1988).

[4] R. Kouhia, P. Marjamäki, J. Kivilahti. *On the implicit integration of rate-dependent inelastic constitutive models*, Int. J. Numer. Meth. Engng., **62**, 1832–1856 (2005).

[5] R. Kouhia. *A time discontinuous Petrov-Galerkin method for the integration of inelastic constitutive equations*, In P. Neittaanmäki, T. Rossi, K. Majava and O. Pironneau (eds.) ECCOMAS 2004 CD-ROM proceedings.

[6] J. Alberty, C. Carstensen. *Numerical analysis of time-dependent primal elastoplasticity with hardening*, SIAM J. Numer. Anal., **37** (4), 1271–1294 (2000).

[7] J. Alberty, C. Carstensen. *Discontinuous Galerkin time discretization in elastoplasticity: motivation, numerical algorithms, and applications.* Comp. Meth. Appl. Mech. Eng., **191**, 4949–4969 (2002).

[8] K. Hackl, M. Schmidt-Baldassari. *Time integration algorithms for evolution equations in finite strain plasticity.* In W.A.. Wall et al., editors, *Trends in Computational Structural Mechanics*, pages 128–139, CIMNE, Barcelona, (2001).

[9] M. Schmidt-Baldassari, K. Hackl. *Advanced time integration algorithms for rate independent finite strain elastoplasticity.* Preprint 03-1, Ruhr-Universität Bochum, Lehrstuhl für Allgemeine mechanik, (2003).

[10] K. Eriksson, D. Estep, P. Hansbo, C. Johnsson. *Computational Differential Equations.* Studentlitteratur, 1996.

[11] E. Hairer, G. Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems.* Springer, second ed., 1991.

[12] F. Garofalo. *Fundamentals of Creep and Creep-Rupture in Metals.* The MacMillan Co.: New York, 1965.

[13] S.N. Burchett, M.K. Neilsen, D.R. Frear, J.J. Stephens. *Computational continuum modelling of solder interconnects.* In R.K. Mahidhara et al., editors, Design and Reliability of Solders and Solder Interconnects, pp. 171–178, The Minerals, Metals & Materials Society, (1997).

A POSTERIORI ERROR CONTROL IN TERMS OF LINEAR FUNCTIONALS FOR ELLIPTIC PROBLEMS: ESTIMATION AND MESH ADAPTIVITY

A. Hannukainen[1]

Institute of Mathematics, Helsinki University of Technology
P. O. Box 1100, FIN-02015 TKK, Finland
e-mail: antti.hannukainen@hut.fi

ABSTRACT

A general scheme for a posteriori estimation of computational errors in terms of linear functionals for elliptic type boundary value problems is presented. In the framework of this scheme easy-to-code construction of error estimator and error indicator is proposed. Both techniques are studied in several numerical tests, in which the indicator is used to perform mesh adaptivity and the estimator to verify quality of final solution.

1 INTRODUCTION

During the past decades, a posteriori error estimation has become a crucial part of reliable numerical schemes. There exists several main approaches for the construction of such estimates for errors measured in the global energy norm. The most popular approach is to evaluate the negative norm of the residual (see, e.g., [1, 2, 3, 15, 22]). Other relevant approaches include using special post-processing procedures (see [22, 23, 24]) or tools provided by the duality theory [18].

However, engineers are often much more interested in errors over some critical (usually local) part of the domain, or in the errors in some quantities of interest (e.g., for problems in fracture mechanics [19, 20] and electrostatics [17, 21]). Global estimation in the energy norm gives only a general view of the quality of the solution, thus it has limited relevance for most of engineering applications. It is also important to realize, that error measured in the energy norm can be relatively small, while the error over the critical part of the domain is excessively large. For example, in some electrostatic problems, the relative error in the energy norm can be below 1%, while in some parts of the domain it is still above 1000% (see [17]).

Error estimates are often used to provide feedback for adaptive solution strategies aiming to produce meshes minimising the error (see, e.g., [1, 4, 7, 22]). However, if the main interest lies in some local part of the domain or in some quantity of interest, meshes aiming to minimise the corresponding errors can be very different respect to meshes aiming to minimise the global error. Hence, decay of these local errors in adaptive procedures based on error control in the energy norm can be very slow, and more appropriate mesh adaptation methods should be applied.

Consequently, engineering applications require more refined error control techniques for

[1] The author was supported by the project no. 211512 from the Academy of Finland.

estimation of solution quality and providing feedback for adaptive procedures. One common way to obtain such techinques is to introduce a suitable linear functional ℓ related to the crucial part of the domain or to the quantity of interest, and construct a computable a posteriori estimate for $\ell(u - \bar{u})$ (here u denotes the exact solution and $\bar{u}$ is the approximate solution). This approach has been the focus of active research work for the past ten years. Several techniques for such a type of estimation have been presented in [1, 11, 13, 16, 17, 18, 19, 20, 21].

In this note, we describe a general error estimation framework for the error measured in terms of linear functionals. This approach was first suggested in [13] and [18], and it is based on a decomposition of the error $\ell(u - \bar{u})$ into two terms, in which the first term is directly computable and the second term requires separate estimation (see Theorem 1). Such a decomposition naturally leads to two different techniques for constructing error indicators and error estimators.

The construction of indicators is based on the superconvergence property often observed for averaged gradients of the finite element solution for linear elliptic problems [5, 10, 14]. This property actually requires the approximation $\bar{u}$ to be computed with the finite element method, thus the construction applies only to exact finite element solutions. Moreover, the resulting estimates only serve as preliminary, but fast-computable, error indicators. In fact, they provide a good indication of the error $\ell(u - \bar{u})$ with low computational costs, but do not guarantee real estimation of the error from above or from below.

Estimators are constructed by reducing error control in terms of linear functionals to the error control in the energy norm. With suitable error control methods in the energy norm, this technique is capable of building guaranteed two-sided (upper and lower) bounds for the error $\ell(u - \bar{u})$. These bounds are completely independent of the method used to obtain the (admissible) approximation $\bar{u}$. The bounds are also consistent, i.e., they can be made arbitrarily close to the true error. However, the price for having sufficiently sharp guaranteed bounds is that more computer resources (and time) are required compared to corresponding computatonal costs needed for indicators.

The presented methods are complementary, e.g., the indicator can serve as feedback for mesh adaptation (see [11] for a definition of the relevant adaptive strategy). The estimator can be employed later on, when guaranteed estimation of quality of the solution is really essential.

Effectivity and main features of both error estimation techniques are demonstrated in several numerical tests. These tests aim to demonstrate how error over local part of the domain behaves in adaptive procedure based on the presented indicator. The obtained result is compared with error behavior in adaptive procedure based on the error measured in the energy norm. Our aim is to emphasize the fact, that adaptive procedure in the energy norm is not always suitable, namely when interest lies in error over local part of the domain. Finally, quality of the last solution obtained with adaptive procedure is verified by computing two-sided bounds using the estimator.

Some previous numerical tests on adaptivity based on same type of indicator can be found, e.g., form [11, 19] and results on the effectivity of the estimator in [9]. Similar comparison between the two different error measures, but using different error control tools is performed in [1, 21].

2 MODEL PROBLEM

In order to describe the approach in short we shall deal only with a simple scalar elliptic model problem, whose classical formulation reads: Find a function u such that

$$-\mathrm{div}(A\nabla u) = f \quad \text{in} \quad \Omega, \quad u = 0 \quad \text{on} \quad \partial\Omega. \tag{1}$$

In what follows, $\Omega \subset \mathbf{R}^d$ is a bounded connected domain with a Lipschitz continuous boundary $\partial\Omega$, $f \in L_2(\Omega)$, the matrix of coefficients A is symmetric and positive definite, with bounded entries $a_{ij} \in L_\infty(\Omega)$, $i,j = 1,\ldots,d$, i.e.,

$$c_2|\xi|^2 \geq A(x)\xi \cdot \xi \geq c_1|\xi|^2 \qquad \forall \xi \in \mathbf{R}^d \quad \forall x \in \Omega. \tag{2}$$

We use the weak formulation of the problem (1), which reads : Find $u \in H_0^1(\Omega)$ such that

$$a\,(u,w) = F\,(w) \quad \forall w \in H_0^1(\Omega), \tag{3}$$

where bilinear form and linear functional are defined as

$$a\,(v,w) = \int_\Omega A\nabla v \cdot \nabla w \; dx \qquad F\,(w) = \int_\Omega fw \; dx. \tag{4}$$

Under the above conditions, the solution of problem (3) exists and is unique [8].

3 ERROR CONTROL IN TERMS OF LINEAR FUNCTIONALS

Let $\bar{u}$ be some function from $H_0^1(\Omega)$ considered as an approximation (computed by the finite element method or some another numerical technique) of solution u to problem (3). In truly reliable simulations it is necessary to control the error $e := u - \bar{u}$ during computational process in some suitable way(s).

In what follows, our interest lies in the value of e measured in terms of some bounded linear functional ℓ, i.e., in estimation of

$$\ell(e) = \ell(u - \bar{u}), \tag{5}$$

in the following also called the error. Note that estimates for such an error also allow estimation of the value $\ell(u)$ (called quantity of interest or goal-oriented quantity [1, 16, 17, 19, 20, 21]), since $\ell(u) = \ell(u - \bar{u}) + \ell(\bar{u})$, in which $\ell(\bar{u})$ is computable and $\ell(u - \bar{u})$ is estimated.

Example 1: One example of a possible linear functional for the error control is

$$\ell(w) \;=\; \int_\Omega \varphi w \; dx, \;\; w \in H_0^1(\Omega), \tag{6}$$

where $\varphi \in L_2(\Omega)$ and $\mathrm{supp}\,\varphi = \omega \subseteq \Omega$. In this case, the value $\ell(e)$ gives a certain information about the local behaviour of the error in subdomain ω, and its reliable estimation can help to control local errors and perform a mesh adaptivity (see, e.g., [11, 12, 13]).

3.1 Principal error decomposition

In estimation of the error (5), a common practice is to employ an auxiliary (called adjoint or dual) problem [1, 4, 13, 16, 21], whose right-hand side is formed by the linear functional: Find $v \in H_0^1(\Omega)$ such that

$$a\,(v,w) = \ell\,(w) \quad \forall w \in H_0^1(\Omega). \tag{7}$$

The solution of problem (7) is known to exist and to be unique [8]. Similarly to problem (1), called the primal problem, we can rarely find the exact solution v of problem (7) analytically. Hence, in practice, only the approximation $\bar{v} \in H_0^1(\Omega)$ is available.

Theorem 1. The following error decomposition holds

$$\ell(u - \bar{u}) = E_0(\bar{u}, \bar{v}) + E_1(e, e_\ell), \tag{8}$$

where

$$E_0(\bar{u}, \bar{v}) = F(\bar{v}) - a(\bar{v}, \bar{u}), \quad E_1(e, e_\ell) = a(e, e_\ell), \tag{9}$$

and $e_\ell = v - \bar{v}$.

Proof: Using weak formulation (3) and (7) we observe that

$$\begin{aligned} \ell(e) &= a(e_\ell, e) - a(\bar{v}, \bar{u}) + a(\bar{v}, u) \\ &= F(\bar{v}) - a(\bar{v}, \bar{u}) + a(e_\ell, e) = E_0(\bar{u}, \bar{v}) + E_1(e_\ell, e). \end{aligned}$$ □

The term E_0 in (8) contains only known approximate solutions $\bar{u}$ and $\bar{v}$, so it is directly computable. Estimate for the error $\ell(e)$ only requires estimation of the term E_1. This estimation is performed in two different ways: by using superconvergence property of averaged gradient (indication) and by reducing the problem to the error estimation in the energy norm and applying suitable two-sided estimates (estimation).

3.2 Error indicator constructed via averaged gradients

In construction of error indicators we assume, that $\bar{u}$ and $\bar{v}$ are computed by the finite element method. To emphasize this situation we denote these finite element approximations by u_h and v_τ, where the symbols h and τ are discretization parameters related to finite element schemes used to approximately solve problems (3) and (7), respectively.

It is well known that averaged gradients of finite element approximations for linear elliptic problems usually demonstrate a superconvergence effect (see, e.g., [5, 10, 14]). This suggests to replace the unknown gradients in (9) by easily computable averaged gradients and finally to estimate (5) by the following error indicator

$$\tilde{E}(u_h, v_\tau) := E_0(u_h, v_\tau) + \tilde{E}_1(u_h, v_\tau), \tag{10}$$

where

$$\tilde{E}_1(u_h, v_\tau) = \int_\Omega \mathrm{A}(G_\tau(\nabla v_\tau) - \nabla v_\tau) \cdot (G_h(\nabla u_h) - \nabla u_h)\, dx, \tag{11}$$

and G_h and G_τ are suitable gradient averaging operators. A particular choice of averaging technique defines a concrete form of the above indicator.

Remark 1: Asymptotic convergence of estimator (10) is analysed in [13], where it is proved, under some regularity conditions on the meshes, that

$$\tilde{E} \to \ell(u - u_h) \quad \text{as} \quad v_\tau \to v.$$

Later, such an asymptotic behaviour was supported by various numerical tests in [11] even if meshes used to solve primal and adjoint problems are nonuniform.

3.3 Error estimator constructed via reduction to the energy norm

Estimation of the term E_1 in (8) can be reduced to error estimation in the energy norm, $|||v||| = \sqrt{a(v,v)}$, with two techniques. The first method is a direct application of the Cauchy-Schwartz inequality, wich implies

$$-|||e_\ell|||e||| \leq E_1(e_\ell, e) \leq |||e_\ell|||e|||. \tag{12}$$

However, over/underestimation of E_1 can be substantial. A more refined estimate is obtained by noticing that

$$2E_1(e_\ell, e) = |||\alpha e + \frac{1}{\alpha} e_\ell|||^2 - \alpha^2 |||e|||^2 - \frac{1}{\alpha^2} |||e_\ell|||^2, \tag{13}$$

which is valid for any positive α. The above identity contains the errors in the energy norm for both primal and adjoint problems. For the first term in the right-hand side of (13), we observe that

$$|||\alpha e + \frac{1}{\alpha} e_\ell|||^2 = ||| \left(\alpha u + \frac{1}{\alpha} v \right) - \left(\alpha \bar{u} + \frac{1}{\alpha} \bar{v} \right) |||^2. \tag{14}$$

It is clear that function $\alpha u + \frac{1}{\alpha} v$ is the solution of the following (mixed) problem: Find $u_\alpha \in H_0^1(\Omega)$ such that

$$a(u_\alpha, w) = \alpha F(w) + \frac{1}{\alpha} \ell(w) \quad \forall w \in H_0^1(\Omega), \tag{15}$$

and function $\alpha \bar{u} + \frac{1}{\alpha} \bar{v}$ can be considered as an approximation of u_α.

By using either identity (12) or (13), we can apply any two-sided a posteriori error estimates in the energy norm to obtain two-sided estimates for the error (5). Let us denote the required two-sided estimates as

$$M_F^\ominus \leq |||e|||^2 \leq M_F^\oplus, \quad M_\ell^\ominus \leq |||e_\ell|||^2 \leq M_\ell^\oplus, \quad M_{\alpha,F,\ell}^\ominus \leq |||\alpha e + \frac{1}{\alpha} e_\ell|||^2 \leq M_{\alpha,F,\ell}^\oplus. \tag{16}$$

Now, based on identity (13), we immediately observe that

$$\frac{1}{2}(M_{\alpha,F,\ell}^\ominus - \alpha^2 M_F^\oplus - \frac{1}{\alpha^2} M_\ell^\oplus) \leq E_1(e, e_\ell), \tag{17}$$

and

$$E_1(e, e_\ell) \leq \frac{1}{2}(M_{\alpha,F,\ell}^\oplus - \alpha^2 M_F^\ominus - \frac{1}{\alpha^2} M_\ell^\ominus), \tag{18}$$

which together with the computable term E_0 provide two-sided estimates for the error (5). In this approach, the main task is to apply suitable a posteriori error estimates in the energy norm and to select parameter α.

Parameter α can be selected, for example, by noticing that

$$\begin{aligned} 2E_1(e_\ell, e) &\leq |||\alpha e + \frac{1}{\alpha} e_\ell|||^2 - \alpha^2 M_F^\ominus - \frac{1}{\alpha^2} M_\ell^\ominus \\ &= 2a(e, e_\ell) + \alpha^2 |||e|||^2 + \frac{1}{\alpha^2} |||e_\ell|||^2 - \alpha^2 M_F^\ominus - \frac{1}{\alpha^2} M_\ell^\ominus \\ &\leq 2a(e, e_\ell) + \alpha^2 (M_F^\oplus - M_F^\ominus) + \frac{1}{\alpha^2} (M_\ell^\oplus - M_\ell^\ominus). \end{aligned} \tag{19}$$

Now, minimising (19) with respect to α yields

$$\alpha = \sqrt[4]{\frac{M_\ell^\oplus - M_\ell^\ominus}{M_F^\oplus - M_F^\ominus}}. \tag{20}$$

4 ERROR ESTIMATES IN THE ENERGY NORM

To compute bounds (17) and (18), we apply the following two-sided estimates for error control in the global energy norm. The upper estimate has the form

$$|||e|||^2 \le \left(1 + \frac{1}{\beta}\right) \frac{c_\Omega^2}{c_1} \|f + \operatorname{div} y^*\|_{0,\Omega}^2 + (1+\beta)\, |||\nabla \bar{u} - A^{-1} y^*|||_A^2, \tag{21}$$

where y^* is any function in $H(\operatorname{div};\Omega)$, parameter $\beta > 0$, $\|\cdot\|_{0,\Omega}$ denotes the standard $L_2(\Omega)$-norm and

$$|||\mathbf{v}|||_A := \int_\Omega A\mathbf{v} \cdot \mathbf{v}\, dx.$$

In the following, this upper bound is denoted as $M^\oplus(u_h, \beta, y^*)$. The constant c_1 in (21) is the smallest eigenvalue of the coefficient matrix A and c_Ω is the constant from the Poincaré-Friedrichs inequality

$$\|v\|_{0,\Omega} \le c_\Omega |v|_{1,\Omega}.$$

In practice, value for this constant is determined by the smallest eigenvalue λ_Ω of the Laplacian for Ω as $c_\Omega = \frac{1}{\sqrt{\lambda_\Omega}}$. Note, that several other existing estimation techniques (e.g., of the residual type) involve several unknown constants. These constants are usually related to patches of the computational meshes and their estimation is very difficult, and, in general leads to a very big overestimation of the error even in simple cases (see [6]). Moreover, such constants have to be recomputed if adaptive computations changing the computational mesh are performed, whereas constants c_Ω and c_1 remain the same under any changes in the computational mesh.

Altough, the upper bound can be applied for every approximation $\bar{u} \in H_0^1$, the practical realization is presented here in the context of finite element method. To compute value for $M^\oplus(u_h, \beta, y^*)$ parameters y^* and β have to be selected. For each y^* optimal (minimizing) β can be obtained with simple minimization of $M^\oplus(u_h, \beta, y^*)$, so in practice only selecting y^* is required.

Finite element computations are often performed using series of successive meshes $\mathcal{T}_{h_1}$, $\mathcal{T}_{h_2}$, $\mathcal{T}_{h_3}, \ldots$, where $h = h_1 > h_2 > h_3 > \ldots$. By using this information, a coarse upper bounds can be obtained using values $y^* = G_\mu(u_\mu) \in H(\operatorname{div}, \Omega)$, where $\mu = h_1, h_2, h_3, \ldots$ and G_μ is some commonly used gradient averaging operator [10]. However, sharper upper-bound requires a real minimisation of $M^\oplus(u_h, \beta, y^*)$ with respect to variables y^* and β. In practice, minimisation requires parameter $y^* \in H(\operatorname{div};\Omega)$ to be selected from some finite dimensional subspace of $H(\operatorname{div};\Omega)$. For convinience, we have chosen $y^* \in (S_{h/2^n})^2$, in which $S_{h/2^n}$ denotes to the subspace of standard linear basis functions connected to n-times uniformly refined mesh T_h. From this subspace, y^* is selected with direct minimisation of $M^\oplus(u_h, \beta, y^*)$, or by finding the minimiser as a solution of the respective system of linear equations.

Lower bound for $|||e|||$ is derived by noticing that

$$\|u - w\|^2 = 2(J(w) - J(u)) \quad \forall w \in H_0^1(\Omega), \tag{22}$$

where $J(w) = \frac{1}{2}a(w,w) - F(w)$ is the energy functional. Solution u for the weak problem (3) minimises value of the energy functional, i.e., $J(u) \leq J(w) \quad \forall w \in H_0^1(\Omega)$. Using this fact, one can obtain a lower bound as follows

$$\|e\|^2 \geq 2(J(\bar{u}) - J(w)), \tag{23}$$

where w is any function from $H_0^1(\Omega)$.

The lower-estimate (23) has a practical meaning only if it provides positive lower bound for the (positive) error $\|e\|$. One usually tries to have several successive approximations u_{h_1}, u_{h_2}, $u_{h_3}, \ldots$, satisfying $J(u_{h_1}) > J(u_{h_2}) > J(u_{h_3}) > \ldots$. This situation immediately suggests a meaningful lower bound of the form

$$\|e\|^2 = \|u - u_h\|^2 \geq 2(J(u_h) - J(u_\mu)) > 0, \tag{24}$$

where $\mu = h_2, h_3, \ldots$.

5 NUMERICAL EXPERIMENTS

In this section, the proposed techniques are applied to two numerical tests performed in planar domains (d=2) in the context of the finite element method. To emphasize this situation, we denote $\bar{u} = u_h$ and $\bar{v} = v_\tau$.

The adaptive procedure applied in the tests is schematically described in Figure 1. In each adaptive step a new mesh $\mathcal{T}_{i+1}$ is formed by refining the old mesh $\mathcal{T}_i$. The refinement is performed by using information provided by the error distribution: elements having contributions bigger than a given treshold $\theta \in [0,1]$ times the maximal value are refined.

The required error distribution can be constructed in several different ways. The aim in following tests is to minimise the error $\ell(u - u_h)$, so the natural choice is to use error distribution for error measured in terms of linear functionals. In fact, other error distributions are no suitable for this kind of error control. For example, in some cases, error distribution for error measured in the energy norm produces very poor meshes for the error $\ell(u - u_h)$.

In the follwing tests, we study this situation by comparing adaptive procedure based on error distribution computed by the indicator (10) in terms of linear functionals and a similar indicator for the error in the energy norm. The indicator in the energy norm is based on the superconvergence phenomenon of averaged gradients, and it is of the form

$$\|u - u_h\|^2 \approx \int_\Omega A\,(G_h(u_h) - \nabla u_h) \cdot (G_h(u_h) - \nabla u_h)\; dx, \tag{25}$$

where G_h is some gradient averaging operator. Indicator in terms of linear functionals (10) has the form

$$\ell(u - u_h) \approx E_0(u_h, v_\tau) + \tilde{E}_1(u_h, v_\tau). \tag{26}$$

Computing value for the indicator requires solving the adjoint problem. We solve this problem on adjoint meshes containg approximately the same number of nodes as the corresponding primal mesh.

The performance of the two error distributions are compared by monitoring the development of error $\ell(u - u_h)$. This monitoring is accomplished by the means of reference error, which is obtained by approximating $\ell(u - u_h) \approx \ell(u_{h/2^n}) - \ell(u_h)$. Here $u_{h/2^n}$ is solution obtained from n-times uniformly refined initial mesh.

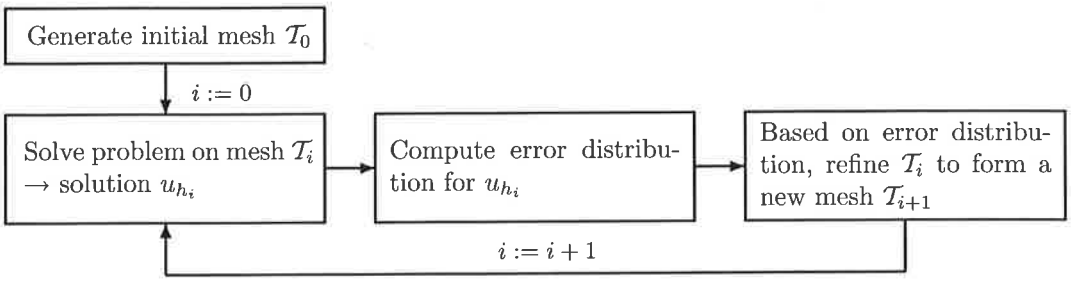


Figure 1: Schematic presentation of the adaptive procedure used in numerical tests

Finally, two-sided bounds are used to verify thr quality of last solution obtained from the adaptive procedure. To illustrate the effect, which adjoint problem has on the quality of the bounds, the adjoint solution is computed on several different meshes. Parameters required in the computation of upper bound for primal, adjoint and mixed problems are denoted as $y_h^*, y_\tau^*, y_\alpha^*$, respectively. The parameters y_h^* and y_τ^* are selected from spaces $(S_{h/2^n})^2, n = 1, 2, \ldots$, and $(S_{\tau/2^n})^2, n = 1, 2, \ldots$, with minimization of $M^{\oplus}(u_h, \beta, y^*)$. To reduce computational costs the parameter y_α^*, used for the mixed problem, is constructed from these values as $y_\alpha^* = \alpha y_h^* + \frac{1}{\alpha} y_\tau^*$.

Parameters required for the lower bound of primal and adjoint problem, denoted as u_μ and v_μ, are solutions from spaces $S_{h/2^{(n+1)}}$ and $S_{\tau/2^{(n+1)}}$, respectively. Parameter $u_{\alpha,\mu}$, required in the lower bound of the mixed problem is obtained by adaptively solving (with error distibution in the energy norm) the mixed problem. This adaptive procedure is stopped, when number of nodes in the generated mesh is equal to number of nodes in the primal mesh.

5.1 Test 1

In this test the failure of adaptive procedure based on error control in the energy norm is demonstrated. The problem (3) is posed in a rectangular domain Ω (see Figure 2) with coefficient matrix $A = I$ and the right hand side function $f = 1$ in Ω_1, $f = 20$ in Ω_2. Here Ω_1 and Ω_2 are partitions of domain Ω, i.e., $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$ and $\Omega_1 \cap \Omega_2 = \emptyset$. The interest in error control lies in the local error over $\omega \subset \Omega$ (see Figure 2) measured in terms of the linear functional (6)

$$\ell(w) = \int_\Omega \varphi w \, dx,$$

where the function $\varphi = 1$ on ω , $\varphi = 0$ on $\Omega \setminus \omega$.

The behavior of the reference error in adaptive refinement process is visualized in Figure 3. The reference error is obtained by using solution from 6-times uniformly refined initial mesh with 24369 nodes and 48128 elements. In this case, the adaptive procedure based on linear functionals yields better convergence rate. Actually, in this case, error control in the energy norm produces no convergence at all. One reason behind this situation can be found from the very different nature of the refinements (see Figures 6 and 7). Procedure based on energy norm refines at those parts of the domain, where the gradient of the solution undergoes rapid changes (cf. Figure 4). Thus, this method produces very little refinement near the zone of interest (see Figure 5 for the initial mesh). On the contrary, approach based on linear functionals produces mesh with heavy refinements near the zone of interest

and yields locally more accurate solution.

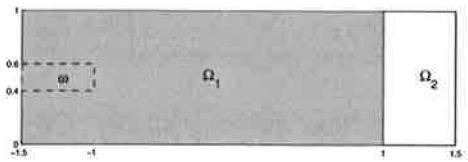


Figure 2: The solution domain Ω, partitions Ω_1, Ω_2 and the zone of interest ω for Test 1.

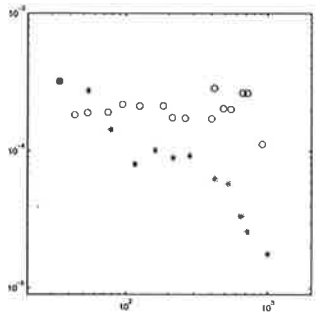


Figure 3: Comparison between two different mesh adaptivity procedures for Test 1. Stars correspond to the error control in terms of linear functionals and circles to the error control in the energy norm.

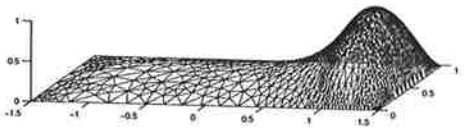


Figure 4: Approximate solution for Test 1 computed on the mesh in Figure 7.

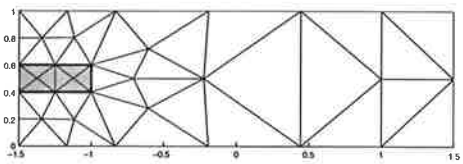


Figure 5: Initial mesh for Test 1 with 34 degrees of freedom.

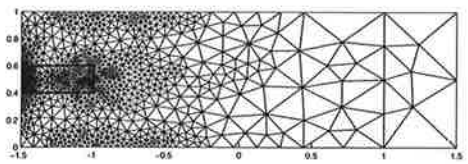


Figure 6: Final mesh (1002 degrees of freedom) from mesh adaptivity in terms of linear functional for Test 1.

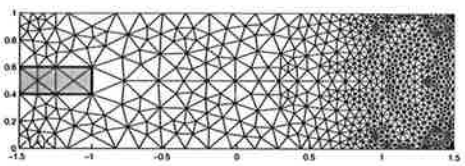


Figure 7: Final mesh (909 degrees of freedom) from mesh adaptivity in terms of the energy norm for Test 1.

Two-sided bounds for the final solution are plotted in Figure 8. The reference error for this solution is $1.76 \cdot 10^{-5}$. The bounds computed by selecting spaces n=1, produce bounds from $-0.0471 \leq e \leq 0.0495$ (the adjoint mesh with 34 nodes) to $-0.0141 \leq e \leq 0.0141$ (the adjoint mesh with 160 nodes). Clearly, better approximate solution to adjoint problem yields better bounds. Even larger improvement can be seen, when bounds are computed on larger space n=2. Now, the obtained bound improve and range from $-0.0220 \leq e \leq 0.0235$ to $-0.0069 \leq e \leq 0.0067$.

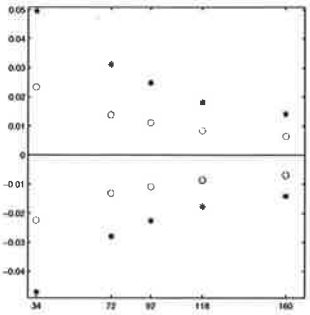


Figure 8: Developement of two-sided bounds with respect to different adjoint meshes for Test 1. Stars correspond to parameter space with n=1 and circles with n=2. Solid line is the reference error.

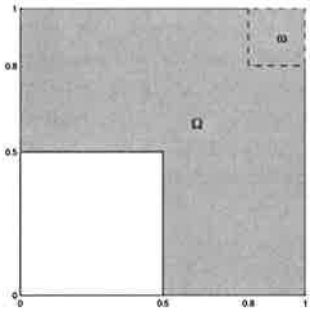


Figure 9: The solution domain Ω and the zone of interest ω for Test 2.

5.2 Test 2

In this test, the L-shaped domain Ω (see Figure 9) is considered. Coefficient matrix $A = I$ and the right hand side function $f \equiv 10$. Again, the interest lies in the local error form the zone of interest $\omega \subset \Omega$ (see Figure 9) measured with the same liner functional as in Test 1.

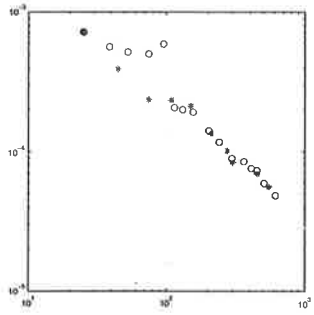


Figure 10: Comparison between two different mesh adaptivity procedures in Test 2. Stars correspond to error control in terms of linear functionals and circles to error control in the energy norm.

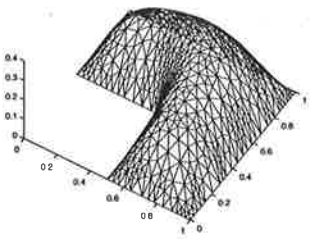


Figure 11: Approximate solution for Test 2 computed on mesh in Figure 14.

The behavior of the reference error with respect to the degrees of freedom in both adaptive refinement processes is plotted in Figure 10. The reference error is obtained by using solution from 6-times uniformly refined initial mesh with 15649 nodes and 30720 elements. In this test problem, both approaches produce very similar convergence. However, the obtained meshes have very different nature. The energy norm based approach refines heavily at the corner singularity (see Figure 14), whereas the linear functional based approach refines mainly near the zone of interest (see Figure 13). Note that the linear functionals also produce some refinement at the corner singularity, probably, to avoid the pollution error.

Two-sided bounds for the solution obtained using the final mesh are plotted in Figure 8. The reference error for this solution is $5.65 \cdot 10^{-5}$. Computations performed using spaces

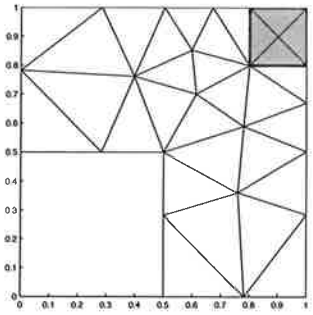


Figure 12: Initial mesh with 25 degrees of freedom for Test 2

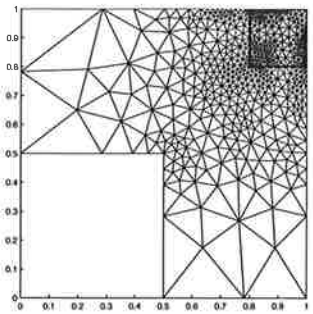


Figure 13: Final mesh (551 degrees of freedom) from mesh adaptivity in terms of linear functional.

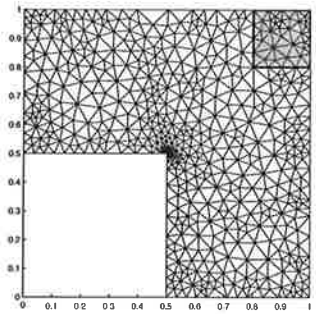


Figure 14: Final mesh (611 degrees of freedom) from mesh adaptivity in terms of the energy norm for Test 2.

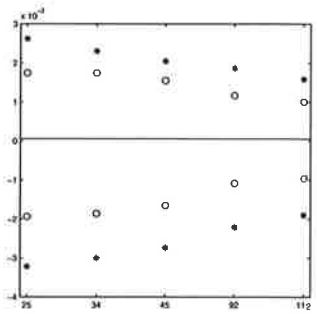


Figure 15: Developement of two-sided bounds with respect to different adjoint meshes for Test 2. Stars correspond to the parameter space with n=1 and circles with n=2. Solid line denotes the reference error.

n=1, produce bounds from $-0.0032 \leq e \leq 0.0026$ (the adjoint mesh with 25 nodes) to $-0.0019 \leq e \leq 0.0016$ (the adjoint mesh with 112 nodes). Again, better approximate solution to adjoint problem yields better bounds and large improvement can be seen, when bounds are computed using space n=2. Now, the obtained bounds improves and range from $-0.0019 \leq e \leq 0.0017$ to $-0.0009 \leq e \leq 0.0010$.

6 ACKNOWLEDGEMENTS

The author is greatful to Dr. Sergey Korotov for his useful suggestions during preparation of this paper.

REFERENCES

[1] M. Ainsworth, J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis.* John Wiley & Sons, Inc., 2000.

[2] I. Babuška, W. C. Rheinbold. Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.*, 15, 736–754, 1978.

[3] I. Babuška, T. Strouboulis. *The Finite Element Method and Its Reliability.* Oxford University Press Inc., New York, 2001.

[4] R. Becker, R. Rannacher. A feed-back approach to error control in finite element methods: Basic analysis and examples. *East-West J. Numer. Math.*, 4, 237–264, 1996.

[5] J. Brandts, M. Křížek. Gradient superconvergence on uniform simplicial partitions of polytopes. *IMA J. Numer. Anal.*, 23, 489–505, 2003.

[6] C. Carstensen, S. A. Funken. Constants in Clément-interpolation error and residual based a posteriori error estimates in finite element methods. *East-West J. Numer. Math.*, 8, 153–175, 2000.

[7] K. Eriksson, D. Estep, P. Hansbo, C. Johnson. Introduction to adaptive methods for differential equations. *Acta Numer.*, 106–158, 1995.

[8] L. C. Evans. *Partial Differntial Equations*. AMS, Providence, 1998.

[9] A. Hannukainen, S. Korotov. Techniques for a posteriori error estimation in terms of linear functionals for elliptic type boundary value problems, *Far East J. Appl. Math.*, 21, 289–304, 2005.

[10] I. Hlaváček, M. Křížek. On a superconvergent finite element scheme for elliptic systems. I, II, III, *Apl. Mat.*, 32 131–154, 200–213, 276–289, 1987.

[11] S. Korotov. A posteriori error estimation of goal-oriented quantities for elliptic type BVPs. *J. Comput. Appl. Math.*, 191, 216–227, 2006.

[12] S. Korotov. New technologies for control of local errors in engineering computations by FEM. *Rakenteiden Mekaniikka (Journal of Structure Mechanics)*, 38, 117–120, 2005.

[13] S. Korotov, P. Neittaanmäki, S. Repin. A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery. *J. Numer. Math.*, 11, 33–59, 2003.

[14] M. Křížek, P. Neittaanmäki. Superconvergence phenomenon in the finite element method arising from averaging gradients. *Numer. Math.*, 45, 105–116, 1984.

[15] C. Lovadina, R. Stenberg, Energy norm a posteriori error estimates for mixed finite element methods. *Math. Comp.* (to appear).

[16] J. T. Oden, S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method. *Comput. Math. Appl.*, 41, 735–756, 2001.

[17] D. Pardo, L. Demkowicz, C. Torres-Verdn, L. Tabarovsky. A goal-oriented hp-adaptive finite element method with electromagnetic applications. Part I: electrostatics. *Internat. J. Numer. Methods Engrg.*, 65, 1269–1309, 2006.

[18] S. Repin. Two-sided estimates of deviation from exact solutions of uniformly elliptic equations. *Amer. Math. Soc. Transl.*, 209, 143–171, 2003.

[19] M. Rüter, S. Korotov, Ch. Steenbock. Goal-oriented error estimates based on different FE-solution spaces for the primal and the dual problem with application to linear elastic fracture mechanics, *Comput. Mech.* (in press).

[20] M. Rüter, E. Stein. Goal-oriented a posteriori error estimates in linear fracture mechanics. *Comput. Methods Appl. Mech. Engrg.*, 195, 251–278, 2006.

[21] P. Šolín, L. Demkowicz. Goal-oriented *hp*-adaptivity for elliptic problems. *Comput. Methods Appl. Mech. Engrg.*, 193, 449–468, 2004.

[22] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley-Teubner, 1996.

[23] O. C. Zienkiewicz, J. Z. Zhu. The superconvergence patch recovery and a posteriori error estimates. Part 1: The recovery technique. *Internat. J. Numer. Methods Engrg.*, 33, 1331–1364, 1992.

[24] O. C. Zienkiewicz, J. Z. Zhu. The superconvergence patch recovery and a posteriori error estimates. Part 2: Error estimates and adaptivity. *Internat. J. Numer. Methods Engrg.*, 33, 1365–1382, 1992.

EPÄJATKUVA GALERKININ MENETELMÄ SEKAELEMENTTITEHTÄVÄLLE

M. JUNTUNEN R. STENBERG

Teknillinen korkeakoulu
Matematiikan laitos
PL 1100, 02015 TKK
e-mail: mika.juntunen@tkk.fi

TIIVISTELMÄ

Esittelemme epäjatkuvan Galerkinin menetelmän yhdistämisen sekaelementtitehtävään. Malliongelmana on sekamuodossa esitetty Poissonin tehtävä. Tehtävän numeerinen ratkaisu muodostetaan epäjatkuvalla Galerkinin menetelmällä, jossa elementit on liitetty toisiinsa Nitschen menetelmällä [1]. Työn pääpaino on a posteriori virhearvion todistamisessa, joka tässä työssä todistetaan Helmholtzin hajotelman [5] avulla. Lisäksi numeeristen esimerkkien avulla tarkastellaan menetelmän ja a posteriori virhearvion toimintaa. Erityisesti näytetään a posteriori virhearvion hyvä soveltuminen adaptiiviseen verkon tihentämiseen.

1 JOHDANTO

Epäjatkuvassa Galerkinin menetelmässä jatkuvuutta voidaan pakottaa monilla eri tavoilla [8]. Eräs nykyään varsin suosittu menetelmä perustuu J.A. Nitschen tapaan käsitellä reunaehtoja [1]. Nitschen menetelmän suurin etu esimerkiksi sakkomenetelmään verrattuna on menetelmän konsistenssi. Pieni haitta Nitschen menetelmässä on että menetelmä ei ole elliptinen, symmetrisessä muodossa, ilman parametria, jolla painotetaan bilineaarimuodon stabiloivia termejä. Parametrin arvo riippuu verkon elementtien muodosta ja valituista vapausasteista, esimerkiksi polynomiasteesta. Parametrin arvo voidaan laskea ennalta, mutta esimerkiksi p-tihennystä (approksimaation parantaminen polynomiastetta kasvattamalla) käytettäessä parametrin arvon muuttuminen on otettava huomioon.

Epäjatkuvan Galerkinin menetelmän ja Nitschen menetelmän yhdistäminen on esitetty ennenkin myös sekaelementtitehtävälle, katso esim. [8]. Tässä työssä näytämme että sekatehtävälle esitetty Nitschen menetelmä on elliptinen riippumatta stabiloivien termien painoista, kun lähtökohtana oleva tehtävä on elliptinen, ja esittelemme menetelmälle a posteriori virhearvion.

Tässä työssä käytämme mallitehtävänä Poissonin tehtävää, joka on esitetty sekamuodossa. Toisin sanoen ratkaisun gradientti on oma muuttujansa. Sekatehtävän käsittely luonnollisesti kasvattaa tehtävän kokoa. Tämä ei kuitenkaan käytännössa ole suuri ongelma, sillä gradientin vapausasteet voidaan kondensoida pois jo lokaalista matriisista, joten varsinaisen ratkaistavan yhtälöryhmän koko pysyy ennallaan. Lisäksi gradientin polynomiaste on yhtä pienempi kuin ratkaisun polynomiaste, joten lokaalin jäykkyysmatriisin kokoamiseen tarvittava työmäärä ei kasva paljoa, vaikka kondensoinnissa vaaditaankin gradientin lokaalien vapausasteiden ratkaisua yhtälöryhmästä.

A posteriori virhearvion todistuksessa käytetään Helmholtzin hajotelmaa, jossa gradientti jaetaan lähteettömään ja pyörteettömään osaan [5], [6]. Helmholtzin hajotelmaan perustuvat todistukset soveltuvat hyvin myös moniin muihin tapauksiin. Tämän tekniikan etuna on mahdollisuus käyttää tavallisia Clément'n interpolantteja vaikka ratkaisua haettaisiinkin epäjatkuvasta avaruudesta sillä interpolantteja tarvitaan vain hajotelman funktioille, jotka voidaan valita jatkuviksi.

Artikkelin sisältö on seuraava: kappaleessa 2 esittelemme tehtävän ja notaatiota ja johdamme Nitschen menetelmän variaatiomuodon. Kappaleessa 3 näytämme että variaatiomuoto on elliptinen riippumatta stabiloivien termien painoista. Kappaleessa 4 esittelemme ja todistamme a posteriori virhearvion. Lopuksi kappaleessa 5 tarkastelemme numeerisia tuloksia.

2 MALLITEHTÄVÄ JA ELEMENTTI APPROKSIMAATIO

Tässä kappaleessa esittelemme mallitehtävän ja johdamme tehtävän variaatiomuodon. Olkoon $\Omega \subset \mathbb{R}^N$ äärellinen alue, jonka reuna $\partial\Omega$ on paloittain sileä. Merkitään alueen kolmiointia $\mathcal{T}_h$:lla. $\mathcal{E}_{\partial\Omega}$ merkitsee elementtien sivuja, jotka ovat reunalla $\partial\Omega$ ja $\mathcal{E}_{\text{int}}$ merkitsee verkon sisäsivuja. Merkitään vielä h_T:lla elementin $T \in \mathcal{T}_h$ kokoa ja h_E:lla elementin sivun $E \in \mathcal{E}_{\text{int}} \cup \mathcal{E}_{\partial\Omega}$ kokoa. Ratkaistaan tehtävää

$$\begin{aligned} -\Delta u &= f \quad \Omega\text{:ssa}, \\ u &= g \quad \partial\Omega\text{:lla}, \end{aligned} \tag{1}$$

missä kuorma $f \in L^2(\Omega)$ ja reunaehto $g \in H^{1/2}(\partial\Omega)$. Esitetään tehtävä ekvivalentissa sekamuodossa

$$\begin{aligned} \nabla \cdot \boldsymbol{\sigma} &= -f \quad \Omega\text{:ssa}, \\ \nabla u &= \boldsymbol{\sigma} \quad \Omega\text{:ssa}, \\ u &= g \quad \partial\Omega\text{:lla}. \end{aligned} \tag{2}$$

Haetaan tälle tehtävälle variaatiomuoto. Määritellään aluksi tarvittavat avaruudet

$$\begin{aligned} V_h &:= \{ v \in L^2(\Omega) \mid v_{|T} \in \mathcal{P}_n \ \forall T \in \mathcal{T}_h \} \quad \text{ja} \\ W_h &:= \{ \boldsymbol{v} \in [L^2(\Omega)]^N \mid \boldsymbol{v}_{|T} \in [\mathcal{P}_{n-1}]^N \ \forall T \in \mathcal{T}_h \}, \end{aligned} \tag{3}$$

missä $\mathcal{P}_n$ merkitsee n asteisia polynomeja. Yllä reuna-arvot pitää käsittää Ω:ssa määriteltyjen funktioiden jälkinä $\partial\Omega$:lla. Kertomalla keskimäinen yhtälö (2):ssa testifunktiolla $\boldsymbol{\tau} \in W_h$ ja integroimalla alueen yli saadaan

$$(\boldsymbol{\sigma}, \boldsymbol{\tau})_\Omega - (\nabla u, \boldsymbol{\tau})_\Omega = 0. \tag{4}$$

Kertomalla ylin yhtälö (2):ssa testifunktiolla $v \in V_h$ ja osittaisintegroimalla saadaan

$$\begin{aligned} (-f, v)_\Omega = (\nabla \cdot \boldsymbol{\sigma}, v)_\Omega &= \sum_{T \in \mathcal{T}_h} (\nabla \cdot \boldsymbol{\sigma}, v)_T \\ &= \sum_{T \in \mathcal{T}_h} \{ -(\boldsymbol{\sigma}, \nabla v)_T + (\boldsymbol{\sigma} \cdot \boldsymbol{n}, v)_{\partial T} \} \\ &= \sum_{T \in \mathcal{T}_h} -(\boldsymbol{\sigma}, \nabla v)_T + \sum_{E \in \mathcal{E}_{\text{int}}} (\{\boldsymbol{\sigma} \cdot \boldsymbol{n}\}, [\![v]\!])_E + \sum_{E \in \mathcal{E}_{\partial\Omega}} (\boldsymbol{\sigma} \cdot \boldsymbol{n}, v)_E, \end{aligned} \tag{5}$$

missä olemme käyttäneet hyväksi vuon normaalikomponentin jatkuvuutta ja merkinneet

$$\begin{aligned} \{\boldsymbol{\sigma} \cdot \boldsymbol{n}\} &:= \frac{1}{2}(\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2) \cdot \boldsymbol{n}_1 \\ [\![v]\!] &:= v_1 - v_2. \end{aligned}$$

Yllä alaindeksi viittaa vierekkäisiin elementteihin ja $\boldsymbol{n}_i$ kyseisen elementin ulkonormaaliin. Dirichlet'n ehtoja ei ole asetettu ratkaisuavaruuteen V_h, joten reunaehto pitää asettaa variaatiomuotoon. Lisäksi ratkaisun jatkuvuus pitää pakottaa variaatiomuotoon. Ratkaisu u on jatkuva ja yhtälön (2) perusteella $u_{|\partial\Omega} = g$, joten

$$\sum_{E\in\mathcal{E}_{\text{int}}} -\frac{\gamma}{h_E}\left([\![u]\!],[\![v]\!]\right)_E = 0 \quad \text{ja} \tag{6}$$

$$\sum_{E\in\mathcal{E}_{\partial\Omega}} -\frac{\gamma}{h_E}(u,v)_E = \sum_{E\in\mathcal{E}_{\partial\Omega}} -\frac{\gamma}{h_E}(g,v)_E, \tag{7}$$

missä $\gamma \in \mathbb{R}$, $\gamma > 0$ on vapaa parametri. Yhtälö (6) pakottaa jatkuvuutta ja yhtälö (7) Dirichlet'n reunaehtoa. Mallitehtävä on symmetrinen, joten myös variaatiomuoto on syytä asettaa symmetriseksi. Jälleen jatkuvuuden ja Dirichlet'n reunaehdon perusteella saadaan

$$\sum_{E\in\mathcal{E}_{\text{int}}} \left(\{\boldsymbol{\tau}\cdot\boldsymbol{n}\},[\![u]\!]\right)_E = 0 \quad \text{ja} \tag{8}$$

$$(\boldsymbol{\tau}\cdot\boldsymbol{n},u)_{\partial\Omega} = (\boldsymbol{\tau}\cdot\boldsymbol{n},g)_{\partial\Omega}. \tag{9}$$

Lisäämällä yhtälöt (8) ja (9) variaatiomuotoon, saadaan variaatiomuoto symmetriseksi. Yhdistämällä yhtälöt (5), (6), (7), (8) ja (9) saadaan tehtävän variaatiomuoto, yhtälö (10).

Menetelmä. Etsi $(u_h,\boldsymbol{\sigma}_h) \in V_h \times W_h$ siten että

$$a(u,\boldsymbol{\sigma};v,\boldsymbol{\tau}) = \mathcal{L}(v,\boldsymbol{\tau}) \quad \forall (v,\boldsymbol{\tau}) \in V_h \times W_h, \tag{10}$$

missä

$$\begin{aligned} a(u,\boldsymbol{\sigma};v,\boldsymbol{\tau}) := & \sum_{T\in\mathcal{T}_h} \left[(\boldsymbol{\sigma},\boldsymbol{\tau})_T - (\nabla u,\boldsymbol{\tau})_T - (\boldsymbol{\sigma},\nabla v)_T\right] \\ & + \sum_{E\in\mathcal{E}_{\text{int}}} \left[\left(\{\boldsymbol{\sigma}\cdot\boldsymbol{n}\},[\![v]\!]\right)_E + \left(\{\boldsymbol{\tau}\cdot\boldsymbol{n}\},[\![u]\!]\right)_E\right] \\ & + \sum_{E\in\mathcal{E}_{\partial\Omega}} \left[(\boldsymbol{\sigma}\cdot\boldsymbol{n},v)_E + (\boldsymbol{\tau}\cdot\boldsymbol{n},u)_E\right] \\ & - \sum_{E\in\mathcal{E}_{\text{int}}} \frac{\gamma}{h_E}(u,v)_E - \sum_{E\in\mathcal{E}_{\partial\Omega}} \frac{\gamma}{h_E}\left([\![u]\!],[\![v]\!]\right)_E \end{aligned} \tag{11}$$

ja

$$\mathcal{L}(v,\boldsymbol{\tau}) := \sum_{T\in\mathcal{T}_h} (-f,v)_T + \sum_{E\in\mathcal{E}_{\partial\Omega}} (\boldsymbol{\tau}\cdot\boldsymbol{n},g)_E - \sum_{E\in\mathcal{E}_{\partial\Omega}} \frac{\gamma}{h_E}(g,v)_E. \tag{12}$$

Menetelmän johdon perusteella on selvää että tehtävä on konsistentti eli vahvan tehtävän (2) ratkaisu toteuttaa myös heikon tehtävän (10).

Tehtävän energianormi on

$$|||v,\boldsymbol{\tau}|||^2 := \sum_{T\in\mathcal{T}_h} \left[\|\boldsymbol{\tau}\|^2_{L^2(T)} + \|\nabla v\|^2_{L^2(T)}\right] + \sum_{E\in\mathcal{E}_{\text{int}}} \frac{1}{h_E}\|[\![v]\!]\|^2_{L^2(E)} + \sum_{E\in\mathcal{E}_{\partial\Omega}} \frac{1}{h_E}\|v\|^2_{L^2(E)}. \tag{13}$$

Jotta ehdotettu menetelmä voidaan näyttää jatkuvaksi ja elliptiseksi yllä määritellyssä normissa, tarvitaan seuraava lemma [3].

Lemma 1. *On olemassa positiivinen vakio C_I siten että*

$$h_E\|\boldsymbol{\tau}\|^2_{L^2(\partial T)} \le C_I\|\boldsymbol{\tau}\|^2_{L^2(T)} \quad \forall \boldsymbol{\tau} \in W \text{ ja } \forall T \in \mathcal{T}_h. \tag{14}$$

Huomautus 1. Vakion C_I arvo on helppo laskea esimerkiksi paloittain vakioiden tapauksessa, mutta korkeammille polynomiasteille tehtävä on jo vaikeampi. Itse asiassa juuri tämän vakion arvosta riippuu tehtävälle (1) asetetun Nitschen menetelmän elliptisyyteen tarvittavan parametrin arvo.

Lemman 1 avulla on helppo näyttää että esitetty bilineaarimuoto $a(\cdot,\cdot;\cdot,\cdot)$ ja lineaarinen funktionaali $\mathcal{L}(\cdot,\cdot)$ ovat jatkuvia normissa $|||\cdot,\cdot|||$.

3 ELLIPTISYYS

Tässä kappaleessa näytämme että esittämämme menetelmä on elliptinen. Todistus perustuu Youngin epäyhtälöön ja bilineaarimuodon lineaarisuuteen. Erityisesti näytämme että parametrin $\gamma > 0$ arvo ei vaikuta elliptisyyteen.

Lause 1. *On olemassa positiivinen vakio C siten että*

$$\sup_{(v,\boldsymbol{\tau})\in V_h\times W_h} \frac{a(u,\boldsymbol{\sigma};v,\boldsymbol{\tau})}{|||v,\boldsymbol{\tau}|||} \geq C|||u,\boldsymbol{\sigma}||| \quad \forall (u,\boldsymbol{\sigma})\in V_h\times W_h. \tag{15}$$

Huomautus 2. Lauseen (1) mukaan esitetty bilineaarimuoto on elliptinen, mistä seuraa ratkaisun olemassaolo ja yksikäsitteisyys.

Todistus. Todetaan aluksi että

$$a(u,\boldsymbol{\sigma};-u,\boldsymbol{\sigma}) = \sum_{T\in\mathcal{T}_h}\|\boldsymbol{\sigma}\|^2_{L^2(T)} + \sum_{E\in\mathcal{E}_{\text{int}}}\frac{\gamma}{h_E}\|[\![u]\!]\|^2_{L^2(E)} + \sum_{E\in\mathcal{E}_{\partial\Omega}}\frac{\gamma}{h_E}\|u\|^2_{L^2(E)}. \tag{16}$$

Valitaan $\boldsymbol{\kappa}\in W_h$ siten että $\boldsymbol{\kappa} = \nabla u$, mistä seuraa

$$(\boldsymbol{\kappa},\nabla u)_T = \|\nabla u\|^2_{L^2(T)} \quad \text{ja} \quad \|\boldsymbol{\kappa}\|_{L^2(T)} \leq \|\nabla u\|_{L^2(T)}. \tag{17}$$

Edellä valitun $\boldsymbol{\kappa}$:n ja Schwarzin epäyhtälön avulla saadaan

$$\begin{aligned}
&a(u,\boldsymbol{\sigma};0,-\boldsymbol{\kappa})\\
&= \sum_{T\in\mathcal{T}_h}\left[-(\boldsymbol{\sigma},\boldsymbol{\kappa})_T + (\nabla u,\boldsymbol{\kappa})_T\right] - \sum_{E\in\mathcal{E}_{\text{int}}}(\{\boldsymbol{\kappa}\cdot\boldsymbol{n}\},[\![u]\!])_E - \sum_{E\in\mathcal{E}_{\partial\Omega}}(\boldsymbol{\kappa}\cdot\boldsymbol{n},u)_E\\
&\geq \sum_{T\in\mathcal{T}_h}\left[\|\nabla u\|^2_{L^2(T)} - \|\boldsymbol{\sigma}\|_{L^2(T)}\|\boldsymbol{\kappa}\|_{L^2(T)}\right]\\
&\quad - \sum_{E\in\mathcal{E}_{\text{int}}}\frac{1}{2}\left[h_E^{1/2}\|\boldsymbol{\kappa}_1\cdot\boldsymbol{n}_1\|_{L^2(E)}h_E^{-1/2}\|[\![u]\!]\|_{L^2(E)} + h_E^{1/2}\|\boldsymbol{\kappa}_2\cdot\boldsymbol{n}_1\|_{L^2(E)}h_E^{-1/2}\|[\![u]\!]\|_{L^2(E)}\right]\\
&\quad - \sum_{E\in\mathcal{E}_{\partial\Omega}} h_E^{1/2}\|\boldsymbol{\kappa}\cdot\boldsymbol{n}\|_{L^2(E)}h_E^{-1/2}\|[\![u]\!]\|_{L^2(E)}.
\end{aligned} \tag{18}$$

Käyttämällä edelliseen lemmaa 1 ja Youngin epäyhtälöä saadaan

$$\begin{aligned}
a(u,\boldsymbol{\sigma};0,-\boldsymbol{\kappa}) \geq &\sum_{T\in\mathcal{T}_h}\|\nabla u\|^2_{L^2(T)} - \frac{1}{2\delta}\|\boldsymbol{\sigma}\|^2_{L^2(\Omega)} - \frac{\delta}{2}\sum_{T\in\mathcal{T}_h}\|\nabla u\|^2_{L^2(T)}\\
&- \frac{C_I\delta}{2}\sum_{T\in\mathcal{T}_h}\|\nabla u\|^2_{L^2(T)} - \frac{1}{2\delta}\sum_{E\in\mathcal{E}_{\text{int}}}\frac{1}{h_E}\|[\![u]\!]\|^2_{L^2(E)}\\
&- \frac{C_I\delta}{2}\sum_{T\in\mathcal{T}_h}\|\nabla u\|^2_{L^2(T)} - \frac{1}{2\delta}\sum_{E\in\mathcal{E}_{\partial\Omega}}\frac{1}{h_E}\|u\|^2_{L^2(E)}.
\end{aligned} \tag{19}$$

Valitsemalla Youngin epäyhtälöstä tullut parametri $\delta > 0$ sopivasti saadaan

$$\begin{aligned} a(u,\boldsymbol{\sigma};0,-\boldsymbol{\kappa}) \geq & -C_1\|\boldsymbol{\sigma}\|^2_{L^2(\Omega)} + C_2 \sum_{T\in\mathcal{T}_h} \|\nabla u\|^2_{L^2(T)} \\ & - C_3 \sum_{E\in\mathcal{E}_{\text{int}}} \frac{1}{h_E}\|[\![u]\!]\|^2_{L^2(E)} - C_4 \sum_{E\in\mathcal{E}_{\partial\Omega}} \frac{1}{h_E}\|u\|^2_{L^2(E)}, \end{aligned} \tag{20}$$

missä vakiot C_1, C_2, C_3 ja C_4 ovat positiivisia. Lineaarisuuden perusteella, yhdistämällä kaavat (16) ja (20), saadaan

$$\begin{aligned} a(u,\boldsymbol{\sigma},-u,\boldsymbol{\sigma}-\epsilon\boldsymbol{\kappa}) \geq & (1-\epsilon C_1)\|\boldsymbol{\sigma}\|^2_{L^2(\Omega)} + \epsilon C_2 \sum_{T\in\mathcal{T}_h} \|\nabla u\|^2_{L^2(T)} \\ & + (\gamma - \epsilon C_3) \sum_{E\in\mathcal{E}_{\text{int}}} \frac{1}{h_E}\|[\![u]\!]\|^2_{L^2(E)} + (\gamma-\epsilon C_4) \sum_{E\in\mathcal{E}_{\partial\Omega}} \frac{1}{h_E}\|u\|^2_{L^2(E)}. \end{aligned} \tag{21}$$

Valitsemalla parametri ϵ siten että

$$\epsilon > 0,\ \epsilon < \frac{1}{C_1},\ \epsilon < \frac{\gamma}{C_3} \text{ ja } \epsilon < \frac{\gamma}{C_4} \tag{22}$$

saadaan yhtälöstä (21)

$$a(u,\boldsymbol{\sigma},-u,\boldsymbol{\sigma}-\epsilon\boldsymbol{\kappa}) \geq C|||u,\boldsymbol{\sigma}|||^2 \tag{23}$$

riippumatta parametrin γ arvosta. Muuttujan $\boldsymbol{\kappa}$ ominaisuuksien (17) perusteella on selvää että

$$|||-u,\boldsymbol{\sigma}-\epsilon\boldsymbol{\kappa}||| \leq C|||u,\boldsymbol{\sigma}|||. \tag{24}$$

Sijoittamalla yhtälöt (23) ja (24) yhtälön (15) vasemmalle puolelle saadaan väite. □

4 A POSTERIORI VIRHEARVIO

Tässä kappaleessa esittelemme ja todistamme a posteriori virhearvion. Virhearvion todistus perustuu Helmholtzin hajotelmaan, jossa vektori jaetaan pyörteettömään ja lähteettömään osaan. Tämän tekniikan hyvä puoli on että tarvitsemme interpolantteja vain jatkuville funktioille, vaikka ratkaisu onkin vain paloittain jatkuva. Yksinkertaisuuden vuoksi käymme todistuksen läpi $\mathbb{R}^2$:ssa ($N = 2$).

Lause 2. *On olemassa positiivinen vakio C siten että*

$$|||u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||| \leq C\left(\sum_{T\in\mathcal{T}_h}\eta_T^2\right)^{1/2}, \tag{25}$$

missä

$$\begin{aligned} \eta_T^2 := & h_T^2\|\nabla\cdot\boldsymbol{\sigma}_h + f\|^2_{L^2(T)} + \|\boldsymbol{\sigma}_h - \nabla u_h\|^2_{L^2(T)} \\ & + h_E\,\|[\![\boldsymbol{\sigma}_h\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\mathcal{E}_{int})} + \frac{1}{h_E}\|[\![u_h]\!]\|^2_{L^2(\partial T\cap\mathcal{E}_{int})} + \frac{1}{h_E}\|u_h-g\|^2_{L^2(\partial T\cap\mathcal{E}_{\partial\Omega})}. \end{aligned} \tag{26}$$

Lauseen 2 todistuksessa tarvitsemme seuraava lausetta [7].

Lause 3. *Vektorille $\boldsymbol{\tau}\in[L^2(\Omega)]^N$ on olemassa $\psi\in H_0^1(\Omega)$ ja $q\in H^1(\Omega)/\mathbb{R}$ siten että*

$$\boldsymbol{\tau} = \nabla\psi + \operatorname{curl} q \quad \textit{ja} \quad \|\boldsymbol{\tau}\|^2_{L^2(\Omega)} = \|\nabla\psi\|^2_{L^2(\Omega)} + \|\operatorname{curl} q\|^2_{L^2(\Omega)}\,. \tag{27}$$

Lausessa 3 on käytetty operaattoria curl, joka on määritelty

$$\operatorname{curl} v := \begin{pmatrix} -\frac{\partial v}{\partial x_2} \\ \frac{\partial v}{\partial x_1} \end{pmatrix}, \tag{28}$$

kun $v \in H^1(\Omega)$ ja $\Omega \subset \mathbb{R}^2$. Edellistä vastaavasti määritellään reunan tangentti

$$\boldsymbol{t} := \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} -n_2 \\ n_1 \end{pmatrix}, \tag{29}$$

missä $\boldsymbol{n}$ on sivun ulkonormaali, ja operaattori $\nabla\times$

$$\nabla \times \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} := \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2}. \tag{30}$$

Todistus. (*Lause 2*) Koska tarkka ratkaisu on jatkuva ja toteuttaa reunaehdot, saadaan

$$\begin{aligned} |||u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h|||^2 = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|^2_{L^2(\Omega)} + \sum_{T \in \mathcal{T}_h} \|\nabla u - \nabla u_h\|^2_{L^2(T)} \\ + \sum_{E \in \mathcal{E}_{\text{int}}} \frac{1}{h_E} \|[\![u_h]\!]\|^2_{L^2(E)} + \sum_{E \in \mathcal{E}_{\partial\Omega}} \frac{1}{h_E} \|g - u_h\|^2_{L^2(E)}. \end{aligned} \tag{31}$$

Yhtälön (31) kaksi jälkimmäistä termiä kuuluvat indikaattoriin η_T, joten estimoitavaksi jää vain yhtälön (31) kaksi ensimmäistä termiä. Aloitetaan ensimmäisestä termistä. Normin määritelmän ja lauseen 3 perusteella

$$\begin{aligned} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{L^2(\Omega)} &= \sup_{\boldsymbol{\tau} \in [L^2(\Omega)]^N} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\tau})_\Omega}{\|\boldsymbol{\tau}\|_{L^2(\Omega)}} \\ &\leq \sup_{\psi \in H^1_0(\Omega)} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \nabla\psi)_\Omega}{\|\nabla\psi\|_{L^2(\Omega)}} + \sup_{q \in H^1(\Omega)/\mathbb{R}} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \operatorname{curl} q)_\Omega}{\|\nabla q\|_{L^2(\Omega)}}. \end{aligned} \tag{32}$$

Ryhdytään tarkastelemaan yhtälön (32) ensimmäistä termiä. Koska $\psi \in H^1_0(\Omega)$, tälle on olemassa jatkuva ja elementeittäin n-asteinen interpolaatiopolynomi $I_h\psi$, joka häviää alueen reunalla. Lisäksi interpolantti toteuttaa

$$\sum_{T \in \mathcal{T}_h} h_T^{-1} \|\psi - I_h\psi\|_{L^2(T)} + \sum_{E \in \mathcal{E}_{\text{int}}} h_E^{-1/2} \|\psi - I_h\psi\|_{L^2(E)} \leq C\|\nabla\psi\|_{L^2(\Omega)}. \tag{33}$$

Yhtälöstä

$$a(0, \boldsymbol{\sigma}_h; I_h\psi, 0) = \mathcal{L}(I_h\psi, 0) \tag{34}$$

saadaan osittaisintegroimalla

$$\sum_{T \in \mathcal{T}_h} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \nabla I_h\psi)_T = 0. \tag{35}$$

Edellä saadun ortogonaalisuuden perusteella

$$
\begin{aligned}
&(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, \nabla\psi)_\Omega = \sum_{T\in\mathcal{T}_h} (\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, \nabla(\psi - I_h\psi))_T \\
&= \sum_{T\in\mathcal{T}_h} \left[-(\nabla\cdot(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h), \psi - I_h\psi)_T + ((\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\cdot\boldsymbol{n}, \psi - I_h\psi)_{\partial T}\right] \\
&= \sum_{T\in\mathcal{T}_h} (\nabla\cdot\boldsymbol{\sigma}_h + f, \psi - I_h\psi)_T + \sum_{E\in\mathcal{E}_{\text{int}}} (\llbracket(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\cdot\boldsymbol{n}\rrbracket, \psi - I_h\psi)_E \\
&\le \sum_{T\in\mathcal{T}_h} h_T \|\nabla\cdot\sigma_h + f\|_{L^2(T)}\, h_T^{-1}\|\psi - I_h\psi\|_{L^2(T)} \\
&\quad + \sum_{E\in\mathcal{E}_{\text{int}}} h_E^{1/2}\,\|\llbracket\boldsymbol{\sigma}_h\cdot\boldsymbol{n}\rrbracket\|_{L^2(E)}\, h_E^{-1/2}\,\|\psi - I_h\psi\|_{L^2(E)} \\
&\le C\left\{\sum_{T\in\mathcal{T}_h} h_T^2\,\|\nabla\cdot\boldsymbol{\sigma}_h + f\|^2_{L^2(T)} + \sum_{E\in\mathcal{E}_{\text{int}}} h_E\,\|\llbracket\boldsymbol{\sigma}_h\cdot\boldsymbol{n}\rrbracket\|^2_{L^2(E)}\right\}^{1/2} \|\nabla\psi\|_{L^2(\Omega)}.
\end{aligned}
\tag{36}
$$

Yhtälön (32) ensimmäinen termi on nyt saatu rajoitettua indikaattorilla η_T. Ryhdytään tarkastelemaan toista termiä. Otetaan funktiolle $q \in H_0^1(\Omega)/\mathbb{R}$ interpolantti $I_h q$, joka on jatkuva ja toteuttaa ehdon (33). Variaatiomuodon määritelmästä saadaan

$$
\sum_{T\in\mathcal{T}_h} (\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, \operatorname{curl} I_h q)_T = a(0, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h; 0, \operatorname{curl} I_h q) = 0 \tag{37}
$$

ja tämän avulla

$$
\begin{aligned}
&(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, \operatorname{curl} q)_\Omega = \sum_{T\in\mathcal{T}_h} (\nabla u - \boldsymbol{\sigma}_h, \operatorname{curl}(q - I_h q))_T \\
&= \sum_{T\in\mathcal{T}_h} \left[(\nabla u - \nabla u_h, \operatorname{curl}(q - I_h q))_T + (\nabla u_h - \boldsymbol{\sigma}_h, \operatorname{curl}(q - I_h q))_T\right] \\
&= R_1 + R_2.
\end{aligned}
\tag{38}
$$

Osittaisintegroimalla ja käyttämällä tulosta $\nabla\times\nabla v = 0$ saadaan

$$
\begin{aligned}
R_1 &= \sum_{T\in\mathcal{T}_h} \left[-(\nabla\times\nabla(u-u_h), q - I_h q)_T + (\nabla(u-u_h)\cdot\boldsymbol{t}, q - I_h q)_{\partial T}\right] \\
&= \sum_{E\in\mathcal{E}_{\text{int}}} (\llbracket\nabla(u-u_h)\cdot\boldsymbol{t}\rrbracket, q - I_h q)_E + \sum_{E\in\mathcal{E}_{\partial\Omega}} (\nabla(u-u_h)\cdot\boldsymbol{t}, q - I_h q)_E \\
&= \sum_{E\in\mathcal{E}_{\text{int}}} (\llbracket\nabla u_h\cdot\boldsymbol{t}\rrbracket, q - I_h q)_E + \sum_{E\in\mathcal{E}_{\partial\Omega}} (\nabla(u-u_h)\cdot\boldsymbol{t}, q - I_h q)_E = S_1 + S_2.
\end{aligned}
\tag{39}
$$

Schwarzin epäyhtälö summille ja interpolaatiotulos (33) antaa

$$
\begin{aligned}
S_1 &\le \left(\sum_{E\in\mathcal{E}_{\text{int}}} h_E\,\|\llbracket\nabla u_h\cdot\boldsymbol{t}\rrbracket\|^2_{L^2(E)}\right)^{1/2} \left(\sum_{E\in\mathcal{E}_{\text{int}}} \frac{1}{h_E}\|q - I_h q\|^2_{L^2(E)}\right)^{1/2} \\
&\le C\left(\sum_{E\in\mathcal{E}_{\text{int}}} \frac{1}{h_E}\|\llbracket u - u_h\rrbracket\|^2_{L^2(E)}\right)^{1/2} \|\nabla q\|_{L^2(\Omega)}.
\end{aligned}
\tag{40}
$$

Edellä on käytetty myös tulosta

$$\|[\![\nabla u \cdot t]\!]\|_{L^2(E)} \leq C \frac{1}{h_E} \|[\![u]\!]\|_{L^2(E)}. \tag{41}$$

Yhtälön (40) perusteella indikaattori η_T rajoittaa S_1:stä. Samaan tapaan saadaan

$$S_2 \leq C \left(\sum_{E \in \mathcal{E}_{\partial\Omega}} \frac{1}{h_E} \|u_h - g\|^2_{L^2(E)} \right) \|\nabla q\|_{L^2(\Omega)}. \tag{42}$$

Yhdistämällä (40) ja (42) nähdään että indikaattori η_T rajoittaa termiä R_1. Edelleen Schwarzin epäyhtälö summille antaa

$$\begin{aligned} R_2 &\leq C \left(\sum_{T \in \mathcal{T}_h} \|\nabla u_h - \boldsymbol{\sigma}_h\|^2_{L^2(T)} \right)^{1/2} \left(\sum_{T \in \mathcal{T}_h} \|\mathrm{curl}(q - I_h q)\|^2_{L^2(T)} \right)^{1/2} \\ &\leq C \left(\sum_{T \in \mathcal{T}_h} \|\nabla u_h - \boldsymbol{\sigma}_h\|^2_{L^2(T)} \right)^{1/2} \|\nabla q\|^2_{L^2(\Omega)}. \end{aligned} \tag{43}$$

Nyt on saatu

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|^2_{L^2(\Omega)} \leq C \sum_{T \in \mathcal{T}_h} \eta_T^2. \tag{44}$$

Vielä pitää rajoittaa yhtälön (31) toinen termi. Yhtälön (44) ja η_T:n määritelmän mukaan

$$\begin{aligned} \sum_{T \in \mathcal{T}_h} \|\nabla(u - u_h)\|^2_{L^2(T)} &= \sum_{T \in \mathcal{T}_h} \|\boldsymbol{\sigma} - \nabla u_h\|^2_{L^2(T)} \\ &\leq \sum_{T \in \mathcal{T}_h} \left[\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|^2_{L^2(T)} + \|\boldsymbol{\sigma}_h - \nabla u_h\|^2_{L^2(T)} \right] \leq C \sum_{T \in \mathcal{T}_h} \eta_T^2. \end{aligned} \tag{45}$$

Yhdistämällä yhtälöt (31), (44) ja (45) saadaan väite. □

5 NUMEERISET TULOKSET

Tässä kappaleessa tarkastelemme esitetyn menetelmän numeerista käyttäytymistä. Aluksi toteamme optimaalisen konvergenssin vapausasteiden lukumäärän suhteen ja tämän jälkeen testaamme a posteriori virhearvioon perustuvaa adaptiivista verkon tihentämistä.

Mallitehtävä on sekamuodossa esitetty Poissonin tehtävä yksikköneliössä

$$\begin{aligned} \nabla \cdot \boldsymbol{\sigma} &= -\sin(\pi x)\sin(\pi y) && (x,y) \in (0,1) \times (0,1), \\ \nabla u &= \boldsymbol{\sigma} && (x,y) \in (0,1) \times (0,1), \\ u &= \sin(2\pi x)/10 && \Gamma = \{(x,y) \mid y = 1,\ x \in (0,1)\}, \\ u &= 0 && \partial\Omega \setminus \Gamma. \end{aligned} \tag{46}$$

Tehtävän tarkka ratkaisu voidaan laskea Fourier-sarjoilla ja se on

$$u(x,y) = \frac{1}{2\pi^2} \sin(\pi x)\sin(\pi y) + \frac{1}{10\sinh(2\pi)} \sin(2\pi x)\sinh(2\pi y)\ . \tag{47}$$

Tehtävän elementtiratkaisuja on esitetty kuvassa 1. Kaikkissa laskuissa käytetään stabiilisuusparametrin arvoa $\gamma = 1$. Kuvasta näkyy ratkaisun epäjatkuvuus erityisesti lineaaristen

Taulukko 1: A posteriori virhearvion tarkkuus. Taulukossa on suureen $\frac{\left(\sum_{T\in\mathcal{T}_h}\eta_T^2\right)^{1/2}}{|||u-u_h,\sigma-\sigma_h|||}$ arvoja adaptiivisessa tihennyksessä sekä lineaarisille että toisen asteen elementeille.

polynomiaste	1. verkko	1. tihennys	2. tihennys
1	1.0	1.3	1.6
2	1.3	1.9	2.3

elementtien tapauksessa. Samoin nähdään että jo toisen asteen elementeillä epäjatkuvuutta on huomattavasti vähemmän, johtuen tarkan ratkaisun sileydestä.

Kuvassa 2 esitetään menetelmän konvergenssi $|||\cdot|||$-normissa vapausasteiden lukumäärän funktiona lineaarisille ja toisen asteen elementeille. Tasaisessa tihennyksessä pitää olla [9] $|||u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||| \sim N^{-1/2}$ lineaarisille ja $|||u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||| \sim N^{-1}$ toisen asteen elementeille, kun N on vapausasteiden lukumäärä. Kuvasta 2 nähdään että menetelmä toteuttaa a priori tuloksen hyvin tarkasti.

Lisäksi kuvassa 2 esitetään virhe myös adaptiivisen verkon tihennyksen tapauksessa. Adaptiivisessa tihennyksessä tihennetään 30% elementeistä ja tihennettävät elemenetit valitaan a posteriori virhearvion perusteella. Kuvasta 2 nähdään että adaptiivisella tihennyksellä saavutetaan tasaista tihennystä parempi konvergenssi. Toisin sanoen ratkaisun adaptiivisella tihennyksellä saavutetaan haluttu tarkkuus pienemmällä vapausasteiden määrällä.

Edellä adaptiiviset tihennykset on tehty ilman toleranssia, mutta käytännössä kannattaa käyttää esimerkiksi seuraavaa tihennysalgoritmia. Pyritään saavuttamaan

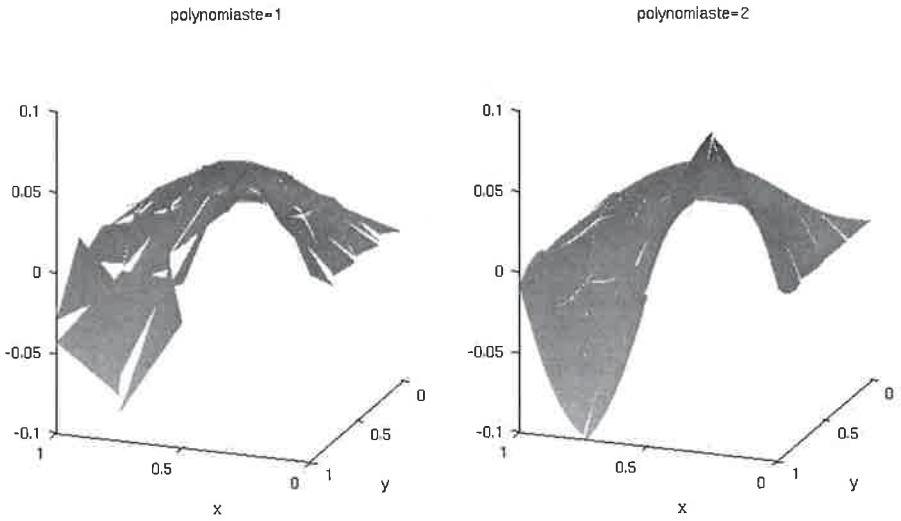
$$C|||u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||| \le \left(\sum_{T\in\mathcal{T}_h}\eta_T^2\right)^{1/2} \le TOL\ ,$$

mikä saavutetaan esimerkiksi tihentämällä elementit, joille

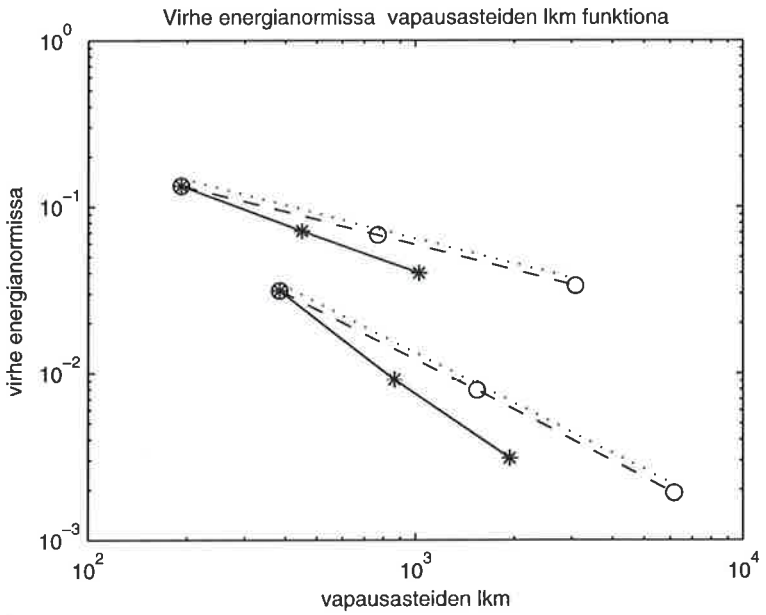
$$\eta_T \ge \frac{TOL^2}{\mathrm{lkm}(elementit)}\ .$$

Edellä vakio C on tuntematon, joten varmoja rajoja ei saavuteta, mutta käytännössä vakion C arvo on yleensä luokkaa 1. Kun elementeistä tihennetään aina 30%, saadaan vakiolle C taulukon 1 arviot.

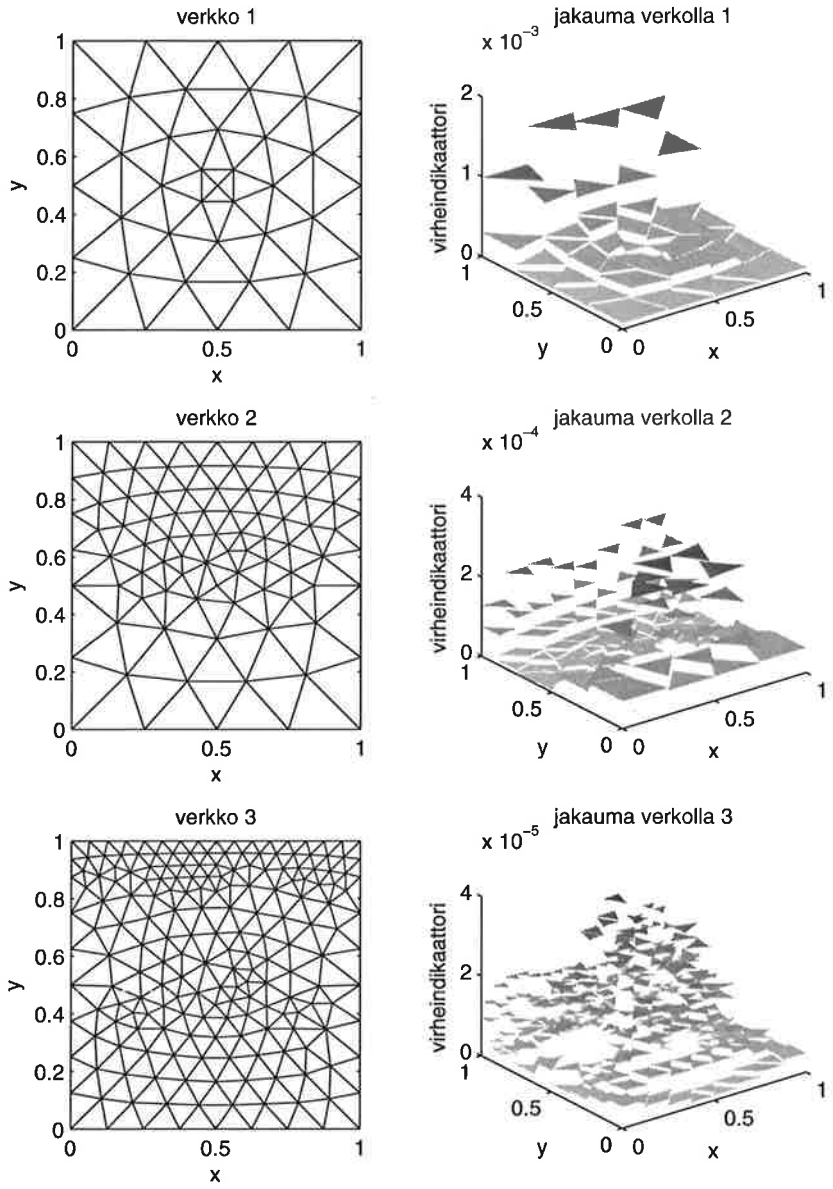
Kuvassa 3 näytetään a posteriori virhearvion jakauma ja adaptiivisella tihennyksellä syntyneitä verkkoja lineaarisilla elementeillä laskettaessa. Kuvista nähdään että virheindikaattori huomaa reunalla Γ ja alueen keskellä sijaitsevat suuremmat ratkaisun muutokset ja tihentää näissä alueissa.



Kuva 1: Esimerkkitehtävän ratkaisuja kuvan 3 ensimmäisellä verkolla.



Kuva 2: Esimerkkitehtävän virheen konvergenssi energianormissa vapausasteiden lukumäärän funktiona. Ylempi sarja on laskettu lineaarisilla elementeillä ja alempi sarja toisen asteen elementeillä. Pallot merkitsevät tasaista tihennystä ja tähdet a posteriori virheen perusteella tehtyä adaptiivista tihennystä. Pisteviivat ovat referenssisuorat $N^{-1/2}$ ja N^{-1}, jotka vastaavat a priori virhearvion antamia oletuskonvergensseja Ch lineaarisille ja Ch^2 toisen asteen elementeille.



Kuva 3: Adaptiivisen tihennyksen kulkua. Ylhäältä alas on verkko jolta aloitetaan ja kaksi adaptiivisen tihennyksen tuottamaa verkkoa. Vasemmalla on käytetyt verkot ja oikealla on a posteriori virheindikaattorin antama virheen jakauma kyseisellä verkolla. Adaptiivisessa tihennyksessä on tihennetty 30% elementeistä.

VIITTEET

[1] J.A. Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* **36**, 9–15, (1970/71)

[2] R. Stenberg. Mortaring by a method of J.A. Nitsche. Computational Mechanics; New Trends and Applications S. Idelsohn, E. Oñate and E. Dvorkin (Eds.) *@CIMNE, Barcelona, Spain, 1998*

[3] R. Becker, P. Hansbo, and R. Stenberg. A finite element methods for domain decomposition with non-matching grids. *Mathematical Modelling and Numerical Analysis* **37**, 209–225, (2003)

[4] D.N. Arnold. An Interior Penalty Finite Element Method with Discontinuous Elements. *SIAM Journal on Numerical Analysis* **19**, 742–760, (1982)

[5] E. Dari, R. Duran, C. Padra, and V. Vampa. A Posteriori Error Estimators For Nonconforming Finite Element Methods. *Mathematical Modelling And Numerical Analysis* **30** (1996) 385–400

[6] C. Lovadina, and R. Stenberg. Energy Norm A Posteriori Error Estimates For Mixed Finite Element Methods. *To apper in Math. Comp.*

[7] V. Girault, and P. Raviart. Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms. *Springer-Verlag* 1986

[8] F. Brezzi, B. Cockburn, L.D. Marini, and E. Süli. Stabilization Mechanism in Discontinuous Galerkin Finite Element Methods *Comput. Methods Appl. Mech. Engrg.* **195** (2006) 3293–3310

[9] P.G. Ciarlet. The Finite Element Methods for Elliptic Problems *North-Holland* 1987

A POSTERIORI ESTIMATES FOR COMPUTATIONAL ERROR CONTROL IN TERMS OF LINEAR FUNCTIONALS FOR LINEAR ELASTICITY PROBLEMS

S. KOROTOV

Institute of Mathematics
Helsinki University of Technology
P. O. Box 1100, FI–02015 TKK, Finland
e-mail: sergey.korotov@hut.fi

ABSTRACT

This paper is devoted to construction of a posteriori error estimators aimed at control of computational errors for problems in the linear elasticity theory. The error control is performed in terms of linear (continuous) functionals, which can be, in particular, designed to verify the difference error between the exact solution and its approximation in local subdomains of special interest. The approach employed has been earlier analysed in the author's works [17, 19] for a class of linear elliptic problems. It is based on usage of an auxiliary (so-called adjoint) problem. In the framework of this approach the original and adjoint problems are solved on noncoinciding meshes and averaging of gradients is used to evaluate the term in the estimator that cannot be evaluated directly. In the present paper, we consider the case of an elliptic system of partial differential equations arising in the theory of linear elasticity, where the averaging procedures are now applied to the field of strains (or stresses). Series of numerical tests performed shows that the error estimator proposed is asymptotically exact, and also demonstrates that in many cases sufficiently accurate evaluation of the error in terms of linear functionals can be obtained even if the number of nodes in the adjoint mesh is essentially less than in the primal one.

1 INTRODUCTION

Linear elasticity problems, arising in many technical applications, are among the most interesting engineering problems. Various numerical methods developed for finding approximate solutions to these problems are well-known and are widely used in the computational practice, see, e.g., [8, 9, 22, 25, 27]. Nowadays the finite element method seems to be the most popular among all other numerical techniques used for this type of elliptic problems.

However, calculations always require a reliable control of accuracy of approximations obtained. Development of analytical and computational tools for such a control is the main purpose of a posteriori error estimation analysis. Various approaches for deriving estimates for elliptic type boundary value problems for errors measured in the energy norm have been suggested by many authors (see, for example, works [1, 2, 3, 4, 12, 14, 16, 23, 26] and references therein), also useful error indicators for adaptive mesh refinement have been constructed by special post-processing procedures (see, e.g., [26, 27]).

In the recent years, a new line in a posteriori error estimation has been actively devel-

oping. It is based on the concept of the error control in terms of special problem–oriented criteria (see, e.g., [1, 4, 14, 15, 21, 24]) instead of (or in addition to) the error control in global energy norms. Error estimates of such a type are strongly motivated by the needs of real–life problems, in which analysts are often interested not in the value of the overall error, but mainly in various local errors over certain "subdomains of special interest" or relative to some interesting characteristics (e.g., to the so-called J-integral in the fracture mechanics, see [24] and references therein). A possible way for estimation of such errors is to introduce a linear functional ℓ associated with "problem-oriented criterion" and to obtain an estimate for the value $\ell(u-\bar{u})$, where u is the exact solution and $\bar{u}$ is the approximate one. Known methods (see, e.g., [1, 5, 10, 11, 21]) find estimates of $\ell(u-\bar{u})$ by employing an auxiliary (adjoint) problem, whose right-hand side is formed by the functional ℓ.

For linear elasticity problems, the basic example of the error control via linear functionals is an estimation of the following integral

$$\int_S \mathbf{\Phi}(s) \cdot (\mathbf{u}(s) - \bar{\mathbf{u}}(s))\, ds,$$

where S is a certain subdomain (or a line) in the problem domain Ω and $\bar{\mathbf{u}}$ is an approximation for the displacement field $\mathbf{u}$. More special (and problem-oriented) examples of error control, via another functionals, can be found in the fracture mechanics, see, e.g., [24].

In the present paper, we use a new way of estimating the computational errors via linear functionals, as proposed in our earlier work [19] (see also [11, 17, 18]) for the linear scalar elliptic problem, and apply it to the case of elliptic system in the linear elasticity theory. The approach is essentially based on two principles: (a) original and adjoint problems are solved on non-coinciding meshes, and (b) term presenting the product of errors arising in the primal and adjoint problems is estimated with help of one of various gradient recovery techniques widely used in various applied problems (see [6, 7, 13, 26, 27]). This differs our approach from the others, where the finite element approximations of both, the primal and adjoint problems, are computed in the same finite-dimensional subspaces. We verify the effectivity and demonstrate the asymptotic convergence of the constructed estimator in a series of numerical tests.

2 PROBLEM FORMULATION AND ERROR DECOMPOSITION

Consider an elastic body occupying a bounded domain $\Omega \subset \mathbf{R}^d, d = 2, 3$, with a Lipschitz continuous boundary $\partial\Omega = \overline{\Gamma}_1 \cup \overline{\Gamma}_2$, where $\Gamma_1 \neq \emptyset$ and Γ_2 are disjoint and relatively open parts of $\partial\Omega$. The problem consists of finding the vector–valued function $\mathbf{u}$ (displacement) and the tensor–valued function σ (stress) such that

$$-\mathrm{div}\ \sigma = \mathbf{f} \qquad \text{in} \quad \Omega, \tag{1}$$

$$\sigma = \mathbf{L}\,\epsilon, \quad \epsilon(\mathbf{u}) = \frac{1}{2}\ \nabla\mathbf{u} + (\nabla\mathbf{u})^T\ , \tag{2}$$

$$\mathbf{u} = \mathbf{u}^0 \qquad \text{on} \quad \Gamma_1, \tag{3}$$

$$\sigma\,\mathbf{n} = \mathbf{g} \qquad \text{on} \quad \Gamma_2, \tag{4}$$

where the symbol $\mathbf{n}$ stands for the unit outward normal to $\partial\Omega$, $\mathbf{f}$ and $\mathbf{g}$ denote the given volume and surface loads, respectively, $\mathbf{u}^0$ is the prescribed displacement on Γ_1, and the tensor $\mathbf{L}$ will be defined later.

In what follows, we denote the scalar product of vectors $\mathbf{a}, \mathbf{b} \in \mathbf{R}^d$ by the dot, $\mathbf{a} \cdot \mathbf{b} := \sum_{i=1}^d a_i b_i$. Similarly, the scalar product of symmetric tensors $\tau, \kappa \in \mathbf{M}_{sym}^{d\times d}$ is denoted by $\tau : \kappa := \sum_{i,j=1}^d \tau_{ij}\kappa_{ij}$. The norm of a vector $\mathbf{a}$ is denoted by $\|\mathbf{a}\|$, $\|\mathbf{a}\| := (\mathbf{a}\cdot\mathbf{a})^{1/2}$, the norm of a tensor τ is denoted by $\|\tau\|$ and is equal to $(\tau : \tau)^{1/2}$.

In the above, $\mathbf{L} = \mathbf{L}(x) = (L_{ijkl}(x))_{i,j,k,l=1}^d$ stands for the fourth order tensor of elastic coefficients, which satisfy the following symmetry condition

$$L_{jikl} = L_{ijkl} = L_{klij} \ , \qquad i,j,k,l = 1,\dots,d \ , \tag{5}$$

and the condition that there exists a positive constant C_1 such that

$$\mathbf{L}(x)\,\tau : \tau \geq C_1\|\tau\|^2 \quad \forall \tau \in \mathbf{M}_{sym}^{d\times d} \tag{6}$$

almost everywhere in Ω. We also assume that

$$L_{ijkl} \in L_\infty(\Omega),\ \mathbf{f} \in (L_2(\Omega))^d,\ \mathbf{g} \in (L_2(\Gamma_2))^d,\ \mathbf{u}^0 \in (H^1(\Omega))^d. \tag{7}$$

By definition,

$$\mathbf{L}\,\tau : \kappa := \sum_{i,j,k,l=1}^d L_{ijkl}\tau_{ij}\kappa_{kl}. \tag{8}$$

The weak formulation of problem (1)–(4) is given below. It will be called the *primal problem* and denoted as $(\mathcal{PP})$ in what follows.

Primal Problem $(\mathcal{PP})$: Find $\mathbf{u} \in \mathbf{V}^0 + \mathbf{u}^0$ such that

$$a(\mathbf{u},\mathbf{w}) \ = \ F(\mathbf{w}) \qquad \forall \mathbf{w} \in \mathbf{V}^0, \tag{9}$$

where

$$\mathbf{V}^0 = \{\mathbf{v} \in (H^1(\Omega))^d \mid \mathbf{v} = 0 \quad \text{on} \quad \Gamma_1\} \ , \tag{10}$$

$$a(\mathbf{v},\mathbf{w}) := \int_\Omega \mathbf{L}\,\epsilon(\mathbf{v}) : \epsilon(\mathbf{w})\,dx, \tag{11}$$

and

$$F(\mathbf{w}) := \int_\Omega \mathbf{f}\cdot\mathbf{w}\,dx + \int_{\Gamma_2} \mathbf{g}\cdot\mathbf{w}\,ds. \tag{12}$$

Since the above defined bilinear form $a(\cdot,\cdot)$ is continuous and coercive (due to the Korn's inequality, see, e.g., [20]), the primal problem (9)–(12) is uniquely solvable.

Let $\bar{\mathbf{u}} \in \mathbf{V}^0 + \mathbf{u}^0$ be an approximation of $\mathbf{u}$ (e.g., obtained by some numerical technique). Our task is to derive an estimate for the following quantity

$$\ell(\mathbf{u} - \bar{\mathbf{u}}), \tag{13}$$

where $\ell : \mathbf{V}^0 \to \mathbf{R}$ is a linear continuous functional selected for the error control. A common way to obtain an estimate for $\ell(\mathbf{u} - \bar{\mathbf{u}})$ is to introduce an auxiliary problem (often called the *adjoint problem* and to which we refer as $(\mathcal{AP})$) in the following way.

Adjoint Problem ($\mathcal{AP}$): Find $\mathbf{v} \in \mathbf{V}^0$ such that

$$a(\mathbf{v}, \mathbf{w}) \;=\; \ell(\mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V}^0. \tag{14}$$

Under the above assumptions (5)–(7), the adjoint problem is uniquely solvable. However, in general, the exact (weak) solution of ($\mathcal{AP}$) cannot be found in analytical form and only its approximation $\bar{\mathbf{v}} \in \mathbf{V}^0$ is usually available.

Lemma 1: The following error decomposition holds

$$\ell(\mathbf{u} - \bar{\mathbf{u}}) \;=\; E_0(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \;+\; E_1(\bar{\mathbf{u}}, \bar{\mathbf{v}}), \tag{15}$$

where

$$E_0(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \; F(\bar{\mathbf{v}}) - \int_\Omega \mathbf{L}\,\epsilon(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{v}})\,dx, \tag{16}$$

and

$$E_1(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \int_\Omega \mathbf{L}\,\epsilon(\mathbf{u} - \bar{\mathbf{u}}) : \epsilon(\mathbf{v} - \bar{\mathbf{v}})\,dx. \tag{17}$$

P r o o f : Decomposition (15) immediately follows from the following obvious integral identities

$$\ell(\mathbf{u} - \bar{\mathbf{u}}) \;=\; \int_\Omega \mathbf{L}\,\epsilon(\mathbf{v}) : \epsilon(\mathbf{u} - \bar{\mathbf{u}})\,dx =$$

$$= \int_\Omega \mathbf{L}\,\epsilon(\mathbf{v} - \bar{\mathbf{v}}) : \epsilon(\mathbf{u} - \bar{\mathbf{u}})\,dx + \int_\Omega \mathbf{L}\,\epsilon(\bar{\mathbf{v}}) : \epsilon(\mathbf{u} - \bar{\mathbf{u}})\,dx,$$

and

$$\int_\Omega \mathbf{L}\,\epsilon(\bar{\mathbf{v}}) : \epsilon(\mathbf{u} - \bar{\mathbf{u}})\,dx \;=\; \int_\Omega \mathbf{L}\,\epsilon(\mathbf{u} - \bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{v}})\,dx =$$

$$= F(\bar{\mathbf{v}}) - \int_\Omega \mathbf{L}\,\epsilon(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{v}})\,dx.$$

Remark 1: We note that the term E_0 is directly computable once $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ are known, whereas the term E_1 should be still estimated.

Remark 2: The terms $E_0(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ and $E_1(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ can also be represented in terms of stresses as follows:

$$E_0(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \; F(\bar{\mathbf{v}}) - \int_\Omega \mathbf{L}^{-1}\,\sigma(\bar{\mathbf{u}}) : \sigma(\bar{\mathbf{v}})\,dx, \;\; E_1(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \int_\Omega \mathbf{L}^{-1}\,\sigma(\mathbf{u} - \bar{\mathbf{u}}) : \sigma(\mathbf{v} - \bar{\mathbf{v}})\,dx. \tag{18}$$

3 CONSTRUCTION OF ERROR ESTIMATOR

Let $\mathbf{V}^h$ and $\mathbf{V}^\tau$ be two finite-dimensional subspaces of $\mathbf{V}^0$, not necessarily coinciding, that are based on finite element meshes $\mathcal{T}_h$ and $\mathcal{T}_\tau$. We shall use $\mathbf{V}^h$ and $\mathbf{V}^\tau$ for construction of the finite element approximations for problems ($\mathcal{PP}$) and ($\mathcal{AP}$), respectively, i.e., pose the following two problems:

Problem $(\mathcal{PP}^h)$: Find $\mathbf{u}^h \in \mathbf{V}^h + \mathbf{u}^0$ such that

$$a(\mathbf{u}^h, \mathbf{w}^h) = F(\mathbf{w}^h) \qquad \forall \mathbf{w}^h \in \mathbf{V}^h, \tag{19}$$

Problem $(\mathcal{AP}^\tau)$: Find $\mathbf{v}^\tau \in \mathbf{V}^\tau$ such that

$$a(\mathbf{v}^\tau, \mathbf{w}^\tau) = \ell(\mathbf{w}^\tau) \quad \forall \mathbf{w}^\tau \in \mathbf{V}^\tau. \tag{20}$$

Both above problems are, obviously, uniquely solvable.

Further, setting $\bar{\mathbf{u}} = \mathbf{u}^h$ and $\bar{\mathbf{v}} = \mathbf{v}^\tau$ in the error decomposition (15)–(17), we obtain

$$\ell(\mathbf{u} - \mathbf{u}^h) = E_0(\mathbf{u}^h, \mathbf{v}^\tau) + E_1(\mathbf{u}^h, \mathbf{v}^\tau). \tag{21}$$

It is well-known that the finite element solution of the linear elasticity problems possesses certain superconvergence properties (see, e.g., [13]), which makes it possible to prove that its averaged gradient often presents a good image of the true one. This fact made a posteriori error indicators based on the gradient averaging techniques very popular. It is natural to exploit this property in the estimation of the term E_1, which contains unknown gradients of the solutions of primal and adjoint problems.

Thus, let

$$\mathbf{G}^h, \mathbf{G}^\tau : (L_\infty(\Omega))^d \to (H^1(\Omega))^d$$

be some gradient averaging operators related to the meshes $\mathcal{T}_h$ and $\mathcal{T}_\tau$, respectively.

In the simplest case of linear finite elements, both gradient averaging operators can be defined as a mapping of a piecewise constant gradient ($\nabla \mathbf{u}^h$ or $\nabla \mathbf{v}^\tau$) into a vector-valued piecewise linear function by setting each its nodal value as the mean (or weighted mean) value of the gradient values on all elements incident with the corresponding nodal point (cf. [13]).

By means of $\mathbf{G}^h$ we can define an averaging operator for the strains (still denoted by the symbol $\mathbf{G}^h$) as follows

$$\mathbf{G}^h(\epsilon(\mathbf{u}^h)) := \frac{1}{2} \left[\mathbf{G}^h(\nabla \mathbf{u}^h) + (\mathbf{G}^h(\nabla \mathbf{u}^h))^T \right], \tag{22}$$

and, similarly,

$$\mathbf{G}^\tau(\epsilon(\mathbf{v}^\tau)) := \frac{1}{2} \left[\mathbf{G}^\tau(\nabla \mathbf{v}^\tau) + (\mathbf{G}^\tau(\nabla \mathbf{v}^\tau))^T \right]. \tag{23}$$

Then, the term $E_1(\mathbf{u}^h, \mathbf{v}^\tau)$ can be replaced by the explicitly computable term

$$\widetilde{E}_1(\mathbf{u}^h, \mathbf{v}^\tau) = \int_\Omega \mathbf{L} \left(\mathbf{G}^h(\epsilon(\mathbf{u}^h)) - \epsilon(\mathbf{u}^h) \right) : \left(\mathbf{G}^\tau(\epsilon(\mathbf{v}^\tau)) - \epsilon(\mathbf{v}^\tau) \right) dx, \tag{24}$$

that leads to the error estimator in the form

$$\widetilde{E}(\mathbf{u}^h, \mathbf{v}^\tau) := E_0(\mathbf{u}^h, \mathbf{v}^\tau) + \widetilde{E}_1(\mathbf{u}^h, \mathbf{v}^\tau), \tag{25}$$

which is directly computable once the approximations $\mathbf{u}^h$ and $\mathbf{v}^\tau$ are defined.

Remark 3: We note that $E_0(\mathbf{u}^h, \mathbf{v}^\tau)$ asymptotically contains the major part of $\ell(\mathbf{u} - \mathbf{u}^h)$ as $\mathbf{v}^\tau \to \mathbf{v}$. For the special case $\mathbf{V}^h \equiv \mathbf{V}^\tau$ the term E_0, obviously, vanishes.

The above construction suggests the following numerical strategy, including suggestion for the relevant mesh adaptive procedure:

(a) Define $\mathbf{V}^\tau$ taking into an account the nature of the functional ℓ (e.g., by putting extra trial functions in a subdomain associated with it), and calculate $\mathbf{v}^\tau$,

(b) Define $\mathbf{V}^h$ and calculate $\mathbf{u}^h$,

(c) Calculate $E_0(\mathbf{u}^h, \mathbf{v}^\tau)$ directly and use post–processed values of $\nabla \mathbf{u}^h$ and $\nabla \mathbf{v}^\tau$ to estimate $E_1(\mathbf{u}^h, \mathbf{v}^\tau)$ replacing unknown strains (stresses) by easily computable averaged strains (stresses),

(d) The estimator $\tilde{E}$ is, in fact, an integral over Ω, i.e.,

$$\tilde{E}(\mathbf{u}^h, \mathbf{v}^\tau) := \sum_{T \in \mathcal{T}_h^{(i)}} I_T, \tag{26}$$

where each contribution I_T is a value of the integral taken over a particular element T of the current mesh $\mathcal{T}_h^{(i)}$. To construct the next primal mesh $\mathcal{T}_h^{(i+1)}$ in order to decrease the error, we propose the following adaptive procedure. First, we find the maximum among all modulus $|I_T|$'s and, secondly, mark up those elements T's which have their contributions larger than the "user-given threshold" $\theta \in [0, 1]$ times that maximum value. Refining the marked elements (and making the mesh conforming), we get the next mesh, $\mathcal{T}_h^{(i+1)}$.

Note that computations made in the item (a) can be further used for estimation of errors of an approximate solution obtained on another mesh and with different $\mathbf{f}$, $\mathbf{u}^0$, $\mathbf{g}$.

4 NUMERICAL EXPERIMENTS

In all the tests in this section, we shall deal with the plane stress problem with the following parameters: Young's modulus $E = 10^6$, Poisson's ratio $\nu = 0.3$, and volume forces $\mathbf{f} = (0, 0)$. The mesh generator and the solver of the PDE Toolbox of Matlab are used in all the calculations. We shall always estimate the error

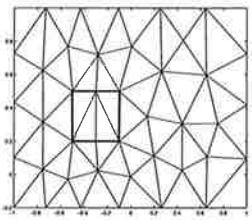
$$\ell(\mathbf{u} - \mathbf{u}^h) = \int_\omega \mathbf{\Phi} \cdot (\mathbf{u} - \mathbf{u}^h)\, dx, \tag{27}$$

where $\mathbf{\Phi} = (1, 1)^T$ in a small subdomain $\omega \subseteq \Omega$ and vanishes outside of ω. The symbol I_{eff} denotes the so-called *effectivity index*

$$I_{eff} := \frac{\tilde{E}(\mathbf{u}^h, \mathbf{v}^\tau)}{\ell(\mathbf{u} - \mathbf{u}^h)}, \tag{28}$$

which shows how good the estimator's values are.

Test 1: In the first test, the solution domain $\Omega := (-1.0, 1.0) \times (-0.2, 1.0)$ and $\omega := (-0.5, -0.1) \times (0.2, 0.5)$, see Fig. 1 (left). We assume that $\mathbf{u}^0 = (0, 0)$ on the lower part of the boundary. Further, let the load $\mathbf{g} = (0, -10^4)$ acts on the upper part of the boundary and it is assumed that $\mathbf{g} = (0, 0)$ on the remaining parts of $\partial\Omega$. The finite element solution $\mathbf{u}^h$ is calculated on primal mesh $\mathcal{T}_h$ with 44 nodes presented in Fig. 1 (left). One example of corresponding adjoint mesh is given in Fig. 1 (right). All the other adjoint meshes in the test are obtained from $\mathcal{T}_h$ by the standard red-refinement procedure used one, two, etc, times. The results of calculations are summarized in Table 1. The first and the second columns of the table contain the numbers of nodes in the primal and adjoint meshes, respectively.



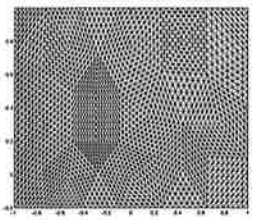


Figure 1: Primal mesh $\mathcal{T}_h$ (left) with 44 nodes (with zone of interest marked by the bold line) and an example of adjoint mesh $\mathcal{T}_\tau$ (right) with 2193 nodes for Test 1.

For the gradient averaging we used a computationally cheap procedure of mean arithmetic averaging of gradients at each node according to the number of elements in the patch surrounding this node.

Table 1: Performance of estimator $\tilde{E}$ for Test 1

Pr	Ad	$E_0 \times 10^3$	$\tilde{E}_1 \times 10^3$	$\tilde{E} \times 10^3$	$\ell(\mathbf{u}-\mathbf{u}^h) \times 10^3$	I_{eff}
44	44	-0.000000	-0.003766	-0.003766	-0.002644	1.42
44	153	-0.002413	-0.001228	-0.003641	-0.002644	1.38
44	569	-0.002679	-0.000443	-0.003122	-0.002644	1.18
44	2193	-0.002669	-0.000150	-0.002819	-0.002644	1.07
44	8609	-0.002644	-0.000048	-0.002692	-0.002644	1.02

It is worth to mention that Ω in this example has no reentrant corners, i.e., the exact solution of the problem is sufficiently smooth almost everywhere in Ω. In addition, the meshes used are "quasi-uniform". Both circumstances provide "strongly expressed" superconvergence of the averaged gradients (cf. [13]). Thus, the term E_0 dominates, $\tilde{E}_1$ tends to zero, and the estimator $\tilde{E}$ tends to the exact value of the error as we use more and more fine adjoint meshes for the estimation. The other types of gradient averaging give essentially the same results.

Test 2: In this test, we consider a more complicated problem when the solution domain has a reentrant corner. Thus, let Ω and ω be as presented in Fig. 2. The Dirihlet conditions $\mathbf{u}^0 = (0, \pm 1)$ are prescribed on the upper and the lower parts of Ω, respectively. In the remaining parts of $\partial\Omega$ we impose homogeneous Neumann conditions. The finite element solution $\mathbf{u}^h$ is calculated on the mesh $\mathcal{T}_h$ with 171 nodes, see Fig. 2 (left), and the corresponding von Mises stress distribution is given in Fig. 2 (right). The results of calculations for the estimator are presented in Table 2 (see also Fig. 3), where the first and the second columns contain the numbers of nodes in primal and adjoint meshes, respectively. Several corresponding adjoint meshes are given in Fig. 4. In this test we use for averaging purposes the weighted mean of gradients with respect to areas of relevant triangles since trianglar elements in the mesh are quite different in value in diferent parts of the solution domain and it is natural to balance the gradients' contributions according to areas of respective triangles. However, the results with another averagings, including a simple mean averaging, are only slightly different.

We note that in this example the solution has a singularity at the point $(0, 0)$, which is a vertex of the reentrant corner. Our test (cf. Test 4) demonstrates the fact that even in the presence of high gradients of the solutions, an usage of a computationally cheap adjoint mesh with a considerable smaller amount of nodes than the number of nodes in the primal

problem leads to quite effective estimates. Also the asymptotic convergence of the estimator to the true error is still observed in this test similarly to Test 1 (see Fig. 3).

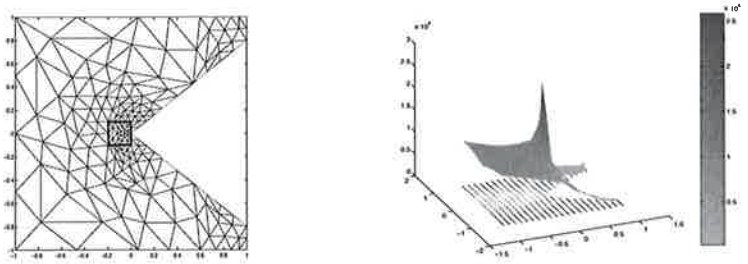


Figure 2: Primal mesh with 171 nodes and zone of interest marked by black bold line (left) and the von Mises stress distribution (right) for Test 2.

Test 3 (adaptivity): In this test, we demonstrate the effectivity of the adaptive procedure proposed in Section 3. We take the same problem as in Test 2 with mesh $\mathcal{T}_h$ having 171 nodes (see Fig. 2 (left)). Further, using a fixed adjoint mesh $\mathcal{T}_\tau$ (with 155 nodes) we perform 4 adaptive steps with $\theta = 0.5$. The results of calculations are presented in Table 3 and the respective meshes are given in Fig. 5. We clearly observe that both, true errors and the estimator's values monotonically decrease, and that the efectivity index is reasonably good even for the case when adjoint mesh is five times coarser than the primal one (155 nodes vs. 776 nodes). It is worth to emphasize that using possibility to solve the adjoint problem on fixed mesh (or at least refining adjoint meshes more slowly than the primal one), we save considerable amount of computational time and memory, which is essential especially if we perform error control for real-life problems (cf. [24] for large computations).

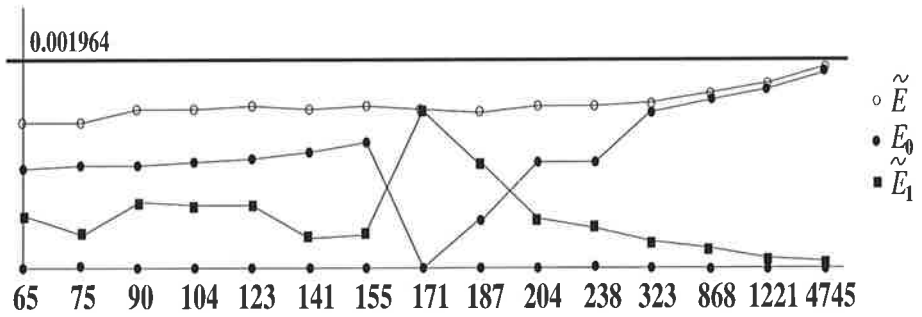


Figure 3: Graphical illustration of behaviour of the estimator and its parts for various choices of adjoint meshes for Test 2 (cf. Table 2). Mesh $\mathcal{T}_h$ has 171 nodes in all computations.

Table 2: Performance of estimator $\tilde{E}$ for Test 2

Pr	Ad	E_0	$\tilde{E}_1$	$\tilde{E}$	$\ell(\mathbf{u}-\mathbf{u}^h)$	I_{eff}
171	65	0.000937	0.000450	0.001387	0.001964	0.71
171	75	0.001053	0.000303	0.001356	0.001964	0.69
171	90	0.000914	0.000646	0.001560	0.001964	0.79
171	104	0.001048	0.000525	0.001573	0.001964	0.80
171	123	0.001057	0.000549	0.001606	0.001964	0.82
171	141	0.001195	0.000318	0.001513	0.001964	0.77
171	155	0.001207	0.000373	0.001580	0.001964	0.80
171	171	0.000000	0.001511	0.001511	0.001964	0.77
171	187	0.000478	0.001053	0.001531	0.001964	0.78
171	204	0.001072	0.000531	0.001603	0.001964	0.82
171	238	0.001093	0.000518	0.001611	0.001964	0.82
171	323	0.001496	0.000241	0.001737	0.001964	0.89
171	868	0.001634	0.000171	0.001805	0.001964	0.92
171	1221	0.001781	0.000086	0.001867	0.001964	0.95
171	4745	0.001893	0.000044	0.001937	0.001964	0.99

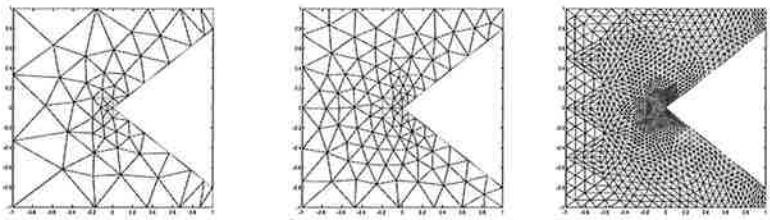


Figure 4: Examples of adjoint meshes with 65, 123, and 1221 nodes for Test 2.

Table 3: The results of adaptivity in Test 3

Pr	Ad	E_0	$\tilde{E}_1$	$\tilde{E}$	$\ell(\mathbf{u}-\mathbf{u}^h)$	I_{eff}
$\mathcal{T}_h^{(0)}(171)$	155	0.001207	0.000373	0.001580	0.001964	0.80
$\mathcal{T}_h^{(1)}(187)$	155	0.000975	0.000373	0.001348	0.001700	0.79
$\mathcal{T}_h^{(2)}(250)$	155	0.000661	0.000288	0.000949	0.001234	0.77
$\mathcal{T}_h^{(3)}(425)$	155	0.000416	0.000233	0.000649	0.000794	0.82
$\mathcal{T}_h^{(4)}(776)$	155	0.000239	0.000165	0.000404	0.000569	0.71

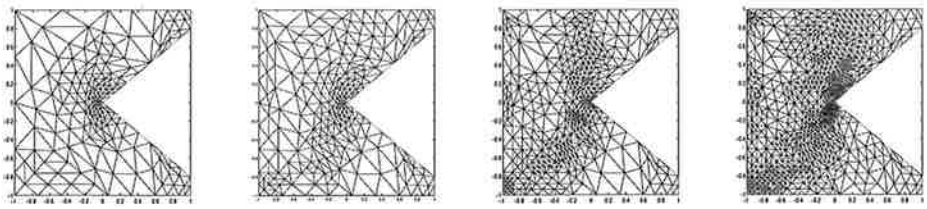


Figure 5: Mesh adaptation in Test 3.

Test 4: In this series of numerical experiments, the solution domains and the zone of interest $\omega = (-0.2, 0) \times (-0.1, 0.1)$ are presented in Fig. 6. We note that the reentrant corner at $(0, 0)$ is getting sharper which makes the solution singularity stronger. The boundary conditions are taken as in Test 2 in all the cases. The finite element solutions $\mathbf{u}^h$ are calculated on the meshes $\mathcal{T}_h$ with 167, 171, and 169 nodes, respectively (see Fig. 6), the employed adjoint meshes are given in Fig. 7. The results of calculations for the estimator are presented in Table 4. We observe that the estimator performs well in all the tests even if the number of nodes in adjoint problem is approximately two times less than the number of nodes in the corresponding primal problem.

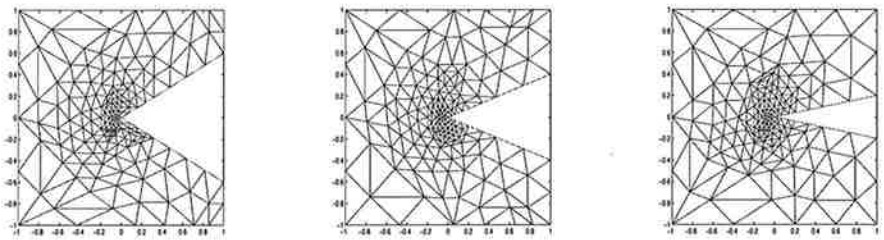


Figure 6: Primal meshes with 167, 171, and 169 nodes for Test 4.

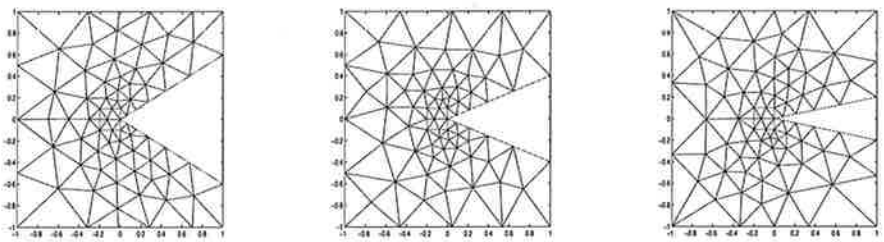


Figure 7: Adjoint meshes with 80, 78, and 82 nodes for Test 4.

Table 4: Performance of estimator $\tilde{E}$ for Test 4

Pr	Ad	E_0	$\tilde{E}_1$	$\tilde{E}$	$\ell(\mathbf{u}-\mathbf{u}^h)$	I_{eff}
167	80	0.000900	0.000441	0.001341	0.001725	0.78
171	78	0.000536	0.000491	0.001027	0.001356	0.76
169	82	0.000913	0.000228	0.001141	0.001405	0.81

ACKNOWLEDGEMENT

The author was supported by the Academy Research Fellowship no. 208628 and project no. 211512 from the Academy of Finland. The author also thanks Mr. Antti Hannukainen for several useful suggestions on the style of the paper.

REFERENCES

[1] M. Ainsworth and J. T. Oden, *A Posteriori Error Estimation in Finite Element Analysis.* John Wiley & Sons, Inc., 2000.

[2] I. Babuška and W. C. Rheinbold, Error estimates for adaptive finite element computations. *SIAM J. Numer. Anal.* (1978) **15**, 736–754.

[3] I. Babuška and T. Strouboulis, *The Finite Element Method and its Reliability.* Oxford University Press Inc., New York, 2001.

[4] W. Bangerth and R. Rannacher, *Adaptive Finite Element Methods for Differential Equations.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.

[5] R. Becker and R. Rannacher, A feed-back approach to error control in finite element methods: Basic approach and examples. *East-West J. Numer. Math.* (1996) **4**, 237–264.

[6] J. Brandts and M. Křížek, Gradient superconvergence on uniform simplicial partitions of polytopes. *IMA J. Numer. Anal.* (2003) **23**, 489–505.

[7] J. Brandts and M. Křížek, Superconvergence of tetrahedral quadratic finite elements. *J. Comput. Math.* (2005) **23**, 27–36.

[8] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods. Second edition. Texts in Applied Mathematics, 15.* Springer-Verlag, New York, 2002.

[9] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam, 1978.

[10] A. Hannukainen, A posteriori error control in terms of linear functionals for elliptic problems: estimation and mesh adaptivity. In "Proc. of Finnish Mechanical Days (SMP-2006), Lappeenranta, Finland ", 1–12 (to appear).

[11] A. Hannukainen and S. Korotov, Techniques for a posteriori error estimation in terms of linear functionals for linear elliptic problems. *Far East J. Appl. Math.* (2005) **21**, 289–304.

[12] A. Hannukainen and S. Korotov, Computational technologies for reliable control of global and local errors for linear elliptic type boundary value problems. Preprint A494 (2006), Helsinki University of Technology (submitted to J. Numer. Anal. Comp. Math.).

[13] I. Hlaváček and M. Křížek, On a superconvergent finite element scheme for elliptic systems. Parts I, II, III. *Apl. Mat.* (1987) **32**, 131–154, 200–213, 276–289.

[14] C. Johnson and P. Hansbo, Adaptive finite elements in computational mechanics. *Comput. Methods Appl. Mech. Engrg.* (1992) **101**, 143–181.

[15] S. Korotov, New technologies for control of local errors in engineering computations by FEM, *Rakenteiden Mekaniikka (Journal of Structure Mechanics)* (2005) **38**, 117–120.

[16] S. Korotov, A posteriori error estimation for linear elliptic problems with mixed boundary conditions. Preprint A495 (2006), Helsinki University of Technology (submitted to BIT).

[17] S. Korotov, A posteriori error estimation of goal-oriented quantities for elliptic type BVPs. *J. Comput. Appl. Math.* (2006) **191**, 216–227.

[18] S. Korotov, Error control in terms of linear functionals based on gradient averaging techniques *Comp. Lett.* (in press).

[19] S. Korotov, P. Neittaanmäki, and S. Repin, A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery. *J. Numer. Math.* (2003) **11**, 33–59.

[20] J. Nečas and I. Hlaváček, *Mathematical Theory of Elastic and Elasto-Plastic Bodies: An Introduction.* Elsevier Scientific Publishing Co., Amsterdam-New York, 1980.

[21] J. T. Oden and S. Prudhomme, Goal-oriented error estimation and adaptivity for the finite element method. *Comput. Math. Appl.* (2001) **41**, 735–756.

[22] J. Pitkäranta and R. Stenberg, Analysis of some mixed finite element methods for plane elasticity equations. *Math. Comp.* (1983) **41**, 399–423.

[23] S. Repin, A posteriori error estimation for variational problems with uniformly convex functionals. *Math. Comp.* (2000) **69**, 481–500.

[24] M. Rüter, S. Korotov and Ch. Steenbock, Goal-oriented error estimates based on different FE-solution spaces for the primal and the dual problem with applications to fracture mechanics. *Comput. Mech.* (in press).

[25] R. Stenberg, A family of mixed finite elements for the elasticity problem. *Numer. Math.* (1988) **53**, 513–538.

[26] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* Wiley-Teubner, 1996.

[27] O. C. Zienkiewicz and J. Z. Zhu, A simple error estimator and adaptive procedure for practical engineering analysis. *Internat. J. Numer. Methods Engrg.* (1987) **24**, 337–357.

Validation of CFD simulations in a mixing tank with stereo-PIV experiments

J. HEIKKINEN[(a)], M. HONKANEN[(b)], J. KORPIJÄRVI[(a)], P. SAARENRINNE[(b)], T.KILPIÖ[(c)]

[(a)]Numerola Oy
PL 126, Väinonkatu 7C, FIN-40101 Jyväskylä
e-mail: johanna.heikkinen@numerola.fi

[(b)]Tampere University of Technology, Energy and Process Engineering
PL 589, FIN-33101 Tampere
e-mail: markus.honkanen@tut.fi

[(c)] Kemira Oyj
PL 171, Typpitie 1, FIN-90101 Oulu
e-mail: teuvo.kilpio@kemira.com

ABSTRACT

This paper presents a comparison of experimental and simulated results of the flow field in a lab-scale mixing tank. Flow structures were measured using the high-speed Stereoscopic Particle Image Velocimetry, which enables an accurate, 3D description of the flow field in the tank. Computational Fluid Dynamics were performed on the tank and two turbulence models - SST and SAS - were utilized. Model validation against experimental data showed that the SAS model was able to capture the main characteristics of the trailing vortex pair, while the vortex trajectories terminated too early in the SST predictions.

1. INTRODUCTION

Flow dynamics in a mixing tank has been investigated by experimental and numerical techniques. The device consists of a supported paddle mixer located in a one litre beaker and a console to adjust the mixer speed. In the current work, the single phase, turbulent flow field of pure tap water was studied in the mixing tank. The goal has been to collect empirical data by Stereoscopic Particle Image Velocimetry (SPIV) technique and to use the experimental data to validate Computational Fluid Dynamics (CFD) simulations of the mixer.

The flow field of a mixing tank can generally be divided into a turbine discharge zone and a bulk region, which covers the rest of the tank. Both the experimental and numerical techniques must resolve the instantaneous fluid flow velocity fields in the turbine discharge zone. The time-averaged Reynolds decomposition cannot resolve flow periodicity caused by the rotating turbine. Therefore, the results are measured for each turbine blade angle. The turbulent fluctuations can be resolved for each instantaneous vector field based on the length scale of flow structures (with a Galilean decomposition) or for each turbine blade angle with an angle-resolved Reynolds decomposition.

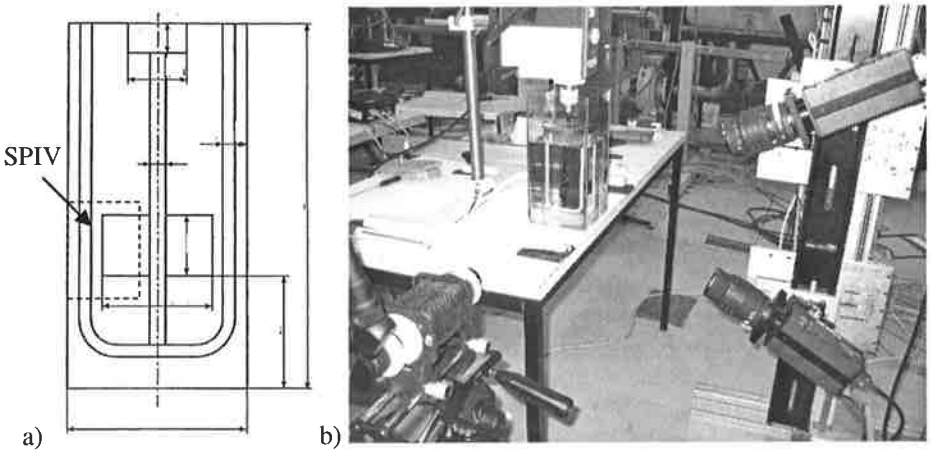


Figure 1. a) Scheme of the mixing tank. Dashed rectangle shows the measurement area. b) Experimental setup.

2. STEREO-PIV EXPERIMENTS

Stereoscopic Particle Image Velocimetry (SPIV, [1,2]) experiments have been carried out with a high-speed PIV system that consists of two high-speed, digital, double-frame CMOS cameras (PCO) that obtain 600 Hz frame rate with a resolution of 1280x1024 pixels and a high-speed Nd-YLF laser (New Wave, Pegasus) with a 10 mJ pulse energy at kHz pulse rate. The cameras are arranged in an angular displacement system that is geometrically calibrated to view a 49x36x1 mm^3 measurement volume in the cross section of the mixing tank. The measurement volume is located in the turbine discharge zone and it is highlighted in Figures 1a and 2 (right). The measurement plane is perpendicular to the plane of two carrier bars. The measurement setup is shown in Figure 1b. The angle between the cameras is 60 degrees. All three components of fluid velocity are acquired in the measurement plane from image sequences of illuminated tracer particle images.

Each instantaneous velocity field consists of 22600 velocity vectors with interval of 280 μm that defines the spatial resolution of the measurement. In case of 160 rpm turbine speed, 1300 velocity fields are resolved at 192 Hz sample rate, which corresponds to the total of 18 rotation cycles. The turbine blade moves 5 degrees between each resolved velocity field. The time delay between the frames of each double-frame image is 0.6 ms, which ensures that the measured velocity fields are truly instantaneous. The measurement does not resolve the smallest turbulence scales at 160 rpm turbine speed. The minimum Kolmogorov time scale is about 1.25 ms, minimum velocity scale is 0.03 m/s and the minimum length scale is about 40 μm. The spatial resolution of measurements is 7 times the Kolmogorov length scale and the time resolution of measurements is 4 times the Kolmogorov time scale. Therefore, the turbulent kinetic energy can be resolved almost totally, but only about half of the turbulent kinetic energy dissipation rate is measured [3]. The large-scale flow structures and their instabilities are measured with high accuracy and the measurement results are utilized to validate the numerical model. The periodic and aperiodic (turbulent) motion of trailing vortices is resolved both in experiments and in simulations.

3. CFD MODELLING

In CFD modelling, conservation equations of mass, momentum and other related variables are set up in order to model the flow field of the system. These partial differential equations are solved by integrating them over small control volumes, into which the computation domain has been discretized (control volume method). In many engineering applications, one is often interested in a steady-state solution rather than an instantaneous flow field that fluctuates due to e.g. changing process conditions and turbulence. Furthermore, an exact description of the transient, turbulent flow field in most practical applications is still too time-consuming to accomplish. Therefore, Reynolds Averaged Navier-Stokes (RANS) approach is commonly used, which leads to two additional equations for turbulent kinetic energy and dissipation.

Many engineering applications are turbulent. Turbulence must therefore be included in the CFD model. The complete spectrum of turbulence length- and time-scales can be exactly solved by Direct Numerical Simulation (DNS). It is well known that with the current and near-future computer capacity, DNS is too heavy in terms of computer power and time. This is why traditionally, turbulence has been modelled by two-equation models together with Boussinesq's eddy viscosity hypothesis that relates the turbulent Reynolds stresses to the velocity gradients of the mean velocity field. The two-equation turbulence models have proven to yield acceptable predictions in a wide range of applications and have therefore a solid status in CFD modelling. However, the most two-equation models fail to correctly predict turbulent vortex shedding, which is of interest in many engineering applications. Alternative methods located somewhere between DNS and 2-equation turbulence models, such as Large Eddy Simulation (LES), Detached Eddy Simulation (DES) and Scale-Adaptive Simulation (SAS) models have been proposed to describe turbulence in computational codes.

In the current work, Computational Fluid Dynamics (CFD) simulations have been performed on a mixing tank. Flow field in this baffled mixing tank has been modelled by ANSYS CFX-10.0 code using two different turbulence models: the shear-stress-transport (SST) model and the Scale-Adaptive Simulation (SAS) both developed by F.R. Menter [4, 5, 6].

The SST turbulence model provides an advanced 2-equation turbulence closure among the traditional 2-equation models as it has been shown to better predict flow separation under adverse pressure gradients. Improvements of the k-ω based SST model are based on the inclusion of transport effects into the formulation of the eddy-viscosity. The transport equations of turbulent kinetic energy (k) and turbulent frequency (ω) are:

$$\frac{\partial(\rho k)}{\partial t}+\frac{\partial(\rho U_i k)}{\partial x_i}=\tilde{P}_k-\beta^*\rho k\omega+\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_i}\right] \tag{1}$$

$$\frac{\partial(\rho\omega)}{\partial t}+\frac{\partial(\rho U_i\omega)}{\partial x_i}=\alpha\rho S^2-\beta\rho\omega^2+\frac{\partial}{\partial x_i}\left[\left(\mu+\frac{\mu_t}{\sigma_\omega}\right)\frac{\partial\omega}{\partial x_i}\right]+(1-F_1)\frac{2\rho}{\sigma_{w2}}\frac{1}{\omega}\frac{\partial k}{\partial x_i}\frac{\partial\omega}{\partial x_i} \tag{2}$$

Here $\tilde{P}_k$ is the turbulence production term, S is the shear strain rate and F_1 is the blending function. Betas and sigmas are model constants, see e.g. [7] for further information. The formulation of the blending function is based on the distance to the nearest surface and on the flow variables.

The SST model actually combines the strengths of the standard k-ε turbulence model and the k-ω model by blending between these two models. The k-ω model is used in the boundary layer, while the standard k-ε turbulence model is utilised in the free stream. As mentioned above, a shortcoming of the 2-equation turbulence models is that they are not able to predict the formation of complex 3D vortices under adverse pressure gradients (e.g. behind obstacles). This has been claimed to be caused by over-predicted turbulence levels rather than their derivation [5]. Menter and Egorov argued that the URANS (Unsteady Reynolds Averaged Navier-Stokes) solution is overly dissipative and therefore prevents the formation of a turbulent cascade [6].

The SAS concept allows a two-equation turbulence model to function in a SAS mode, which results in a LES-like behaviour in the unsteady regions of the flow field. So far, SAS modes of KE1E and the SST turbulence models have been reported [5, 6]. In this work, SST-SAS has been used and is hereby referred to as SAS for simplicity. The SAS method is based on introducing the von Karman length-scale into the turbulence length scale equation originally derived by Rotta [8]. The von Karman length-scale reads [6]:

$$L_{vK} = \kappa \left| \frac{\partial U / \partial y}{\partial^2 U / \partial y^2} \right| \tag{3}$$

Here κ is the von Karman constant. When the SST turbulence model is operated in SAS mode, the transport equation of turbulent frequency becomes [6]:

$$\frac{\partial(\rho\omega)}{\partial t} + \frac{\partial(\rho U_i \omega)}{\partial x_i} = \alpha\rho S^2 - \beta\rho\omega^2 + \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_\omega}\right)\frac{\partial\omega}{\partial x_i}\right] + (1-F_1)\frac{2\rho}{\sigma_{w2}}\underbrace{\left(\frac{1}{\omega}\frac{\partial k}{\partial x_i}\frac{\partial\omega}{\partial x_i} - \frac{k}{\omega^2}\frac{\partial\omega}{\partial x_i}\frac{\partial k}{\partial x_i}\right)}_{SAS-terms} + \xi\kappa\rho S^2 \frac{L}{L_{vK}} \tag{4}$$

It has been shown that the SAS modes of KE1E and SST models are able to predict the formation of 3D vortices in the unsteady flow regions [5, 6]. According to Menter and co-workers [5], the reason why the URANS models usually fail to model the formation of turbulent structures is a result of over-predicted turbulence levels. Namely, the standard turbulence models provide a length-scale proportional to the thickness of shear layer, which does not adjust to the local flow topology. On the contrary, SAS formulation is based on the von Karman length scale that adjusts to the readily resolved turbulence scales, hence the length scale is proportional to the size of local eddies. In the derivation of the SAS model, Menter and co-researchers used Rotta's length scale equation in a modified form. More specifically, in [9] Menter and Egorov note that the omission of second order terms in Taylor series expansion of Rotta's length scale definition, which is based on the assumption of homogenous turbulence, is overly restrictive. Therefore the second order term is included in the SAS formulation of the turbulence length scale.

Model set-up

The computational domain was divided into a static stator and a rotating rotor domain, between which a transient sliding mesh procedure was utilised. General Grid Interface (GGI) connection was used on the domain interface to account for the non-matching nodes. Figure 2 (left) illustrates the stator and rotor domains and the spatial discretization used. The mesh consisted mainly of tetrahedral elements, but three layers of prism elements were added on the surfaces of the mixing vessel, impeller and baffles in order to more accurately predict the boundary layer. In total, the mesh had 542144 nodes equalling to about 2.77M elements.

The rotational speed of the impeller was 160 rpm, corresponding to the Impeller Reynolds number of about 9400. The transient rotor-stator model of ANSYS CFX 10.0 was used to simulate the actual movement of the rotating domain. Two turbulence models, i.e. SST and SAS were utilised. Together with the SST-based turbulence models, ANSYS CFX uses automatic wall functions. The automatic wall functions automatically switch from wall-functions (log law of the wall) to a low-Re near wall formulation when the mesh size normal to the wall is small enough. The transient terms of the discrete finite volume equations were discretized according to the Second Order Backward Euler scheme. For the advection terms, a factor of 0.75 was used to blend between a 2nd order High Resolution Scheme and a 1st order Upwind Difference Scheme. Factor 1 means a full 2nd order scheme.

Initial values for the time-dependent simulations were obtained from a steady-state SST simulation, which converged in less than 400 iterations. The SAS simulation was run for 20 impeller revolutions and the SST model for 16 revolutions before the actual data reported in this paper was collected. The intention was to make sure that a semi steady-state flow was reached and to eliminate the influence of the initial flow field on the calculated results. A time step corresponding to 2.5° impeller rotation was used and data was collected every 5° for about 9 impeller revolutions. The simulations were performed in parallel on four AMD Opteron 2.4 GHz processors each having 4 Gbytes of memory. A CPU time required per iteration was approximately 4 minutes and 10 seconds.

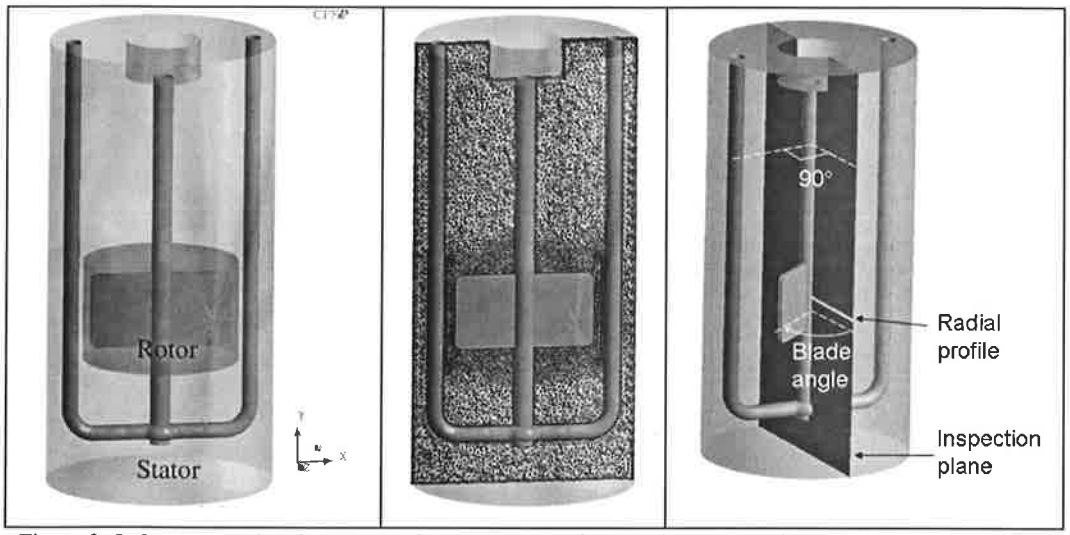


Figure 2. *Left: computational geometry showing rotor and stator domains mesh, middle: computational mesh, right: locations of radial profile and inspection plane.*

4. 3D TRAILING VORTICES

Figure 3 shows the turbulence structures resolved by the simulations. When SAS and SST turbulence models are compared, it is noted that SAS is capable of predicting the formation of 3D vortices that are chaotic in nature. The turbulent structures calculated by the SST model are much smoother in nature and exist to a lesser extent. The isosurface drawn in Figure 3 is the strain rate invariant at 5000 s^{-2} coloured by the ratio of eddy viscosity/dynamic viscosity. It can be seen that the eddy viscosity of the trailing vortices predicted by the SST model is much higher than that of SAS model. Therefore it seems that the URANS model used here (SST) indeed over-predicts the turbulence levels, which prevents the formation of the turbulent cascade as already mentioned in Section 3. On the particular isosurface in Figure 3, according to the SST model the viscosity ratio has a maximum of 73 and an average of 18.9, which are considerably higher than the corresponding values of the SAS model (6.9 and 2.6, respectively).

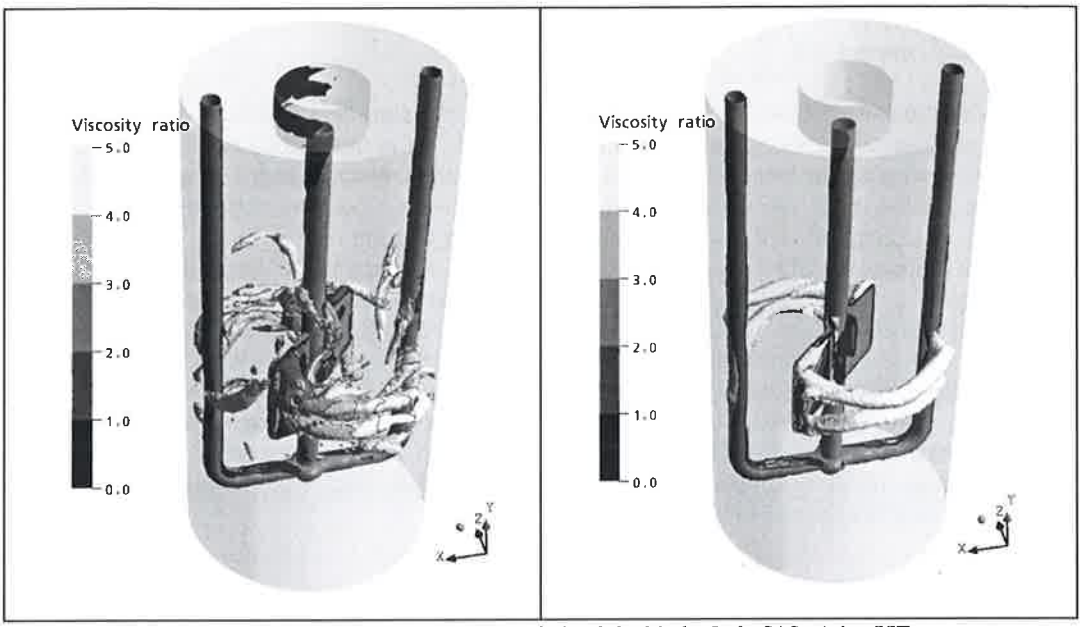


Figure 3. Turbulent structures of the trailing vortices behind the blade. Left: SAS, right: SST.

5. MODEL VALIDATION WITH EXPERIMENTAL DATA

Even if the 3D turbulent structures displayed in Figure 3 are illustrative, they cannot be easily used for model validation. Instead, radial velocity profiles and the trajectories of the vortex mid-points are here used to validate simulation results. In order to compare the modelled and measured flow fields, they were interpolated to a regular grid with a grid spacing of 2 mm and a step of 5 degrees in the turbine blade angle. The evolution of large-scale flow structures *i.e.* trailing vortices in the turbine discharge zone can then be compared.

Trajectories of trailing vortices

The turbine has straight, flat blades. The flow therefore arrives axially into the turbine and detaches it radially. The turbine blade creates trailing vortices to its both tip corners and the vortices generate the discharge zone of the turbine. Interaction of the vortices accelerates their radial motion, while diffusion and viscous forces decelerate it. Initially vortices move axially towards each other, but when their radial acceleration stops, they start moving axially away from each other. When the vortices encounter the boundary layer of the vessel wall, the radial motion decelerates. At latter blade angles > 130° the vortices either bounce back from the boundary layer or vanish depending on the location of the stagnation point on the wall. The trailing vortices are mainly aligned so that the maximum component of the rotation tensor is the tangential component.

The trailing vortices are located from the data as follows. The spatial derivatives of radial (u,x) and axial (v,y) velocity components in the tank cross section are computed with a central difference scheme. The tangential component of rotation is computed from $\omega_z = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$. It has a positive sign for the upper trailing vortex and a negative sign for the lower trailing vortex. The magnitude of ω_z may be high where the flow changes rapidly, which, however, does not always indicate the location of a vortex. Therefore, another vortex detection parameter, the eigenvalues of the velocity gradient tensor [10] is utilized. The negative imaginary part of the eigenvalues of the velocity gradient tensor is called as swirling strength. Swirling strength in the plane is derived as

$$swirling\_strength = 4\left(\frac{du}{dx}\cdot\frac{dv}{dy}-\frac{du}{dy}\cdot\frac{dv}{dx}\right)-\frac{1}{2}\left(\frac{du}{dx}+\frac{dv}{dy}\right)^2 \tag{5}$$

The local maxima of swirling strength exceeding a given threshold indicate the positions of the trailing vortices, whose direction of rotation is defined by ω_t in that particular location. A subgrid accuracy of locating the local maxima of swirling strength is obtained with a 3-point Gaussian fit. The vortex trajectories are composed with a vortex tracking function that utilizes the nearest neighbour-search to follow the vortices in the consecutive velocity fields.

Figure 4 shows the 3D trajectories of trailing vortices detected in the experimental data and in the simulated data calculated by SAS and SST turbulence models. It can be seen that the SAS turbulence model provides results that are similar to the experimental data. Some discrepancies can be noted with small blade angles (0-10°), where the experimental data shows more scatter. On the other hand, the measurement locations close to the blade are more prone to errors than angles further away from the blade. When turbulence is described by the SST model, vortices are not detected at blade angles > 120° but they vanish on the vessel wall boundary layer. This implies that the dissipation of the turbulent structures is overestimated. SAS also slightly over-estimates dissipation compared to experiments.

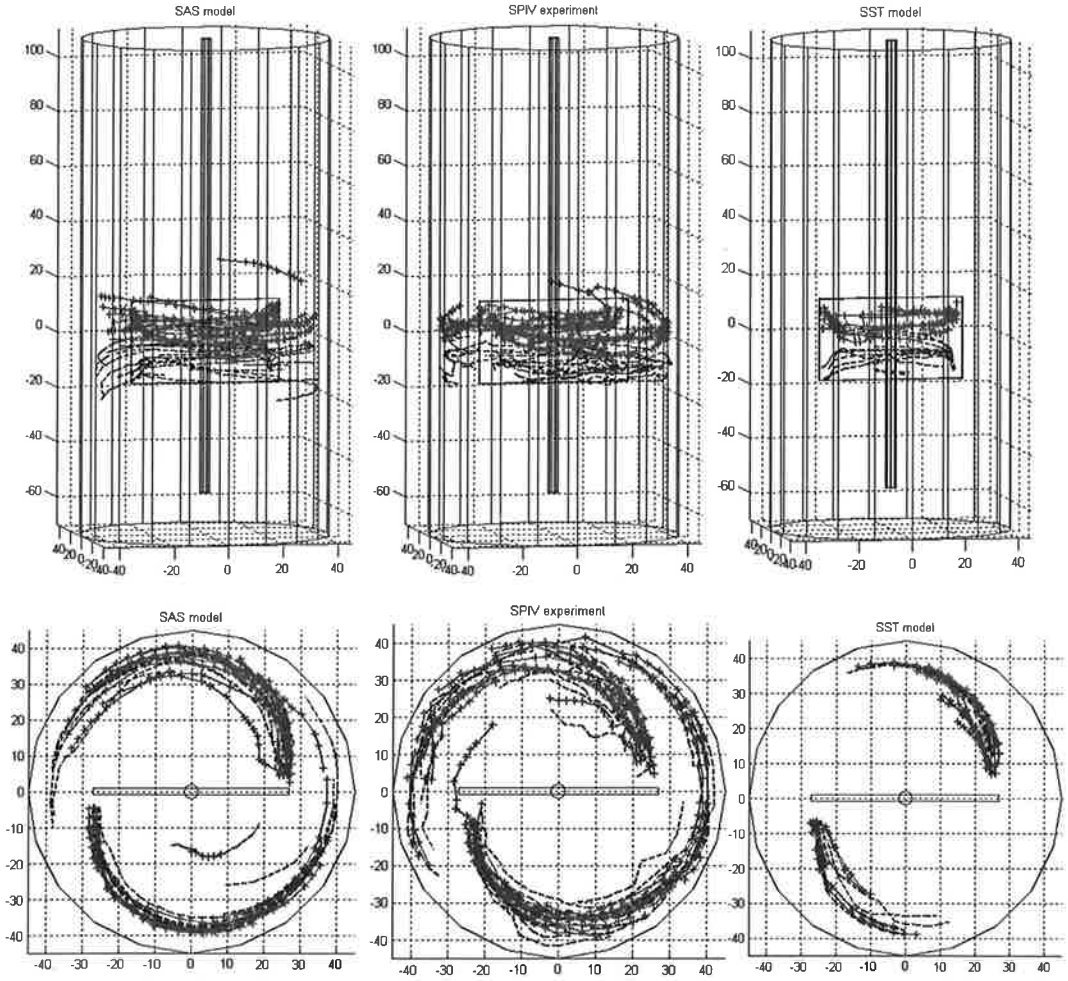


Figure 4. The three-dimensional trajectories of trailing vortices. Left SAS model, middle experimental, right SST model. The upper images show the side view, lower the top view.

Figure 5 depicts the average 3D trajectories of trailing vortices detected in the experimental and CFD data. Again it is noted that the vortex path lines calculated by the SAS turbulence model resemble the experimentally identified trajectories, while the path lines of the SST model terminate when the blade angle exceeds 120°. From Figure 5 one can also see that at blade angles < 20°, the path lines of experimental and SAS data start closer to the vessel wall than one would expect. This can be explained by the fact that vortex "tails" from the previous revolution are present in the instantaneous data, which were included in the averaging process. From Figure 4 one can see that in the SPIV experiments, the vortices from a previous round are detected close to the vessel wall. The large scatter of the SPIV data at small blade angles in Figure 5 is mainly due to vortex tail from previous blade pass. Due to somewhat more dissipation in SAS, the vortices from previous revolution are not detected, compare Figures 4 and 5.

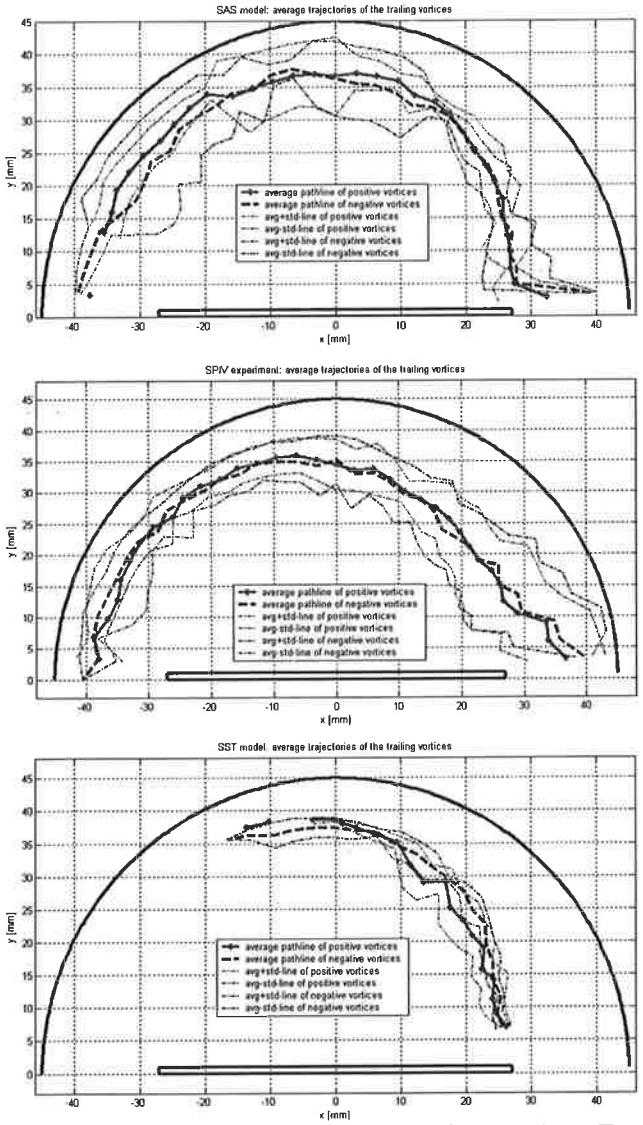


Figure 5. The average path lines of trailing vortices. Top: SAS model, middle: experimental, bottom: SST turbulence model.

In Figure 6 the mean radial and axial velocity components for the vortex midpoints are shown as a function of the turbine blade angle. These profiles hence indicate average speed and direction of the midpoints at each blade angle. Figure 6 (a) reveals that for blade angles smaller than 60°, the experimentally determined radial velocity of the midpoints is lower than what the models predict. On the other hand, when the blade angle exceeds 100°, the midpoints move faster towards the vessel wall in the experimental data than in the simulations. As noted before, no trailing vortices are detected in the SST data for blade angles > 120°.

The average axial velocity of the vortex midpoints in Figure 6 (b) shows whether vortices move up- or downwards in the vessel. According to the CFD simulations, the positive (= upper) vortices have a higher negative axial velocity and the negative (= lower) vortices a higher positive axial velocity. In other words, according the model the midpoints approach each other faster than what the experimental SPIV data implies. The experimental experience gained in this work indicates that the closer the vortices of the vortex pair are to each other, the faster they move towards the wall. This is why the trailing vortices in the SAS model (Figure 5) travel somewhat closer to the vessel wall than in the experiments. It is possible that the model over-predicts the strength of vortices, which would cause the over-estimated axial velocities. On the other hand, the SPIV data at small blade angles includes also the vortex tails from previous blade passes, which decrease the average midpoint velocities.

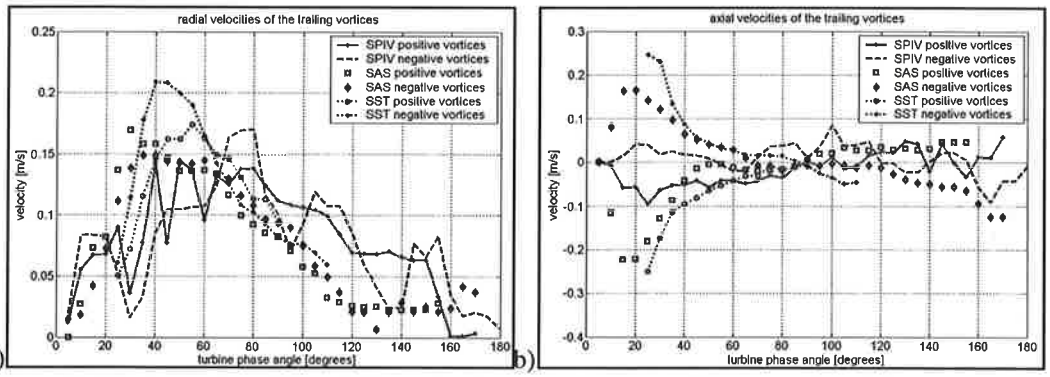


Figure 6. The mean a) radial and b) axial velocity components of the midpoint of the positive and negative trailing vortex at each turbine blade angle.

Velocity profiles in radial direction

Velocity profiles in radial direction are resolved at blade angles from 5 to 180 degrees with a step of 5°. These profiles reported here are taken on a plane perpendicular to the baffles at a height of 70 mm, which is in the middle of the turbine blade height (See also Figure 2, right). The experimental data were averaged from 35 revolutions, while the simulated data just from 9 revolutions. However, as the turbine blade is symmetrical, two data sets per simulated revolution were obtained; hence the modelled radial profiles reported here are averages of 16-18 data sets. In stead of the arithmetic average, median was utilised to analyse the profiles calculated by CFD, as median is somewhat more robust with regard to the extremes of fluctuations.

The radial profiles of radial, axial and tangential velocity at blade angles of 20, 40, 80 and 120° are collected in Figures 7-9. The radial velocity profiles in Figure 7 confirm the conclusions drawn from Figure 6 (a), *i.e.* at small blade angles the models over-predict the radial velocity and at large blade angles under-predict the radial velocity. At small blade angles the differences between SAS and SST model are small, but at large angles SST model underestimates the radial velocity magnitude. The radial profiles of axial velocity depicted in Figure 8 suffer from large scatter.

Therefore the profiles shown here are far from being developed and more data would be needed to draw conclusions. Finally, the radial profiles of tangential velocity (Figure 9) indicate an acceptable agreement between the models and SPIV experiments. Moreover, no large differences in the results by SAS and SST model are observed.

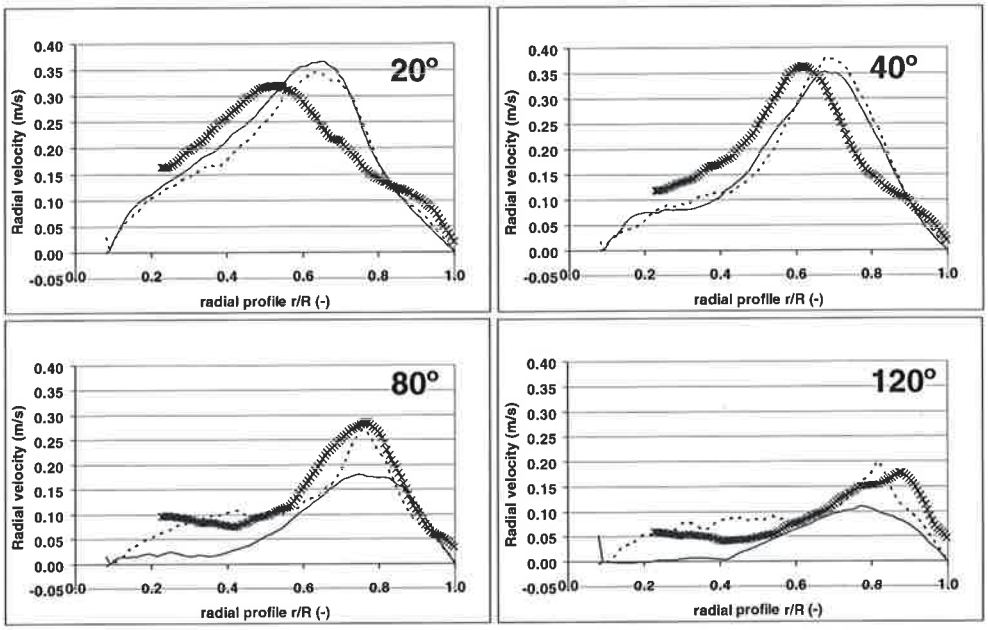


Figure 7. Radial velocity profiles in radial direction at blade angles of 20, 40, 80 and 120°. (x) experimental, (--) SAS model, (—) SST model.

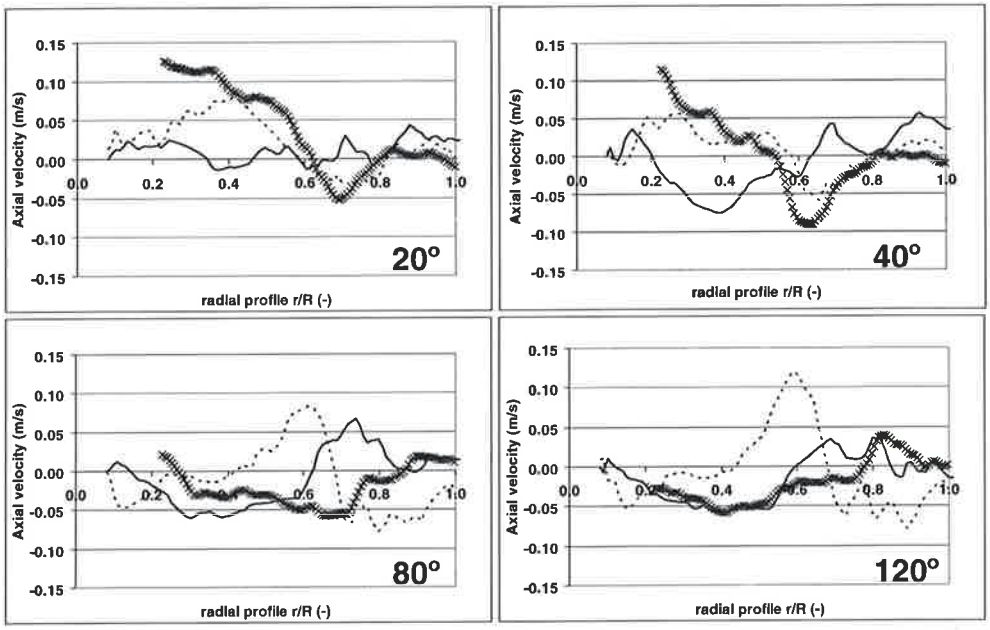


Figure 8. Axial velocity profiles in radial direction at blade angles of 20, 40, 80 and 120°. (x) experimental, (--) SAS model, (—) SST model.

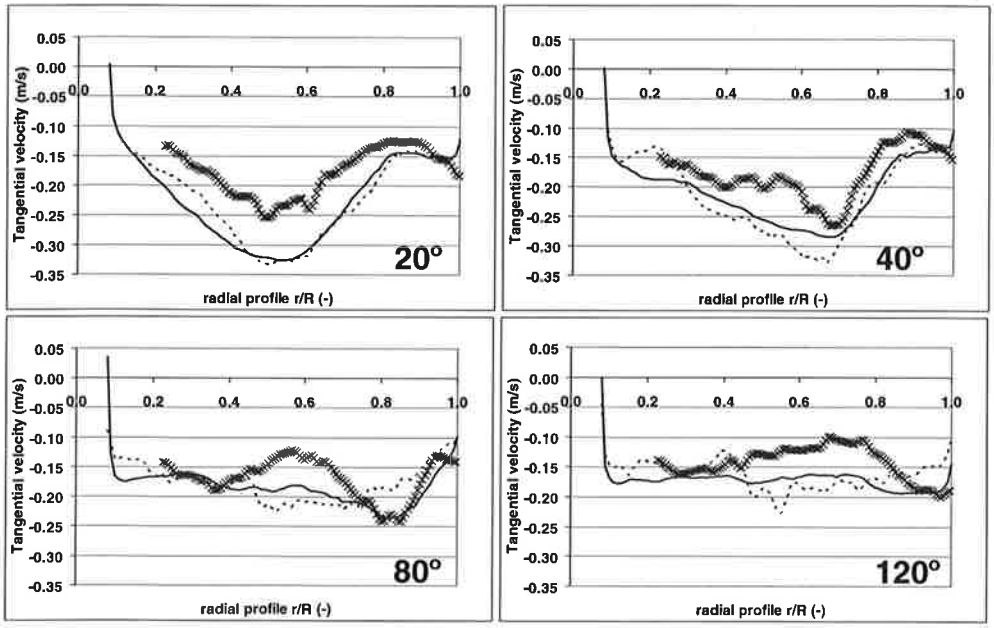


Figure 9. Tangential velocity profiles in radial direction at blade angles of 20, 40, 80 and 120°. (x) experimental, (--) SAS model, (—) SST model.

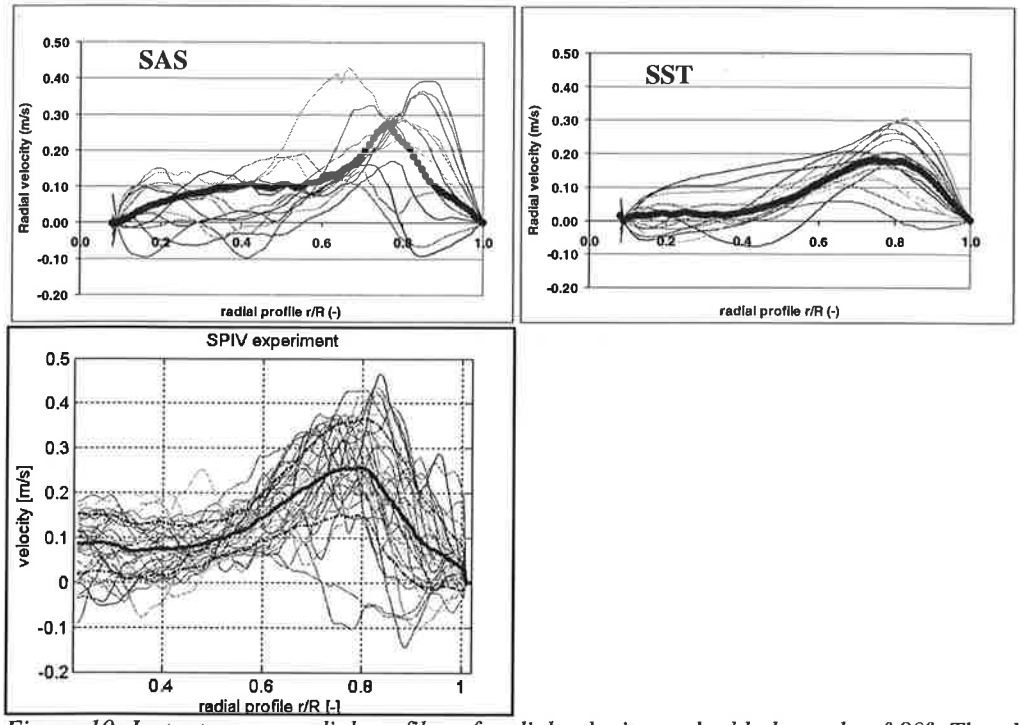


Figure 10. Instantaneous radial profiles of radial velocity at the blade angle of 80°. The phase-resolved average is shown in thick black line. Top left: SAS, top right: SST, bottom: SPIV.

Figure 10 shows multiple instantaneous radial profiles of radial velocity at the blade angle of 80°. It can be seen that fluctuation are large and random, particularly in the data of SAS model and SPIV experiments. The SST model also predicts fluctuations, but these are smoother and somewhat smaller in extent. From this data it is obvious that one would need more data in order to draw definite conclusions.

6. CONCLUSIONS

Flow field in a lab-scale mixing tank has been modelled by CFD and the modelling results have been validated against experimental SPIV data. Unique results has been reported, as the Stereo-PIV technique offers an advanced means to capture the 3D flow field in the mixing tank. In CFD modelling, two turbulence models, i.e. SST and SAS have been used. It has been shown that the SAS model is able to predict the formation of chaotic, 3D trailing vortices in the turbine discharge zone. It therefore shows LES-like behaviour without requiring extensively fine computational meshes. On the contrary, the trailing vortices predicted by the URANS SST model were much smoother and overly dissipative in nature and hardly any smaller-scale vortices were noted, which existed to a high-extent in SAS results.

The CFD model was successfully validated against radial velocity profiles and by following the trajectories of the midpoints of the trailing vortex pair. These showed that the SAS model was able to capture the main path lines of the trailing vortex pair, while in the case of SST model the path lines terminated at a blade angle of about 120°. In general, no large difference between the predictions by SAS and SST model were noted in the radial profiles. The radial profiles showed that an acceptable agreement was obtained in most cases. However, in order to draw definite conclusions, the simulation should be run for more revolutions. Unfortunately this was not possible due to long CPU time required by these transient simulations within this work, but the research will be continued as the presented results are very promising.

REREFENCES

1. C. Willert, *Stereoscopic digital particle image velocimetry for application in wind tunnels.* Meas. Sci. Technol. 8:1465-1479, 1997.
2. A. K. Prasad, *Stereoscopic particle image velocimetry.* Exp. Fluids 29:103-116, 2000.
3. P. Saarenrinne, M. Piirto, H. Eloranta, *Experiences of turbulence measurement with PIV.* Meas. Sci. Technol. 12:1904-1910, 2001.
4. F.R. Menter, *Two-equation eddy-viscosity turbulence models for engineering applications.* AIAA-Journal, 32(8), 1994.
5. F.R. Menter, M. Kuntz and R. Bender, *A scale-adaptive simulation model for turbulent flow predictions.* AIAA Paper 2003-0767, 2003.
6. F.R. Menter and Y. Egorov, *A scale-adaptive simulation model using two-equation models.* AIAA Paper 2005-1095, 2005.
7. F.R. Menter, M. Kuntz and R. Langtry, *Ten years of industrial experience with the SST turbulence model.* International Symposium on Turbulence, Heat and Mass Transfer 4, Antalya, Turkey, October 12-17, 2003.
8. J.C. Rotta, *Turbulente Strömungen.* Teubner Verlag, Stuttgart, 1972.
9. F.R. Menter and Y. Egorov, *Revisiting the turbulent scale equation.* Proc. IUTAM Symposium, Göttingen, Germany, August 12-14, 2004.
10. J. J. Jeong, F. Hussain, *On the identification of a vortex.* J. Fluid Mech. 285:69-94, 1995.

VAIHTOEHTOISIA MENETELMIÄ MIKROSYSTEEMIEN VAIMENNUSOMINAISUUKSIEN LASKEMISEEN

P. RÅBACK[1], M. MALINEN[1], A. PURSULA[1], T. VEIJOLA[2]

[1]Tieteen tietotekniikan keskus CSC
PL 405, FIN-02101 Espoo
e-mail: Peter.Raback@csc.fi

[2]Teknillinen korkeakoulu
PL 3000, FIN-02015 HUT

TIIVISTELMÄ

Mikrosysteemien vaimennuskäyttäytymistä mallinnetaan yleensä Navierin ja Stokesin yhtälöistä johdetuilla yksinkertaistuksilla. Tässä paperissa käsitellään virtauksen ratkaisua käyttäen kolmea eri mallia: linearisoitua Navierin ja Stokesin yhtälöä, Stokesin yhtälöä sekä dimensioreduktiolla saatavaa Reynoldsin yhtälöä. Yhtälöt on toteutettu elementtimenetelmään pohjautuvaan vapaan lähdekoodin Elmer-ohjelmistoon. Yhtälöiden soveltamisesta annetaan käytännön esimerkkejä ja lisäksi käsitellään niiden soveltuvuutta eri tilanteisiin ja yhdistämistä erilaisten geometristen rakenteiden käsittelyyn.

1 JOHDANTO

Mikrosysteemien mallinnus on haastava monifysikaalinen tehtäväkenttä, jossa sähköiset ja mekaaniset voimat ovat vahvasti toisiinsa kytkeytyneitä. Näiden lisäksi tulee usein huomioida myös kaasun virtaus, joka on pääasiallinen vaimennuksen lähde monissa mikromekaanisissa komponenteissa.

Mikrosysteemeissä kaasun virtauksella on tiettyjä erityispiirteitä, jotka eivät yleensä päde suuremmassa mittakaavassa. Pienet mitat tuovat haasteita mallinnukselle, sillä Navierin ja Stokesin yhtälöiden takana oleva jatkumo-oletus ei välttämättä enää päde kovinkaan hyvin. Jatkumoyhtälöiden pätevyysaluetta voidaan kuitenkin jonkin verran laajentaa modifioimalla virtausyhtälöitä tai niiden reunaehtoja. Myös mikrosysteemien rakenteiden geometria on usein poikkeuksellinen sisältäen toistuvia rakenteita sekä hankalia aspektisuhteita. Toisaalta pienet dimensiot myös helpottavat mallitusta, sillä virtauksen voidaan alhaisen Reynoldsin luvun takia lähes aina olettaa olevan laminaaria. Mikrosysteemien erityispiirteenä on usein myös ilmiöiden jaksollisuus, mistä syystä virtausyhtälöidenkin ratkaisua kannattaa hakea aikaharmonisella yritteellä.

Suoraviivaisin tapa kaasun virtauksen aiheuttamien voimien laskentaan olisi ajasta riippuvien Navierin ja Stokesin yhtälöiden ratkaiseminen. Valitettavasti niiden suora soveltaminen on kuitenkin laskennallisesti niin raskasta, että se on järkevää vain poikkeustapauksissa. Yllä kuvattujen virtauksen erityispiirteiden ansiosta täydellisten yhtälöiden soveltaminen ei kuitenkaan yleensä ole edes tarpeellista. Useimmissa tilanteissa voidaan käyttää jotain Navierin ja Stokesin yhtälöistä johdettua erikoistapausta, jonka laskennallinen ratkaisu on

huomattavasti kevyempää. Joskus on myös mahdollista yhdistää eri virtausmalleja siten, että toistuvat pienet yksityiskohdat parametrisoidaan koko systeemiä kuvaaviin virtausyhtälöihin.

Mikrosysteemien suunnitelijan työkalu on yleensä piirisimulaattoriohjelmisto (esim. Aplac tai Spice). Näin ollen virtauslaskennasta saatava tieto olisi mieluiten saatettava kompaktiin muotoon, jota voidaan käsitellä myös piirisimulaattorissa. Yleisessä tapauksessa virtauslaskenta sisältää epälineaarisuuksia, joiden kuvaus yleispäteväksi lumpatuksi malliksi on hyvin epätriviaalia. Yleensä rajoitutaankin tutkimaan kaasun lineaarista käyttäytymistä. Tämä osin pragmaattinen valinta ei käytännössä rajoita mallien tarkkuutta kuin harvoissa poikkeustapauksissa. Myös tässä paperissa rajoitutaan lineaaristen virtausmallien käsittelyyn.

2 MATEMAATTISET MALLIT

2.1 Linearisoitu Navierin ja Stokesin yhtälö

Mikrosysteemien käyttäytymisen kannalta kriittinen kaasun virtaus tapahtuu liikkuvien osien välisissä kapeissa raoissa. Kaasun hitausvoimien ja viskoosien voimien suhdetta kuvaava Reynoldsin luku on näissä niin alhainen, että hitausvoimien vaikutus voidaan jättää huomiotta. Yleensä myös paineen vaihtelut referenssipaineeseen nähden ovat sen verran pienet, että virtausta voidaan riittävällä tarkkuudella kuvata linearisoiduilla Navierin ja Stokesin yhtälöillä [1, 2]

$$\begin{aligned} \rho_0 \frac{\partial \vec{v}}{\partial t} &= -\nabla p + (\mu + \eta)\nabla\nabla \cdot \vec{v} + \eta \Delta \vec{v}, \\ \frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \cdot \vec{v}, \end{aligned} \tag{1}$$

missä $\vec{v}$ on nopeus, ρ tiheys, p paine sekä μ ja η kaasun viskositeetit. Alaindeksillä "0" viitataan tässä tasapainotilan olosuhteisiin.

Jatkuvuusyhtälössä tiheys voi riippua sekä paineesta että lämpötilasta. Kaasun kokoonpuristuvuus on usein merkittävää, eikä tällöin virtauksen voida automaattisesti olettaa olevan isotermistä, joten yhtälöitä tulee täydentää linearisoidulla energiayhtälöllä

$$\rho_0 C_p \frac{\partial T}{\partial t} = \kappa \Delta T + \frac{\partial p}{\partial t}, \tag{2}$$

missä C_p on ominaislämpökapasiteetti vakiopaineessa, T lämpötila ja κ lämmönjohtavuus. Kaasun kitkahäviöistä aiheutuva lämmitys on mikrosysteemeissä niin vähäistä, että sitä ei tarvitse huomioida. Yllä esitetetyn yhtälöryhmän sulkemiseen soveltuu ideaalikaasun tilanyhtälö

$$p = \rho R T, \tag{3}$$

missä R on kaasuvakio.

Esitettyjen yhtälöiden lisäksi tarvitaan sopivat reunaehdot. Tyypillisesti mikrosysteemeissä reunaehdot pinnan normaalin $\vec{n}$ suunnassa määrää seinämän nopeus $\vec{v}_w$ siten, että

$$\vec{v} \cdot \vec{n} - \vec{v}_w \cdot \vec{n} = 0. \tag{4}$$

Tangentin $\vec{t}$ suunnassa ei voida olettaa liukumattomuusehtoja nopeudelle, jos molekyylien törmäysten välinen vapaa matka λ on samaa suuruusluokkaa kuin kanavan korkeus d. Tätä kuvaa ns. Knudsenin luku

$$K_n = \frac{\lambda}{d}. \tag{5}$$

Ideaalikaasuille vapaa matka voidaan arvioida yhtälöstä

$$\lambda = \eta\sqrt{\frac{\pi}{2p\rho}}. \tag{6}$$

Mikäli $0 < K_n < 0.1$, voidaan soveltaa liukuehtoa nopeudelle [3]

$$\vec{v}\cdot\vec{t} - \vec{v}_w\cdot\vec{t} = \frac{2-\sigma_v}{\sigma_v}\frac{1}{\rho\sqrt{2RT_w/\pi}}\vec{f}\cdot\vec{t}, \tag{7}$$

missä $\vec{f}$ on virtauksen seinämään kohdistama pintavoima.

Liukuehtoja vastaava reunaehto kaasun lämpötilalle on [3]

$$T - T_w = \frac{2-\sigma_T}{\sigma_T}\frac{2(\gamma-1)}{\gamma+1}\frac{1}{R\rho\sqrt{2RT_0/\pi}}\vec{q}\cdot\vec{n}, \tag{8}$$

missä T_w on seinämän lämpötila ja $\vec{q}$ lämpövuo yli pinnan ja $\gamma = C_p/C_v$ ominaislämpökapasiteettien suhde. Parametrit σ_v ja σ_T ovat liikemäärän ja energian vaihtoon liittyviä kertoimia. Arvo $\sigma = 1$ kuvaa täydellistä liikemäärän tai energian vaihtoa kaasun ja kappaleen pinnan molekyylien välisissä törmäyksissä. Todelliset arvot ovat hieman tätä pienempiä.

Mikäli $K_n > 0.1$, ei Navierin ja Stokesin yhtälöitä enää pelasta mikään. Tällöin on turvauduttava korkeamman kertaluvun jatkumoyhtälöihin (Burnettin yhtälö) tai molekyylidynaamisiin Monte Carlo -simulaatioihin. Näiden käyttö merkitsee kuitenkin suurta hyppyä sekä yhtälöiden mutkikkuudessa että tarvittavassa laskenta-ajassa.

2.2 Aikaharmoninen Navierin ja Stokesin yhtälö

Usein mikrosysteemeissä on kiinnostavaa tuntea kaasun vaste värähtelevän kappaleen liikkeeseen. Tällöin yhtälöiden (1)–(3) ratkaisua haetaan tasapainopisteen ympärillä aikaharmonisella yritteellä

$$\begin{aligned}
\vec{v}(\vec{x},t) &= \vec{v}(\vec{x})\exp(\mathrm{i}\omega t),\\
\rho(\vec{x},t) &= \rho_0 + \rho(\vec{x})\exp(\mathrm{i}\omega t),\\
p(\vec{x},t) &= p_0 + p(\vec{x})\exp(\mathrm{i}\omega t),\\
T(\vec{x},t) &= T_0 + T(\vec{x})\exp(\mathrm{i}\omega t),
\end{aligned} \tag{9}$$

missä ω on kulmanopeus.

Sijoittamalla yritteen saamme kompleksisen yhtälöryhmän amplitudeille $\vec{v}$, ρ, p ja T.

$$\begin{aligned}
\rho_0\mathrm{i}\omega\vec{v} &= -\nabla p + (\lambda+\eta)\nabla\nabla\cdot\vec{v} + \eta\Delta\vec{v},\\
\mathrm{i}\omega\rho &= -\rho_0\nabla\cdot\vec{v}\\
\rho_0 C_p\mathrm{i}\omega T &= \kappa\Delta T + \mathrm{i}\omega\frac{\gamma-1}{\gamma}p,\\
p &= RT_0\rho + R\rho_0 T
\end{aligned} \tag{10}$$

Yhtälöistä voidaan suoraan eliminoida pois esimerkiksi paine ja tiheys. Käytännössä osoittautuu kuitenkin, että yhtälöitä elementtimenetelmällä ratkaistaessa paineenkaltainen muuttuja kannattaa säilyttää numeerisen stabiilisuuden vuoksi [2].

2.3 Yksinkertaistetut lineaariset yhtälöt

Aikaharmonisista yhtälöistä voidaan kirjoittaa lukuisia yksinkertaistuksia, jotka helpottavat yhtälöiden ratkaisua. Näitä on kattavasti käsitelty viitteessä [1]. Yleisessä tapauksessa ei yhtälöä kuitenkaan voida tästä sieventää, sillä etenkin resonanssipisteen ympäristössä tulokset ovat hyvin herkkiä kaikille parametreille [2].

Usein mikrosysteemien värähtelytaajuudet ovat niin alhaiset, että kaasun liike voidaan ratkaista ajasta riippumattomasta Stokesin yhtälöstä, joka saadan aikaharmonisista yhtälöistä rajalla $\omega, \kappa \to 0$.

$$\begin{aligned} \eta \Delta \vec{v} - \nabla p &= 0, \\ \nabla \cdot \vec{v} &= 0. \end{aligned} \tag{11}$$

Mainittakoon, että rajalla $\mu, \eta, \kappa \to 0$ aikaharmoniset Navierin ja Stokesin yhtälöt lähestyvät Helmholtzin yhtälöä. Puhtaasti akustista Helmholtzin yhtälöä ei kuitenkaan mikrosysteemien yhteydessä juurikaan käytetä, sillä viskoosien voimien merkitys pienissä dimensioissa korostuu.

2.4 Reynoldsin yhtälö

Hyvin kapeissa raoissa kolmiulotteisten yhtälöiden ratkaisu ei enää ole laskennalllisesti järkevää. Laskentaelementtien koon määrittelee virtauskanavan korkeus d. Jos muut dimensiot ovat tähän nähden suuria, nousee elementtien kokonaismäärä helposti hyvin suureksi. Tällöin on usein järkevää laskea virtaukset hyödyntäen laskennallisesti kevyempää dimensioreduktiolla saatavaa Reynoldsin yhtälöä.

Reynoldsin yhtälön johdossa oletetaan laminaarin virtauksen kanavassa olevan täysin kehittynyttä, jolloin liikeyhtälöstä saadaan nopeusprofiilille ainoastaan painegradientista riippuva analyyttinen ratkaisu. Yhdessä jatkuvuusyhtälön kanssa tämä johtaa yhtälöön

$$\frac{\partial}{\partial x}\left(\frac{\rho d^3}{12\eta}\frac{\partial p}{\partial x}\right) = \frac{\partial(\rho d(t))}{\partial t}, \tag{12}$$

joka isotermisen ideaalikaasun tilanyhtälön kanssa antaa Reynoldsin yhtälön paineelle

$$\nabla \cdot \left(\frac{pd^3}{12\eta}\nabla p\right) = \frac{\partial(pd)}{\partial t}. \tag{13}$$

Linearisoimalla yhtälö tasapainopisteen ympärillä saadaan

$$\frac{d_0^3}{12\eta}\Delta p - \frac{d_0}{p_0}\frac{\partial p}{\partial t} = \frac{\partial d}{\partial t}. \tag{14}$$

Jos yhtälön lähteenä toimiva reunan liike on harmonista, saadaan paineen amplitudille kompleksinen yhtälö

$$\frac{d_0^3}{12\eta}\Delta p - \mathrm{i}\frac{\omega d_0}{p_0}p = v_n, \tag{15}$$

missä v_n on normaalin suuntainen nopeuden amplitudi.

Reynoldsin yhtälön reunaehtoina käytetään yleensä annettua ulkoista painetta eli harmonisessa tapauksessa $p = 0$. Kun rako ei ole riittavan pieni muihin dimensioihin verrattuna, on tarpeen käyttää vuoehtoa, jonka avulla voidaan huomioida päätyefektien vaikutus

painejakaumaan

$$\frac{\partial p}{\partial n} = \frac{p}{L(\omega)}, \tag{16}$$

missä L on analyyttisiin tai numeerisiin laskuihin perustuva efektiivinen reunapidennys [4].

Kaasun harvuus Reynoldsin yhtälön tapauksessa voidaan huomioida efektiivisen viskositeetin avulla. Näin päästään huomattavasti korkeampiin Knudsenin lukuihin kuin on mahdollista yleisessä tapauksessa liukureunaehtoja käyttäen. Sovitus

$$\eta_{eff} = \frac{\eta}{1 + 9.638 K_n^{1.159}} \tag{17}$$

antaa vähintään viiden prosentin suhteellisen tarkkuuden välillä $0 < K_n < 880$ [5].

Reynoldsin yhtälöä on perinteisesti laskettu joko isotermisessä tai adiabaattisessa tilanteessa, jossa raon nopeusprofiili voidaan analyyttisesti helposti laskea. On kuitenkin mahdollista huomioida myös inertian ja lämpötilan vaikutus raon nopeusprofiilin muoton, jolloin päädytään jonkin verran mutkikkaampiin yhtälöihin [1].

2.5 Lumpattujen suureiden laskeminen tuloksista

Linearisoitujen yhtälöiden tulos voidaan esittää muodossa, joka ei riipu amplitudista. Sopiva lumpattava suure on esimerkiksi mekaaninen impedanssi eli liikettä vastustavan voiman ja kappaleen nopeuden suhde

$$Z = \frac{F}{v} = \frac{\int_\Gamma f_v \, dA}{v}, \tag{18}$$

missä f_v on liikkeen suuntaan vaikuttava pintavoima, joka Reynoldsin yhtälön tapauksessa on paine. Impedanssi on aikaharmonisten yhtälöiden tapauksessa kompleksinen suure, jonka reaaliosa kuvaa vaimennusta ja kompleksiosa jousivaikutusta. Stokesin yhtälöstä laskettu impedanssi on sen sijaan puhtaasti reaalinen.

Usein impedanssin sijasta käytetään myös admittanssia, joka on yksinkertaisesti impedanssin käänteisluku. Seuraavaa kappaletta silmällä pitäen määritellään akustinen ominaisadmittanssi Y, joka on tilavuusvirran ja voiman suhde eli

$$Y = \frac{V}{F} = \frac{vA}{\int_\Gamma f_v \, dA}. \tag{19}$$

2.6 Kaksitasoinen virtausanalyysi

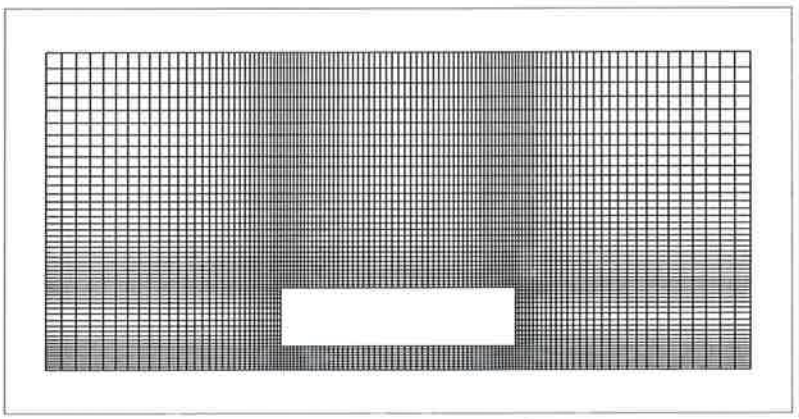
Usein mikrosysteemejä mallinnettaessa kohdataan rei'itettyjä rakenteita. Reikien tarkoitus on yleensä pienentää kaasun vaimennusta ja jousimaista vaikutusta, joka on erityisen voimakas kahden tasomaisen kappaleen lähestyessä toisiaan – eli juuri sellaisissa tilanteissa, joiden mallinnukseen Reynoldsin yhtälö olisi muutoin omiaan. Tällaisten tilanteiden mallinnuksessa voidaan hyödyntää kaksitasoista strategiaa [6, 7]. Lisäämällä Reynoldsin yhtälöön lisätermi, joka huomioi kaasun reikien kautta tapahtuvan liikkeen, voidaan yhtälö kirjoittaa muodossa

$$\frac{d_0^3}{12\eta}\Delta p - \mathrm{i}\frac{\omega d_0}{p_0} p - Y_h p = v_n \tag{20}$$

Laskemalla nyt yhden reiän admittanssi Y_h täyden dimension yhtälöistä ja käyttämällä sitä edelleen Reynoldsin yhtälön ratkaisussa saadaan mutkikkaat rei'itetyt rakenteet kustannustehokkaasti laskettua.

3 TULOKSIA

Yllä esitetyt yhtälöt on toteutettu elementtimenetelmään pohjautuvaan Elmer-ohjelmistoon [8]. Elmer on monipuolinen avoimen lähdekoodin ohjelmisto, johon voi joustavasti lisätä uusia fysikaalisia malleja uusien tarpeiden ilmaantuessa. Reynoldsin yhtälö voidaan ratkaista tavanomaisella Galerkinin menetelmällä, mutta muiden yhtälöiden tapauksessa on testifunktioavaruutta rikastettu kuplafunktioilla ratkaisun stabiloimiseksi [2]. Yhtälöitä on sovellettu lukuisten kaksi- ja kolmidimensioisten tapausten mallintamiseen. Tässä esitetään yhtälöiden soveltamista kahdessa tapauksessa, jotka edustavat laskennalliselta raskaudeltaan kahta ääripäätä. Laskennassa käytetyt kaasun materiaaliparametrit olivat p_0=101300 Pa, T_0=300 K, η=18.5 Pas, κ=0.025 W/mK, ρ_0=1.155 kg/m^3, C_p=1010 J/kgK, γ=1.4, σ=1.



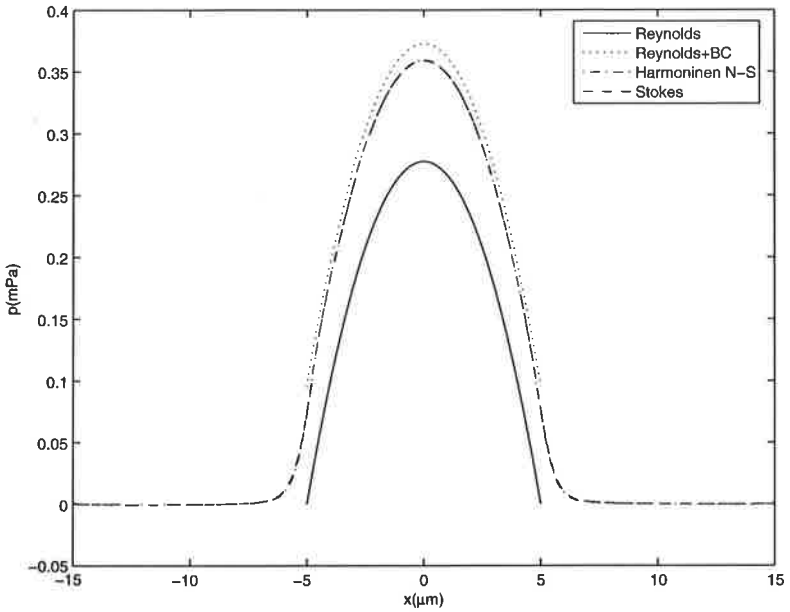
Kuva 1: 2D värähtelijän geometria ja laskentaverkko

3.1 Tason yläpuolella värähtelevä suorakaide

Ensimmäiseksi tutkimme eri mallien antamaa painejakaumaa kaksiulotteisessa tapauksessa, jossa 10 μm×2.5 μm kokoinen suorakaiteen muotoinen kappale värähtelee 1 Hz:n taajuudella ja 0.1 μm:n amplitudilla 1 μm:ä tason yläpuolella. Laskentaan käytetty verkko on esitetty kuvassa 1. Reynoldsin yhtälön tapauksessa laskentaverkko on tasainen 1D verkko. Reunaehtoina on käytetty tason ja värähtelijän reunoilla liukumattomuusreunaehtoja ja muilla reunoilla on oletettu paineen vaihteluiden häviävän.

Testitapauksen tulokset eri yhtälöillä laskettuna on esitetty kuvassa 2. Näin pienellä taajuudella kaasun jousivaikutus on olematon, jolloin aikaharmoniset yhtälöt antavat puhtaasti reaalisen painejakauman. Kuvassa aikaharmonisen Navierin ja Stokesin yhtälöiden ja puhtaan Stokesin yhtälön välillä ei täten ole silmin havaittavaa eroa. Sen sijaan Reynoldsin yhtälön ratkaisu reunaehdolla $p = 0$ poikkeaa merkittävästi edellisistä. Yhtälön (16) mukaisella reunaehdolla on Reynoldsin yhtälön paineen taso hyvin lähellä muiden yhtälöiden antamaa ratkaisua.

Aikaharmonisilla yhtälöillä voidaan suorittaa pyyhkäisy taajuuskaistan yli ja tutkia impedanssin reaali- ja imaginääriosan muutoksia. Kuvassa 3 on esitetty aikaharmonisilla Reynoldsin sekä Naviersin ja Stokesin yhtälöillä lasketut impedanssit välillä 1–10^8 Hz. Nähdään, että pienillä taajuuksilla kaasun jousivaikutus kasvaa suoraan verrannollisesti taajuuteen. Stokesin yhtälön ratkaisu olisi esimerkkitapauksella riittävä ainakin noin 10 kHz:n taajuu-



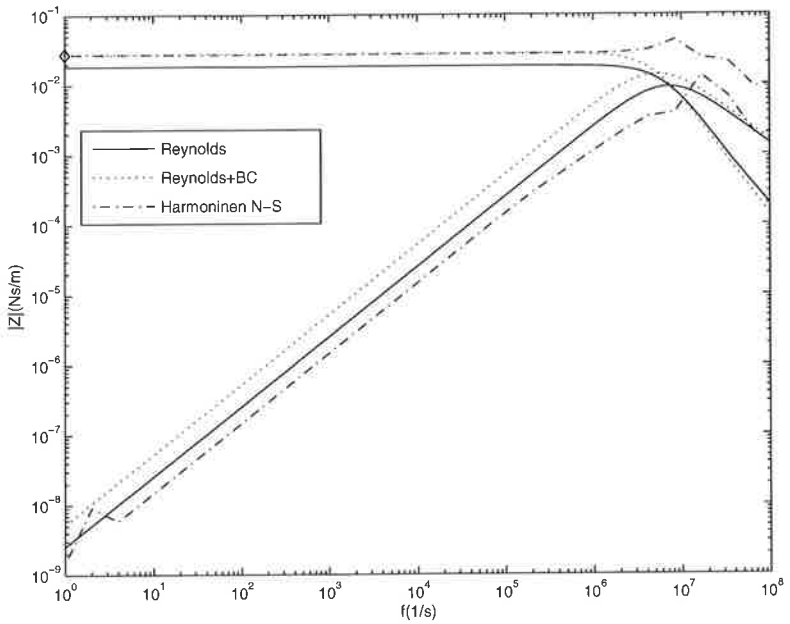
Kuva 2: 2D värähtelijän aiheuttamat painejakaumat pohjatasolla eri malleilla laskettuna

teen saakka. Mallinnetulla värähtelijällä jousivoima ohittaa vaimmennusvoiman vasta noin 10 MHz:n taajuudella. Tällä alueella myös aikaharmonisen Reynoldsin sekä Naviersin ja Stokesin yhtälöiden käyttäytyminen on hyvin erilaista, mikä johtuu muunmuassa lämpötilan sekä laskentaverkon tiheyden vaikutuksesta jälkimmäisessä tapauksessa.

3.2 Tason yläpuolella värähtelevä reikälevy

Toisena esimerkkitapauksena on tason yläpuolella värähtelevä neliömäinen reikälevy, jossa on 64 reikää säännöllisessä hilassa. Reikälevyn särmän pituus on 40 μm ja neliöllisten reikien särmä 1 μm, 2 μm tai 3 μm. Levyn paksuus ja etäisyys tasosta ovat molemmat 1 μm.

Esimerkkitapausta mallinnettiin Stokesin yhtälöllä sekä kaavan (20) modifioidulla Reynoldsin yhtälöllä, jossa reiät on huomioitu paikallisena admittanssina. Stokesin yhtälön tapauksessa laskentaan käytetty 3D verkko sisälsi 250 000 elementtiä, kun Reynoldsin yhtälön tapauksessa käytetty 2D verkko sisälsi ainoastaan 5000 elementtiä. Laskentaan kulunut aika oli Stokesin yhtälöllä noin puoli tuntia ja Reynoldsin yhtälön tapauksessa pari sekuntia (koneena 1 GHz:n Compaq AlphaServer). Aikaharmonisten Navierin ja Stokesin yhtälöiden ratkaisu ei tässä tapauksessa enää järkevässä ajassa onnistunut. Simulaation tuloksena saadut painekentät pohjalevylllä on esitetty kollaasina kuvassa 4. Painejakaumat ovat hyvin muodoltaan hyvin samanlaiset. Reynoldsin yhtälön antama kokonaisvaimennuksen suuruus on eri tapauksissa 5.5-13.5 % pienempi. Tämä johtuu osin siitä, että Reynoldsin yhtälössä ei huomioida reikälevyn reunoihin ja päälle kohdistuvaa voimaa. Molemmissa malleissa käytettiin kaasun harvuudesta johtuvia korjauksia. Tarkemmin simulaatiot on kuvattu viitteessä [7].



Kuva 3: Aikaharmonisella Reynoldsin sekä Naviersin ja Stokesin yhtälöillä lasketut impedanssit taajuuden funktiona. Reaaliosat ovat alussa vakiota ja imaginääriosa kasvaa lineaarisesti taajuuden funktiona.

4 POHDINNAT

Paperissa sovellettiin kolmea vaihtoehtoista yhtälöä kaasun vaimennuksen laskemiseksi mikrosysteemeissä. Matalilla taajuuksilla ja tasamaisissa tapauksissa yhtälöt antavat likimain yhtenevät tulokset. Mikrosysteemeissä tyypillisillä taajuuksilla jousivoimalla on merkitystä vain kapeissa raoissa, joilla on suuri aspektisuhde. Näiden mallintamiseen Reynoldsin yhtälö soveltuu erinomaisesti. Jos aspektisuhde on verrattain pieni, ei kaasun loukkuuntumista vastaavassa määrin ilmene, jolloin vaimennuksia voidaan hyvin mallintaa stationäärisellä Stokesin yhtälöllä. Edellisiä yhtälöitä voidaan myös yhdistää kaksitasoisessa analyysissä. Täydellisiä aikaharmonisia Navierin ja Stokesin yhtälöitä tarvitsee pääosin hyödyntää vain radiotaajuualueella sekä kartoitettaessa tai laajennettaessa yksinkertaisempien mallien pätevyysalueiden rajoja. Yksinkertaistettujen mallien käyttö saattaa pudottaa laskenta-aikaa, jopa useita kertaluokkia. Mikosysteemien vaimennusominaisuuksien tehokas simulointi edellyttääkin kuhunkin tarkoitukseen sopivimpien yhtälöiden käyttöä.

VIITTEET

[1] W.M. Beltman. Viscothermal wave propagation including acousto-elastic interaction, Part I: Theory. *Journal of Sound and Vibration*, **227**, 555–586, 1999.

[2] M. Malinen, M. Lyly, P. Råback, A. Kärkkäinen, and L. Kärkkäinen. *A finite element method for the modeling of thermo-visoucs effects in acoustics*, ECCOMAS 2004, Jyväskylä, Finland, 24-28 June 2004.

[3] G.E. Karniadakis and A. Beskok. *Micro Flows, Fundamentals and Simulation.* Springer, 2002.

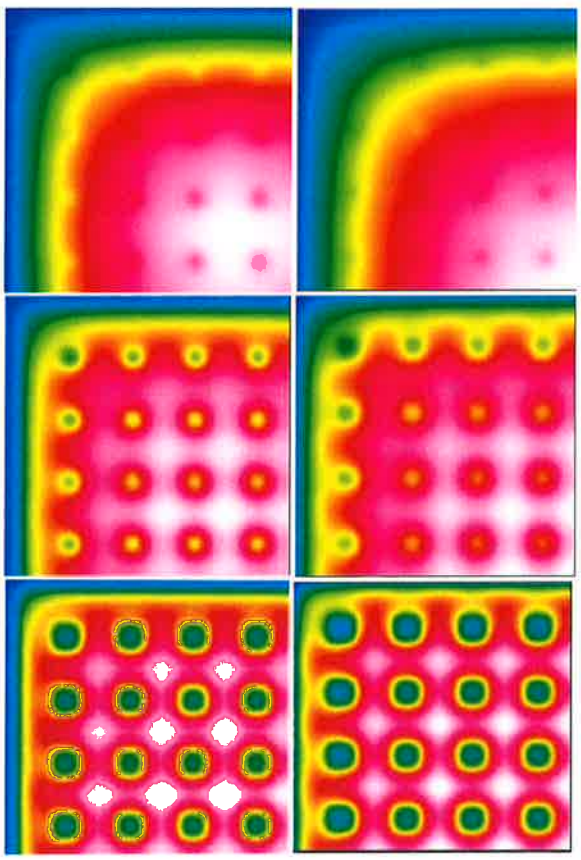
[4] T. Veijola, A. Pursula and P. Råback. *Extending the validity of existing squeezed-film damper models with elongations of surface dimensions*, Journal of Micromechanics and Microengineering, **15**, 1624–1636, 2005.

[5] T. Veijola, H. Kuisma, J. Lahdenperä and T. Ryhänen. *Equivalent-circuit model of the squeezed gas film in a silicon accelerometer*, Sensors and Actuators A, **48**, 239–248, 1995.

[6] P. Råback, A. Pursula, V. Junttila, and T. Veijola. *Hierarchial finite element simulation of perforated plates with arbitrary hole geometries*, NANOTECH 2003, San Francisco, 23-27 February 2003, **1**, 194–197.

[7] T. Veijola and P. Råback. *A Method for solving arbitrary MEMS perforation problems with rare gas effects*, NANOTECH 2005, Anaheim, May 8-12, 2005, **3**, 561–564.

[8] *Elmer finite element software homepage*, http://www.csc.fi/elmer



Kuva 4: Painejakaumat laskettuna 3D Stokesin yhtälöstä (vasemmalla) sekä 2D Reynoldsin yhtälöstä (oikealla). Reikien halkaisijat ovat 1 μm (ylhäällä), 2 μm (keskellä) sekä 3 μm (alhaalla). Vain symmetrinen neljännes levystä on simuloitu.

EFFECT OF PARALLEL SYSTEM ROTATION ON HEAT TRANSFER IN LAMINAR PIPE FLOW

A. HELLSTEN

Helsinki University of Technology, Laboratory of Aerodynamics
P.O.Box 4400 FIN-02015 TKK
e-mail: antti.hellsten@tkk.fi

ABSTRACT

Laminar flow and heat transfer in straight pipe with circular cross section rotating about an offset parallel axis is studied computationally. There are two kinds of secondary flow mechanisms in this flow situation: a Coriolis driven and a centrifugal-buoyancy driven secondary flow. The secondary flows may have a significant influence on the heat-transfer rate and the pressure loss. In this study, some results for constant wall temperature cases with moderate rotational Reynolds number and Grashof number are first presented. However, the main focus is in situations where the axially constant wall temperature varies circumferentially. Such wall temperature difference is shown to have a stabilizing effect on the buoyancy-driven secondary flow if the inner side is hotter than the outer side, and a destabilizing effect in the opposite case. The stabilizing effect is not strong enough to dissipate the secondary flow even though the wall-temperature difference is of the same magnitude as the difference between the inflow temperature and the average wall temperature. According to the computational results, the secondary flow pattern remains almost similar to the situation with circumferentially constant wall temperature.

1 INTRODUCTION

Pipe flows in systems rotating about the axis parallel but offset to the pipe axis are found in many rotating machines. The flow geometry is illustrated in Fig. 1. This kind of pipe flow may be used for cooling or heating the rotating machines. The system rotation and temperature differences give rise to secondary flow phenomena which influence the heat-transfer rate and also the pressure loss. There are only few studies found in the literature on these kinds of problems [1–4]. Soong and Yan have computationally studied laminar flow in rectangular ducts with constant wall temperature and also constant wall heat flux [3].

Practical cooling or heating flows are usually turbulent, and turbulence is known to be influenced by the system rotation and buoyancy. These interactions render the problem much more complicated. The basic motivation behind this study is in a turbulent heating-flow application found in heated cylinders in paper-machine calanders. However, because the turbulent problem is excessively complicated, a corresponding laminar flow is studied first to gain understanding about the secondary-flow phenomena in these kinds of flows. A special feature in the cylinder heating problem is that the wall-temperature may be significantly lower on the outer side of the pipe as the overall heat flux is directed radially outwards. The difference $\Delta T_c = T_{w\,\max} - \overline{T}_w$ is comparable with the difference between

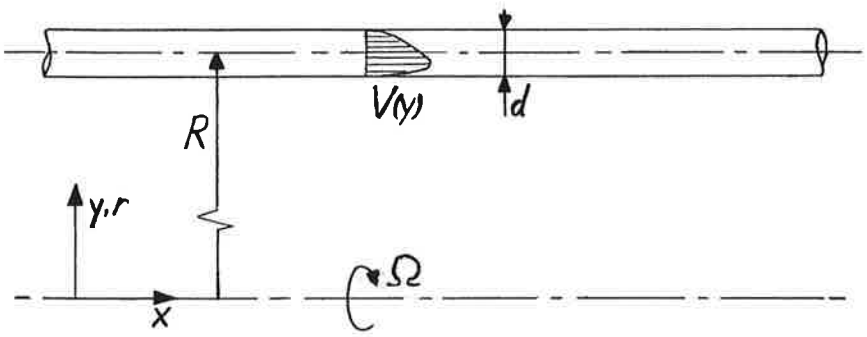


Figure 1: A schematic illustration about the flow geometry.

the inlet temperature T_i and the average wall temperature $\overline{T}_w$, $\Delta T = T_i - \overline{T}_w$. To the author's knowledge, such situations have not been studied so far. Prior to this study it was not known whether the wall temperature difference changes the secondary flow system remarkably or not.

In this work, some computational results for circular tube with constant wall temperature are first presented, but the main focus is in situations where the axially constant wall temperature varies circumferentially as described above. It will be shown that non-uniform circumferential wall temperature distributions with cooler outer side have a stabilizing effect on the buoyancy-driven secondary flow, and distributions with hotter outer side have a destabilizing effect. However, the stabilizing effect is not strong enough to dissipate the secondary flow in the studied situations spanning up to $A = \Delta T_c/\Delta T = 1$. According to the computational results, the secondary flow pattern remains almost similar to the situation with circumferentially constant wall temperature and the secondary-flow energy is only moderately reduced.

2 EQUATIONS AND SCALING

The generation mechanisms of the secondary flows can be identified from the transverse momentum equations written in rotating coordinate system. The equations are scaled such that the axial velocity u is scaled by the bulk velocity U as $\tilde{u} = u/U$, and the transverse velocity components are scaled by a viscous velocity scale ν/d as $v^* = vd/\nu$ and $w^* = wd/\nu$, where ν is the kinematic viscosity and d is the pipe diameter. The nondimensional axial coordinate $\tilde{x} = x/(Re\, d)$ while the nondimensional transverse coordinates are defined by $y^* = y/d$ and $z^* = z/d$. Furthermore, the pressure field is split in three parts as

$$p(x,y,z) = \rho U^2 \tilde{p}(x) + p_\Omega(y,z) + \frac{\rho\nu^2}{d^2} p^*(y,z) \tag{1}$$

where $\rho U^2 \tilde{p}$ is the axial pressure field which drives the flow, p_Ω is the mean pressure field generated by the centrifugal acceleration, and finally the remaining $\rho\nu^2 p^*/d^2$ is related to the secondary flows. The mean centrifugal acceleration and p_Ω can be removed from the equations since they balance each other exactly. This scaling and splitting of pressure leads to non-dimensional equations where the bulk Reynolds number appears only in axial diffusion terms, which can be dropped from these equations similarly as from the boundary-layer equations. Thus, we obtain the parabolized non-dimensional momentum equations

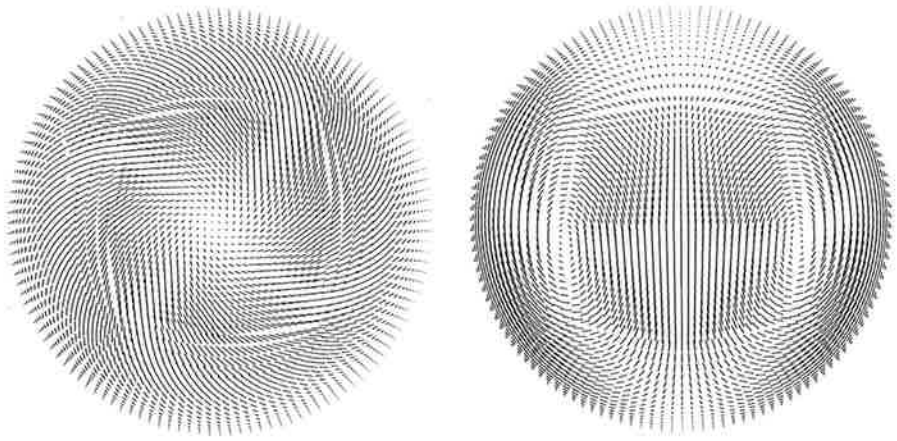


Figure 2: Typical secondary flow pattern near the pipe entrance (left) and far downstream (right)

which involve only two nondimensional groups

$$\begin{aligned}
\frac{\partial \tilde{u}^2}{\partial \tilde{x}} + \frac{\partial \tilde{u}v^*}{\partial y^*} + \frac{\partial \tilde{u}w^*}{\partial z^*} &= -\frac{\partial \tilde{p}}{\partial \tilde{x}} + \frac{\partial^2 \tilde{u}}{\partial y^{*2}} + \frac{\partial^2 \tilde{u}}{\partial z^{*2}} \\
\frac{\partial \tilde{u}v^*}{\partial \tilde{x}} + \frac{\partial v^{*2}}{\partial y^*} + \frac{\partial v^*w^*}{\partial z^*} &= -\frac{\partial p^*}{\partial y^*} + \frac{\partial^2 v^*}{\partial y^{*2}} + \frac{\partial^2 v^*}{\partial z^{*2}} + 2Jw^* + Gr\frac{y}{R} \\
\frac{\partial \tilde{u}w^*}{\partial \tilde{x}} + \frac{\partial v^*w^*}{\partial y^*} + \frac{\partial w^{*2}}{\partial z^*} &= -\frac{\partial p^*}{\partial z^*} + \frac{\partial^2 w^*}{\partial y^{*2}} + \frac{\partial^2 w^*}{\partial z^{*2}} - 2Jv^* + Gr\frac{z}{R}
\end{aligned}$$

The last two terms in the transverse equations are the Coriolis and centrifugal terms which depend on the rotational Reynolds number J and the Grashof number Gr, respectively. These nondimensional groups are defined as follows

$$J = \frac{\Omega d^2}{\nu}, \qquad Gr = \frac{\beta \Delta T \Omega^2 R d^3}{\nu^2}$$

Ω is the system rotation rate, β is the coefficient of thermal expansion, and R is the offset between the rotation axis and the pipe axis. Note that the centrifugal terms depend on y/R and z/R instead of y^* and z^*. This means that any change in R only changes the value of Gr while all the other factors and terms in the equations remain unchanged. It is assumed that the flow does not depend on R/d at least if it is not very small. In this study, R/d is always kept larger than 4.

The Coriolis acceleration produces secondary flow mainly in the entrance region of the pipe where the boundary layer grows. Initially, the resulting secondary flow pattern is a single spiral vortex, see Fig. 2 (left). As the flow develops and the changes in axial direction become small, the Coriolis-induced secondary flow starts to disappear. This is because the Coriolis-related production of vorticity depends on the derivatives in the axial direction [3]. However, Soong and Yan have shown for rectangular cross sections that the Coriolis acceleration may have significant effects quite far downstream if $J > 200$ [3].

The centrifugal acceleration causes buoyancy in the non-uniform temperature field. This produces secondary-flow pattern involving two counter-rotating vortices, see Fig. 2 (right). This secondary-flow system has strong potential to enhance the heat transfer rate as long as there are significant temperature differences in the pipe cross section.

3 COMPUTATIONAL METHOD

The computations were carried out using the Navier-Stokes solver called FINFLO [5,6]. The FINFLO code, developed at Helsinki University of Technology, is based on the finite-volume approach and utilizes structured multi-block grids. The solution method is an implicit time integration. A multigrid cycling is used to accelerate convergence. The inviscid fluxes are evaluated using Roe's flux-difference splitting with a MUSCL-type discretization, while the central-differencing scheme is used in the calculation of the viscous fluxes. No simplifications, such as the thin-layer approximation, are subjected to the viscous terms.

The solution method used in the FINFLO version utilized in this study is density-based and therefore limited to flows with Mach number of about 0.2 and larger. The essentially incompressible flow considered in this study was emulated by setting the Mach number to 0.2. The fluid is calorically perfect gas with the Prandtl number $Pr = 0.72$. The Mach-number limitation prevented simulations with higher values of Gr and J. A pressure-based algorithm, e.g. an artificial compressibility method is needed in the next phase of the study considering the real turbulent flow in the heating pipes in calander cylinders featuring much higher values of Gr and J.

The computations were made in the rotating coordinate system implemented in the FINFLO solver. The computational domain extended $115d$ downstream from the pipe entry. R/d had different values between 4.3 and 17.3 in different cases. Constant momentum, temperature and other quantities were specified at the inflow boundary and the outflow boundary was approximated assuming fully developed flow. The usual no-slip conditions with given temperature distributions were specified on the pipe walls.

The computational grid is constructed in special fashion to avoid singularity on the axis. The cross section is gridded using a nearly rectangular 32×32 core block surrounded by four 16×32 blocks, see Fig. 3. The axial direction is divided into 128 control volumes, and the length of the first one at the entry is $0.38d$ and the last one at the outlet is $1.75d$. In total, the grid consists of 393,216 control volumes.

The influence of the truncation error on the solutions was studied by comparing the pressure losses computed with the present grid and two subsequently coarser grids obtained by removing every second grid line in each direction. The differences between the finest and second finest grids were about 1% whereas the corresponding difference between the second and third grids were from 4% to 5%. Based on these observations, it is believed that the grid dependence is negligibly small.

4 RESULTS

4.1 Constant wall temperature

The influence of the secondary flows, especially the buoyancy-driven secondary flow is clearly seen in the computational results. The circumferentially averaged value of the Nusselt number is studied with different values of Gr keeping J at a relatively small constant value of 113. The Nusselt number is defined as

$$Nu = \frac{q_w d}{k_w (T_b - \overline{T}_w)} \tag{2}$$

with the bulk temperature being defined following Soong and Yan [3] as

$$T_b = T_i - \frac{\pi \int_S \rho u (T_i - T)\, \mathrm{d}S}{4 \int_S \rho u \, \mathrm{d}S} \tag{3}$$

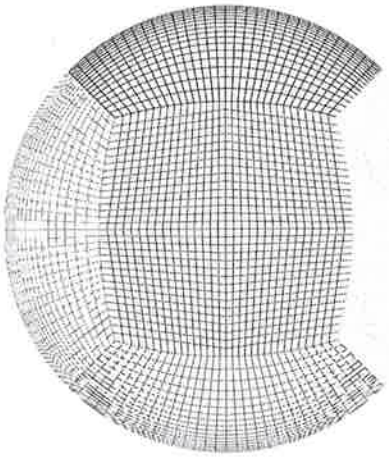


Figure 3: The computational grid over the cross section of the pipe.

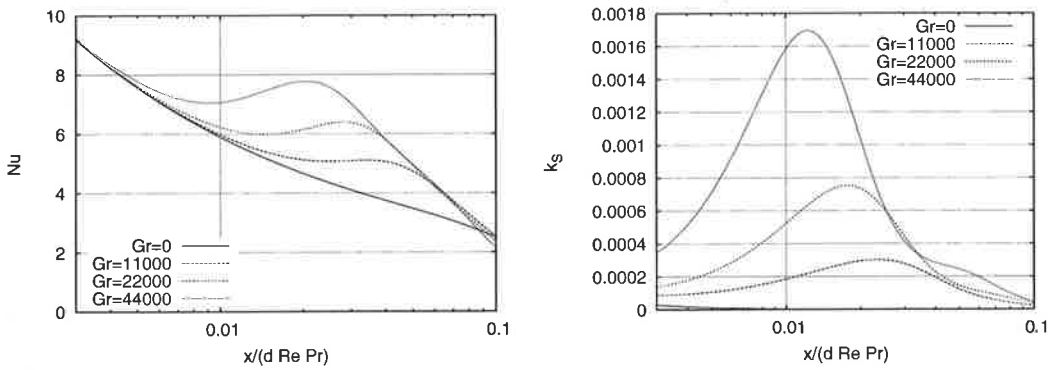


Figure 4: Circumferentially averaged Nusselt number (left) and the relative secondary-flow energy (right) with constant wall temperature and different values of Gr. $J = 113$.

Here q_w and k_w are the circumferentially averaged heat flux and the coefficient of thermal conductivity at the wall. The Nusselt-number distributions are shown in Fig. 4 (left) as a function of $x/(dRePr) = \tilde{x}/Pr$. The Nusselt number clearly increases with increasing Gr. Fig. 4 (right) shows the relative kinetic energy of the transverse motion, i.e. the secondary-flow energy

$$k_S = \frac{\int_S \rho(v^2 + w^2)\,\mathrm{d}S}{\int_S \rho(u^2 + v^2 + w^2)\,\mathrm{d}S} \tag{4}$$

integrated over each cross section. The maximum secondary-flow energy quite rapidly increases with increasing Gr. The secondary motion becomes weaker downstream as the temperature differences decrease since constant wall temperature was specified.

4.2 Circumferentially variable wall temperature

Three simulations with circumferentially variable wall temperature have been performed with $Gr = 44,000$ and $J = 113$. The wall temperature is defined as

$$T(\varphi) = \overline{T}_w + A\Delta T \cos\varphi \tag{5}$$

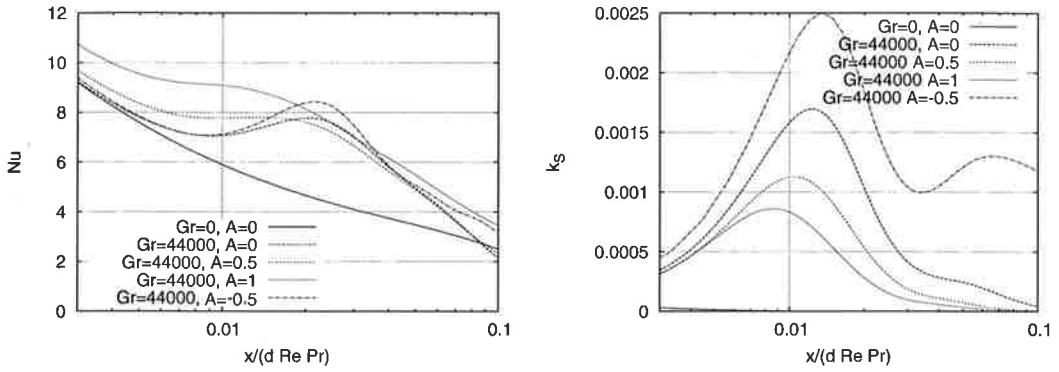


Figure 5: Circumferentially averaged Nusselt number (left) and the relative secondary-flow energy (right) with constant and variable wall temperature $Gr = 44,000$. $J = 113$.

with A being a constant parameter. The angular coordinate φ has its origin in the inner intersection of the main radius and the pipe wall. Thus the maximum wall temperature occurs at the innermost point of the wall if $A > 0$. The wall temperature is independent of the axial coordinate. This kind of circumferential variation in the wall temperature has a stabilizing effect on the buoyancy-driven secondary flow when heating flows are considered. In cases of cooling flows, the opposite situation takes place. The aim of these simulations is to study how strong the stabilizing influence is.

Computations have been performed for two stabilizing cases, $A = 0.5$ and $A = 1$. In the latter case, the highest wall-temperature equals to the inflow temperature. The Nusselt number and the relative secondary-flow energy are given in Fig. 5. Judging from the secondary-flow energy, the stabilizing effect is not strong enough to dissipate the secondary motion. The vortices are somewhat weaker, but they still exist. The increased temperature difference on the outer side of the pipe enhance the heat flux more than the weakening of the secondary flow inhibits it. Therefore the Nusselt number does not decrease with increasing A, actually it slightly increases. This means that the buoyancy-driven secondary flow remains an important factor in these kinds of flows despite of the stabilizing circumferential wall-temperature distribution.

Wall temperature distributions with hotter outer side have a destabilizing effect. Only one case, $A = -0.5$ is computed. The increased secondary-flow energy can be seen in Fig. 5 (right). There is a clear qualitative difference to the cases with $A > 0$ in the downstream region $\tilde{x}/Pr > 0.03$. The secondary flow energy has a second maximum, which is not seen in the results with $A > 0$.

The effects of the stabilizing and destabilizing circumferential wall-temperature distributions on the structure of the secondary flow pattern is also of interest. These patterns in cases of $A = 0$, $A = 0.5$, $A = 1$ and $A = -0.5$ are shown in Fig. 6 near the location of maximum secondary-flow energy $\tilde{x}/Pr = 0.015$, and in Fig. 7 in more downstream location $\tilde{x}/Pr = 0.05$ near the second maximum found in the case with $A = -0.5$. The variations in the vortex strength are visible in the plots, but the structure of the vortices are quite similar in each case. There is slightly more asymmetry in $\tilde{x}/Pr = 0.015$ than in the more downstream location. This is explained by the influence of the Coriolis-driven secondary motion which becomes weaker as the flow develops.

5 CONCLUSIONS

Laminar flow and heat transfer in straight pipe with circular cross section rotating about an offset parallel axis was studied computationally. This study is the first step of a larger study concerning the modelling of turbulent flow in this particular geometry. The motivation behind the study is the need to simulate heating flows in heated cylinders of paper-machine calanders. The predictive capability and reliability of turbulence models in the cylinder-heating flow problem will be assessed in the future work.

Results for circular tube with both constant and circumferentially varying but axially constant wall temperature were presented. The Grashof number Gr ranged from zero up to 44,000 and the rotational Reynolds number J was 113 in rotating cases. It was shown that non-uniform circumferential wall temperature distributions with cooler outer side have a stabilizing effect on the buoyancy-driven secondary flow, and distributions with hotter outer side have a destabilizing effect. The stabilizing effect is not strong enough to dissipate the secondary flow in the studied situations spanning up to $A = \Delta T_c / \Delta T = 1$. In case of cooler outer side, the secondary flow pattern remains almost similar to the situation with circumferentially constant wall temperature and the secondary-flow energy is only moderately reduced. Respectively, in case of hotter outer side, the secondary flow energy is somewhat enhanced, and in contrast to the stabilizing case, the secondary-flow energy has two maxima. However, qualitatively the secondary flow pattern remains quite similar to the constant wall-temperature case.

6 ACKNOWLEDGEMENTS

This work has been carried out in the TURBU research project funded by the Finnish Technology Agency (Tekes), Metso Paper Oy, Numerola Oy, Andritz Oy, Finflo Oy, Fortum Nuclear Services Oy, and Patria Aviation Oy. The support from Tekes and all the companies is gratefully acknowledged. The author also wants to thank Lic. Tech. Reijo Pietikäinen of Metso Paper Oy for the interesting research problem and for fruitful discussions.

REFERENCES

[1] Y. Mori and W. Nakayama. Forced convective heat transfer in a straight pipe rotating around a parallel axis (1st report, laminar region). *International Journal of Heat and Mass Transfer*, 10:1179–1194, 1967.

[2] W. D. Morris and J. L. Woods. Heat transfer in the entrance region of tubes that rotate about a parallel axis. *Journal of Mechanical Engineering Science*, 20(6):319–325, 1978.

[3] C. Y. Soong and W. M. Yan. Development of secondary flow and convective heat transfer in isothermal/iso-flux rectangular ducts rotating about a parallel axis. *International Journal of Heat and Mass Transfer*, 42:497–510, 1999.

[4] W. M. Yan. Mixed convection heat and mass transfer in rectangular ducts rotating about a parallel axis. *International Journal of Heat and Mass Transfer*, 42:2955–2965, 1999.

[5] T. Siikonen. An application of Roe's flux-difference splitting for the $k - \epsilon$ turbulence model. *International Journal for Numerical Methods in Fluids*, 21(11):1017–1039, 1995.

[6] P. Kaurinkoski and A. Hellsten. FINFLO: the parallel multi-block flow solver. Report A-17, Helsinki University of Technology, Laboratory of Aerodynamics, Espoo, Finland, 1998. ISBN 951–22–3940–X.

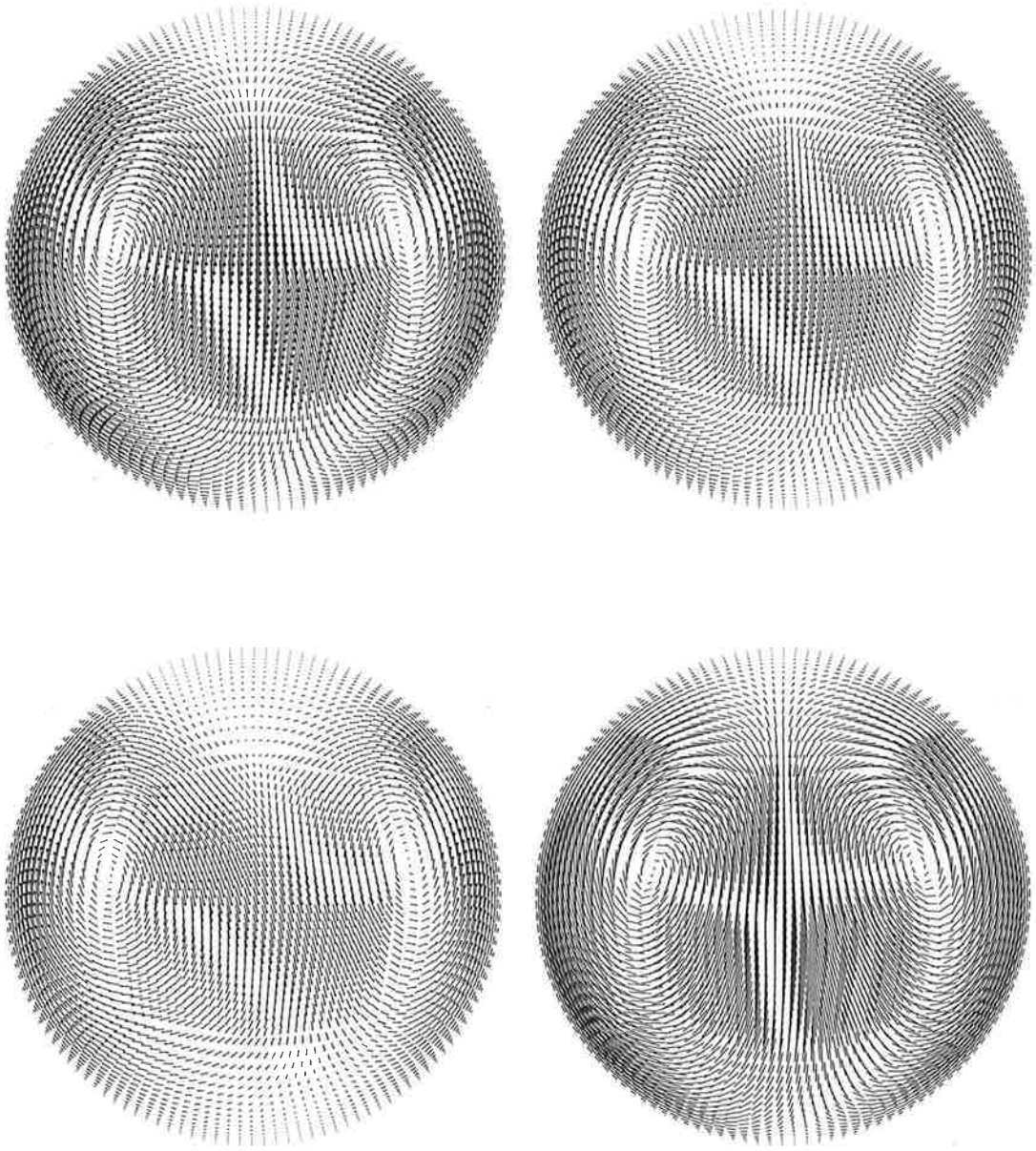


Figure 6: Secondary flow pattern at $\tilde{x}/Pr = 0.015$ in cases of $A = 0$ (upper left), $A = 0.5$ (upper right), $A = 1$ (lower left) and $A = -0.5$ (lower right).

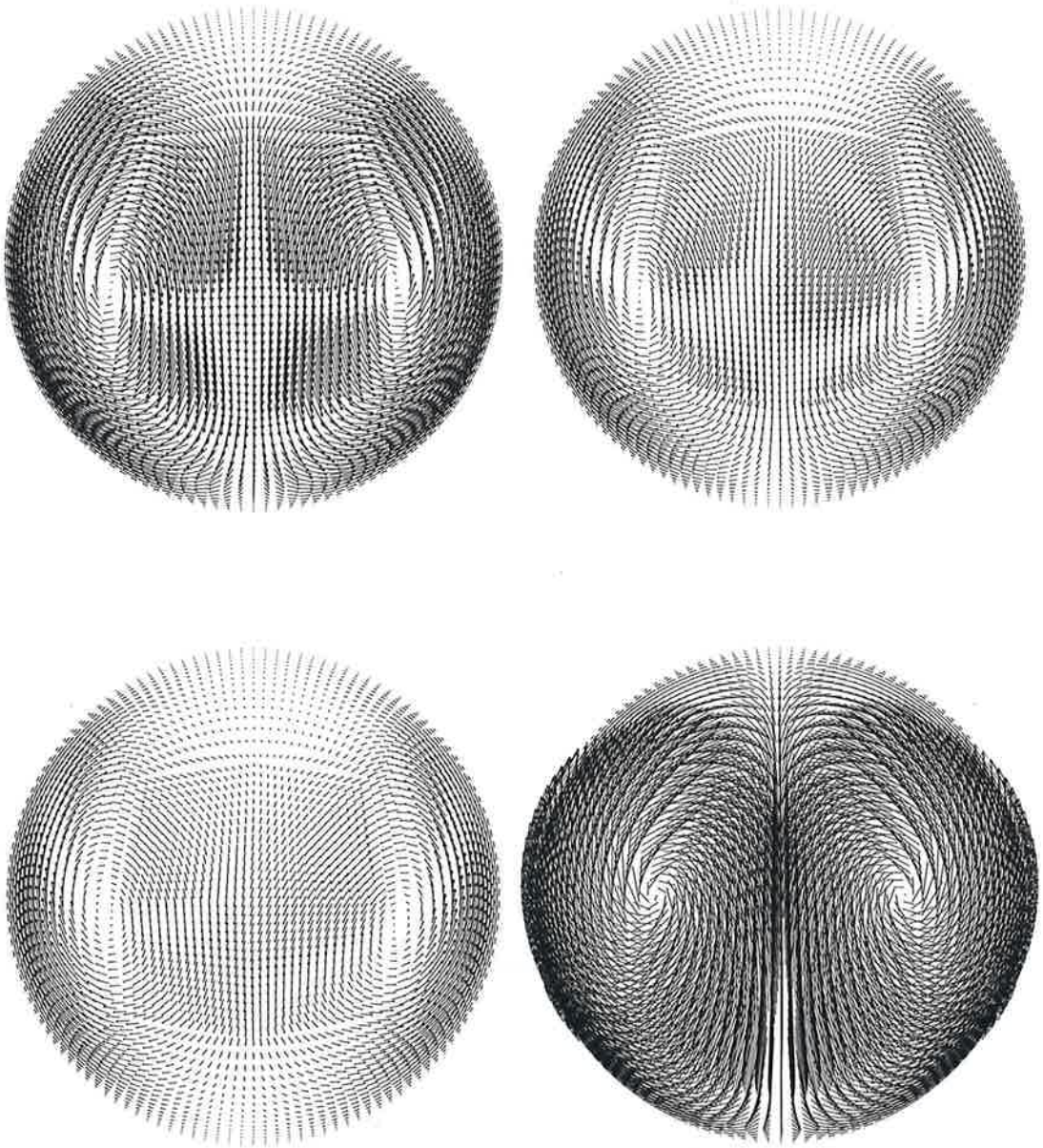


Figure 7: Secondary flow pattern at $\tilde{x}/Pr = 0.05$ in the same cases as shown in Fig. 6. Note that the scaling factor of the vector arrows is here four times the factor used in Fig. 6

VOLUME FLOW RATE OSCILLATION IN TAPERED CHANNELS

T. KARVINEN
Institute of Energy and Process Engineering
Tampere University of Technology
P.O. Box 589
33101 Tampere
timo.karvinen@tut.fi

ABSTRACT

Volume flow rate oscillation in tapered channels of different shapes was studied and compared. Channels were modeled using the equations of acoustics and the impedance method of fluid transients. The boundary conditions are the known pressure amplitude upstream and zero pressure at the channel outlet. It was found that these two methods yield exactly the same results. In the acoustical approach, analytical solutions can be found for certain channel shapes for the pressure and volume flow rate oscillation. An optimization algorithm was employed to obtain an optimal shape by minimizing the volume flow rate oscillation at the channel outlet for a single frequency.

1. INTRODUCTION

Converging pipes or channels are important applications in many fields of industry. The purpose of this study is to examine the volume flow rate oscillation in tapered channels used in paper making machines.

A pump or fan produces harmonic pressure oscillation and at the same time volume flow rate oscillation. This oscillation is driven through the pipeline and finally it travels through the tapered channel where the flow is accelerated. The volume flow rate variation may be detrimental to the quality of the end product, in the case of a paper making machine the machine direction basis weight variation of paper. In our problem the pressure oscillation amplitude and the frequency are known at the upstream end of the channel and the channel discharges into the atmosphere. The frequency range of interest is from few Hertz to 100 Hz. The problem is analogous to the string vibration, in which the string is driven from the other end with the transverse displacement oscillating harmonically and another end is fixed.

The taper angle of channels is small and therefore reflections from the channel wall can be omitted [1] and one-dimensional equations can be used. The channel walls are assumed to be rigid. The flowing fluid is water. Two mathematical approaches are possible. Usually wave propagation problems concerning liquids are solved using the theory of fluid transients, and when the flowing media is gas the equations of acoustics are employed. In this study both approaches are used and the results are compared. These two methods can be compared because the pipes and channels

under examination are relatively short and the friction can be neglected in the equations of fluid transients.

2. GOVERNING EQUATIONS

2.1 Acoustical approach

The pressure pulsation in a channel with the cross-sectional area A a function of the x-coordinate is governed by the modified wave equation [1]

$$\frac{\partial^2 p}{\partial^2 x} - \frac{1}{a^2}\frac{\partial^2 p}{\partial t^2} + \frac{1}{A}\frac{\partial A}{\partial x}\frac{\partial p}{\partial x} = 0 \tag{1}$$

in which p is the pressure, x the x-coordinate, a the wave speed and t the time. If the discussion is restricted to harmonic cases, $\partial / \partial t$ can be replaced with $i\omega$ yielding the modified Helmholtz equation

$$\frac{\partial^2 p}{\partial^2 x} + k^2 p + \frac{1}{A}\frac{\partial A}{\partial x}\frac{\partial p}{\partial x} = 0 \tag{2}$$

in which $k = \omega / a$ is the wave number and $\omega = 2\pi f$ the angular frequency (f frequency). The above equation is also known as the Webster's equation.

2.1.1 Linear change of channel height

The equation for the cross-section area of a linearly changing channel height (constant width) is (see Fig. 1)

$$A(x) = y(0) + bx = b(m + x) \tag{3}$$

in which $b = \left[y(L) - y(0)\right] / L$, $m = y(0) / b$ with $y(0)$ the channel height at the upstream end, y(L) the height at the downstream end and L the channel length. When Eq. (3) is substituted to Eq. (2), it yields

$$\frac{\partial^2 p}{\partial^2 x} + k^2 p + \frac{1}{m + x}\frac{\partial p}{\partial x} = 0 \tag{4}$$

The boundary conditions are the known pressure amplitude P at the upstream end and the zero pressure at the downstream end representing the discharge to the atmosphere.

$$\begin{cases} p(0) = P \\ p(L) = 0 \end{cases} \tag{5}$$

The final solution with the boundary conditions (5) is

$$p(x) = -P \frac{J_0[k(m+x)]\,Y_0[k(m+L)] - J_0[k(m+L)]\,Y_0[k(m+x)]}{J_0[k(m+L)]\,Y_0(km) - J_0(km)\,Y_0[k(m+L)]} \tag{6}$$

where J_0 and Y_0 are the Bessel functions of first and second kind.

The particle velocity is obtained from [1]

$$u = \frac{i}{\rho\omega}\frac{\partial p}{\partial x} \tag{7}$$

where ρ is the fluid density. The volume flow rate oscillation is obtained from

$$Q = uA(x) \tag{8}$$

2.1.2 Exponential change of channel height

If the shape of the channel is exponential, the equation of cross-sectional area is $A(x) = y(0)e^{x/h}$, $h = L/\ln[y(L)/y(0)]$. The Webster's equation (2) becomes now

$$\frac{\partial^2 p}{\partial^2 x} + k^2 p + \frac{1}{h}\frac{\partial p}{\partial x} = 0 \tag{9}$$

The solution of Eq. (9) with the boundary conditions (5) is

$$p(x) = P \frac{e^{\frac{1}{2h}\left[-L-L(1-4h^2k^2)^{\frac{1}{2}}-x-x(1-4h^2k^2)^{\frac{1}{2}}\right]} - e^{\frac{1}{2h}\left[L-L(1-4h^2k^2)^{\frac{1}{2}}+x+x(1-4h^2k^2)^{\frac{1}{2}}\right]}}{-e^{\frac{1}{2h}L\left[-1+(1-4h^2k^2)^{\frac{1}{2}}\right]} + e^{\frac{1}{2h}L\left[1+(1-4h^2k^2)^{\frac{1}{2}}\right]}} \tag{10}$$

2.1.3 Parabolic and 3rd degree curve of channel height

If the channel shape is parabolic or a 3rd degree curve, the equation of cross-sectional area is of the form $A(x) = c_1 x^3 + c_2 x^2 + c_3 x + c_4$. In this case an analytical solution of Eq. (2) cannot be found [2] and the equation must be solved numerically. The numerical solution was obtained with the Matlab's boundary value problem solver.

2.2 Fluid transients approach

The simplified equations of fluid transients in a pipeline are the equations of momentum and continuity [3]

$$\frac{\partial H}{\partial x} + \frac{1}{gA}\frac{\partial Q}{\partial t} + \frac{fQ^n}{2gDA^n} = 0 \tag{11}$$

$$\frac{\partial Q}{\partial x} + \frac{gA}{a^2}\frac{\partial H}{\partial t} = 0 \tag{12}$$

in which H ($p = \rho g H$) denotes the pressure head, g the gravitational acceleration, Q the volume flow rate, f the friction factor, D the pipe diameter and n is an integer depending on the type of flow. The friction term in the equation of motion can be neglected, if the pipeline is short. If partial derivatives with respect to x and t are taken of Eq. (13) and (14) and the equations are combined and then Q is eliminated, it results in the wave equation

$$\frac{\partial^2 H}{\partial x^2} - \frac{1}{a^2}\frac{\partial^2 H}{\partial t^2} = 0 \tag{13}$$

which proves that the one-dimensional wave equation and the equations for oscillatory fluid transients are the same in the absence of friction for a pipe of constant cross-section.

When the transients are steady harmonic, the solution for Eq. (11) and (12) can be obtained with the technique of separation of variables [3]. The complex head and discharge are

$$H(x) = H_U \cosh \gamma x - Z_C Q_U \sinh \gamma x \tag{14}$$

$$Q(x) = \frac{H_U}{Z_C} \sinh \gamma x + Q_U \cosh \gamma x \tag{15}$$

with $\gamma = \sqrt{-\omega^2 / a^2}$ a complex constant, $Z_C = -i\gamma / C\omega$ the characteristic impedance, $C = gA / a^2$ the capacitance and the subscripts U and D denote upstream and downstream ends of the pipe, respectively.

In order to solve the pressure and volume flow rate oscillation in a channel with the cross-sectional area a function of x, the channel is divided into sections of constant cross-section shown in Fig. 1. At the junction of two sections the pressures on both sides of the junction are equal and the continuity condition holds.

The pressure and volume flow rate oscillation can be determined at each point of the channel with the use of Eq. (14), (15) and the boundary conditions. The equations for one channel section are

$$\begin{aligned} H_{U_{i+1}} &= H_{D_{i+1}} \cosh \gamma_{i+1} l + Q_{D_{i+1}} Z_{C_{i+1}} \sinh \gamma_{i+1} l \\ Q_{U_{i+1}} &= \frac{H_{D_{i+1}}}{Z_{C_{i+1}}} \sinh \gamma_{i+1} l + Q_{D_{i+1}} \cosh \gamma_{i+1} l \\ H_{D_i} &= H_{U_{i+1}} \\ Q_{D_i} &= Q_{U_{i+1}} \end{aligned} \tag{16}$$

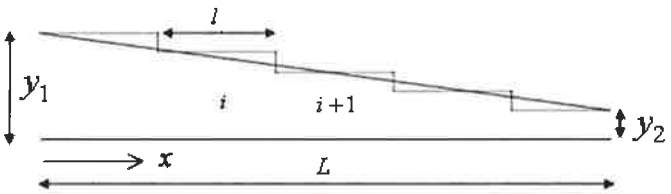


Fig. 1. Dividing linearly tapered channel into sections of constant cross-section.

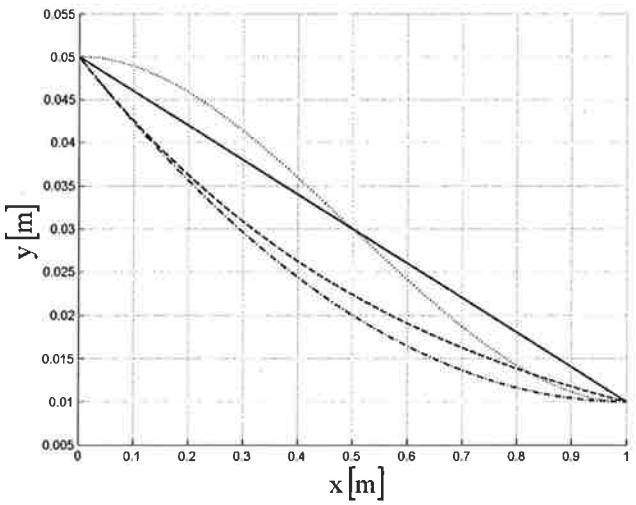


Fig. 2. Channel height of different channel shapes as function of x. '–' linear, '--' exponential, '-·' parabolic, '··' 3rd degree curve.

The method of characteristics is the most widely used solution method for solving the equations of pressure transients. If the transients are harmonic oscillations the impedance method is more practical and especially computationally efficient.

3. CHANNEL SHAPE OPTIMIZATION

Equations for different channel shapes were presented in the previous chapter. It is also possible to find an optimal shape *y(x)* for the channel such that the volume flow rate oscillation at the channel outlet is minimized. The optimization procedure follows the one presented by Bängtsson et al. [4] for the shape optimization of an acoustic horn. The objective function J is

$$J = Q_{out} \tag{17}$$

The value of the objective function depends on the shape of the channel, i.e. on the shape of $y(x)$. The shape is associated to a reference boundary which is the initial guess of the channel shape

$$y(x) = y(x)_{ref} + \alpha_k w_k(x) \tag{18}$$

where α_k is a deflection from the boundary, w_k is a basis function and k is the index of the design variable.

The length and the height at upstream and downstream ends are fixed and the channel height in between these ends must obey certain limits. The optimization problem is

$$\begin{aligned} &\min_{\alpha \in \Omega} J(\alpha) \\ &J(\alpha^*) \leq J(\alpha) \end{aligned} \tag{19}$$

where Ω is the admissible design domain and * denotes the optimal solution.

A gradient-based optimization algorithm is used. The gradient of the objective function reveals to which variations in the geometry the function is the most sensitive. A finite difference formula is used to calculate the derivatives as

$$\frac{\partial J}{\partial \alpha_k} = \frac{J_{new} - J_{old}}{\Delta \alpha_k} \tag{20}$$

For the following iteration, a step is taken in this direction and the procedure is repeated until convergence [5]. The new value of α is calculated from

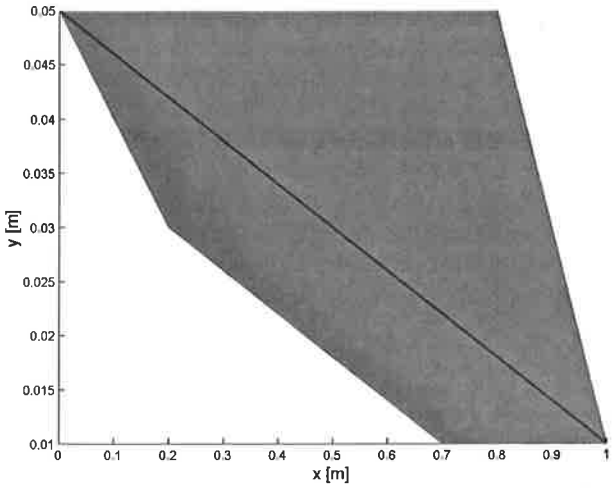


Fig. 3. Admissible design domain Ω and reference curve.

$$\alpha_k^{new} = \alpha_k^{old} - \varepsilon \frac{\partial J}{\partial \alpha_k} \tag{21}$$

where ε is a weighting term. The weighting parameter is used to weight the gradient information and changes with cost function and control variables. There exist more efficient ways to compute the gradients but in this computationally light problem the above described method is sufficient.

The fluid transients approach is used for solving the volume flow rate oscillation in the channel. Eq. (14) and (15) are solved using the new $y(x)$ and the procedure is repeated until the optimal shape is found. Furthermore, we want the curve $y(x)$ to be decreasing, i.e. $y(x_{i+1}) \leq y(x_i)$. The curve $y(x)$ is allowed to be in the domain depicted in Fig. 3.

4. RESULTS

The pressure and volume flow rate oscillation in channels of different shapes were compared using both the acoustical and fluid transients approaches. In the acoustical approach the analytical solutions were used for the available cases. For other cases the Webster's equation (2) was solved using Matlab's boundary value problem solver. The boundary conditions, the channel height at upstream and downstream ends and the channel length were the same in all cases.

The channel dimension were $y(0) = 0.05$ m, $y(L) = 0.01$ m, $L = 1$ m, $f = 60$ Hz corresponding to the blade passing frequency of a pump and the wave speed $a = 1200$ m/s for water ($\rho = 1000$ kg/m^3). The pressure oscillation amplitude at the upstream end was $P = 1$ kPa.

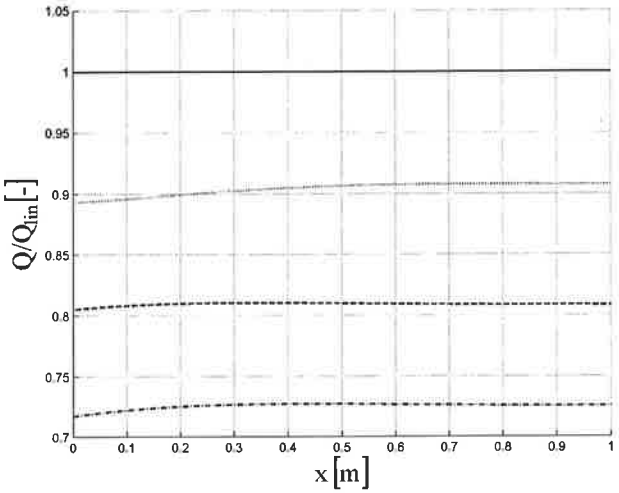


Fig 4. Volume flow rate oscillation in different channels as a function of x normalized with linear channel. '—' linear '--', exponential, '-·' parabolic, '··' 3rd degree curve.

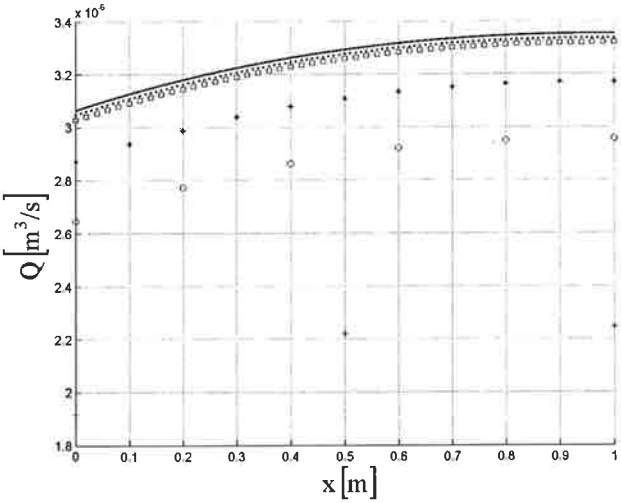


Fig. 5. Volume flow rate oscillation in linear channel for exact '—' and approximate solution of fluid transient method using '+' 2, 'o' 5, '*' 10, '□' 50 and '•' 100 sections.

The volume flow rate oscillation in different channels is compared in Fig. 4. The magnitude of oscillation is highest for the linear channel. The volume flow rate oscillation is lowest for the parabolic channel being approximately 28 % lower than in the linear channel. The exponential and 3rd degree curve channels are placed between linear and parabolic channels in volume flow rate oscillation, the latter having a higher oscillation, only about 10 % less than the linear channel.

Fig. 5 shows the volume flow rate oscillation in a linear channel along the channel length calculated using the method of fluid transients and compared with the exact solution of acoustics of Eq. (6). The plots are for the exact solution and for approximate solutions obtained using 2, 5, 10, 50 and 100 sections. By using very few sections the fluid transients method gives too small

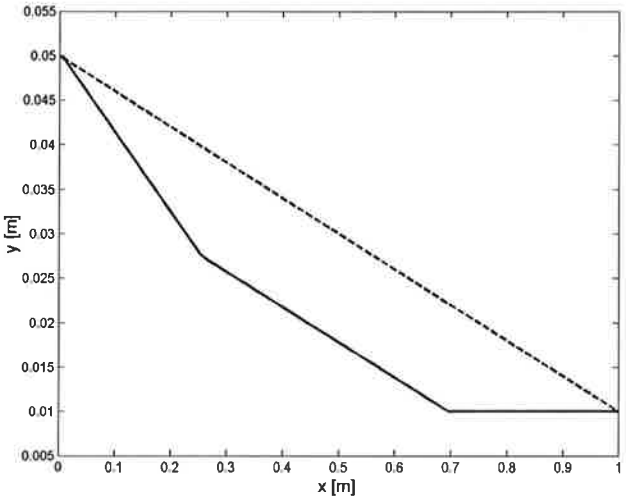


Fig. 6. Initial "--" and optimal "—" design of tapered channel shape.

values of the volume flow rate oscillation, but as the number of segments is increased, the solution converges quickly to that of the exact solution. With 50 sections the error is negligible compared to the exact solution.

The optimization procedure presented in chapter 3 was used to obtain an optimal channel shape minimizing the volume flow rate oscillation at the channel outlet. The linear initial shape and the optimal shape are presented in Fig. 6.

The optimal shape seen in Fig. 6 is very close to the parabolic curve, Fig. 2, and the improvement obtained compared to it is very small. The volume flow rate oscillation is 29 % lower in the optimized channel than in the linear one. The admissible design domain determines greatly what the optimal shape is. Clearly, the channel upper wall must be flat in the vicinity of the outlet. A small change in the pressure gradient favors the low volume flow rate oscillation. At the beginning of the channel the optimal curve descends quickly but it does not reach the admissible design boundary until at approximately 0.25 m from the upstream end. A smoothing technique [5] could be used in the optimization algorithm to avoid the sharp edges which are created by the admissible design boundary.

The optimal shape is found after very few iterations depending on the weighting parameter ε and the reference curve.

The results presented above are for a single frequency of 60 Hz. The changes are negligible in the case of another frequency, only the change of magnitude of volume flow rate oscillation is significant.

5. CONCLUSIONS

The theory of fluid transients and acoustics are combined in this study to analyze tapered channels. The effect of channel shape on the volume flow rate oscillation is examined for known pressure oscillation at the upstream end, which is created by rotating machinery such as pumps. Solutions were obtained by solving the one-dimensional Webster's equation of acoustics. Results were

compared with the impedance method of fluid transients, which gave exactly the same results with a sufficient number of channel sections.

The channel shape affects significantly the volume flow rate oscillation magnitude, when the channel length and end dimensions are the same. The lowest oscillation level at the outlet was obtained for a channel with the parabolic shape, while a linear channel shape gave the highest oscillation level. The channel shape was optimized using an optimization procedure to minimize the volume flow rate oscillation at the channel outlet. The optimal shape is very close to the parabolic channel. The optimal shape also depends greatly on the admissible design domain. A small gradient in the curve close to the outlet favors the low volume flow rate oscillation.

NOMENCLATURE

A = cross-sectional area
a = wave speed
b = slope of linear channel
C = capacitance
c_i = coefficient of polynomial
D = diameter
f = frequency, friction factor
g = gravitational acceleration
H = pressure (piezometric) head
h = constant of exponential channel
i = imaginary unit, channel section
J = objective function
k = wave number
L = length
l = length of channel section
m = constant of linear channel
n = integer depending on flow type
P = pressure amplitude
p = pressure
Q = volume flow rate
t = time
u = particle velocity
w = basis function
x = x-coordinate
y = channel height
Z_C = characteristic impedance

α = deflection from reference curve
ε = weighting term
γ = complex constant
ρ = fluid density
Ω = admissible design domain
ω = angular frequency

Subscripts

D = downstream

out = outlet
U = upstream
1 = upstream
2 = downstream

REFERENCES

[1] Morse, P.M. 1948. Vibration and Sound. McGraw-Hill, New York.

[2] L. M. B. C. Campos. 1984. Some general properties of the exact acoustic fields in horns and baffles. Journal of Sound and Vibration, **95**(2), pp 177-201.

[3] Wylie, E.B., Streeter, V.L. 1978. Fluid Transients. McGraw-Hill, New York.

[4] Bängtsson, E., Noreland, D., Berggren, M. 2003 Shape Optimization of an Acoustic Horn. Computer Methods in Applied Mechanics and Engineering, **192**, pp. 1533–1571.

[5] Mohammadi, B., Pironneau, O. 2001 Applied Shape Optimization for Fluids. Oxford University Press.

FINITE ELEMENT TECHNIQUES APPLIED TO GLACIOLOGICAL PROBLEMS

T. ZWINGER[1], O. GAGLIARDINI[1,2], J. RUOKOLAINEN[1], M. LYLY [1], P. RÅBACK[1]

[1] CSC – Scientific Computing Ltd.
Box 405, FIN-02101 Espoo
e-mail: Thomas.Zwinger@csc.fi

[2] LGGE CNRS-UJF Grenoble I,
BP 96,
38402 Saint Martin d'Hères,
France

ABSTRACT

Earth's cryosphere is a real-time indicator for climatological changes. If we understand the reaction of ice masses to variations in precipitation and surface temperatures or heat fluxes, we immediately can draw conclusions to the climate that caused those changes. Accurate models of the dynamics as well as the thermodynamics of ice-masses (e.g., glaciers, ice sheets, ice shelves, sea ice) are an inevitable tool to get insight into these highly complex processes and interactions taking place. Features such as thermo-mechanical coupling, free surface evolution, non-Newtonian rheologies and tracer dynamics have to be included in such models.

During the recent years, CSC's in-house FEM code Elmer has been adapted to meet with the complexity imposed by modern numerical Glaciology. In this article, we try to draw an outline of the mathematical formulations of the ice flow and the implemented numerical techniques for solving them. Examples and their results shall demonstrate the versatility of the model.

1 MODEL EQUATIONS

Ice is a viscous material and – if though very slowly – flows under the influence of gravity, $\boldsymbol{g}$. This slow motion is reflected by ignoring the acceleration terms in the conservation of linear momentum, leading to the Stokes equation

$$-\mathrm{grad}\, p + \mathrm{div}\ \boldsymbol{S} + \rho \boldsymbol{g} = \boldsymbol{0}. \tag{1}$$

Here, p denotes the pressure and $\boldsymbol{S}$ the deviatoric part of the stress tensor.

At the top of ice sheets and glaciers, fresh snow is converted into firn, which by compression transforms into ice. In contrary to pure ice, which can be considered to be incompressible (with a constant density ρ_{ice}), firn has a variable density. Neglecting the mass contribution of the enclosed air the density is proportional to the ice volume fraction, φ,

$$\rho = \rho_{\mathrm{ice}}\, \varphi. \tag{2}$$

As φ represents an additional degree of freedom, for compressible flow an evolution equation for the ice volume fraction has to be provided. Additionally, a density-pressure coupling parameter, $\kappa_{\mathrm{cp}}(\varphi)$, introduces a general form of the volume balance

$$\operatorname{div} \boldsymbol{v} + \kappa_{\mathrm{cp}}(\varphi) p = 0. \tag{3}$$

In the limit of pure ice, i.e., $\varphi \to 1$, this parameter vanishes, $\kappa_{\mathrm{cp}} \to 0$, restoring a divergence free velocity field, $\boldsymbol{v}$, and hence incompressibility.

A constitutive relation linking the deviatoric stress in Eq. (1) to the deviatoric part of the strain-rate tensor,

$$\boldsymbol{D} = \operatorname{sym} \operatorname{grad} \boldsymbol{v} - \underbrace{\operatorname{tr} (\operatorname{sym} \operatorname{grad} \boldsymbol{v})}_{\equiv \operatorname{div} \boldsymbol{v}}, \tag{4}$$

has to be provided. Different approaches, such as the commonly used (isotropic) Norton-Hoff type of non-linear flow law [1] as well as a linear orthotropic relation accounting for the internal crystal structure of the ice [2] have been implemented in Elmer. As the discussion of the latter would go far beyond the scope of this article, we will restrict ourselves to a general form of the first,

$$\begin{gathered} \boldsymbol{S} = 2\eta \boldsymbol{D}, \\ \eta = \frac{1}{a(\varphi)} \left(2\, E\, A(T')\right)^{-1/n} d_{\mathrm{D}}^{(1-n)/n}. \end{gathered} \tag{5}$$

The function $a(\varphi)$ and the invariant,

$$d_{\mathrm{D}} = \left(\frac{2 \operatorname{tr} (\boldsymbol{D}^2)}{a(\varphi)} + \frac{(\operatorname{div} \boldsymbol{v})^2}{b(\varphi)} \right)^{1/2}, \tag{6}$$

as well as the coupling parameter occurring in (3)

$$\kappa_{\mathrm{cp}} = b(\varphi) \cdot \left[a(\varphi)\, \eta(T', d_{\mathrm{D}}) \right]^{-1}, \tag{7}$$

introduce the aspect of compressibility to the system [3]. As for the limit of pure ice $\varphi \to 1$ those functions take the values $a \to 1$ and $b \to 0$ and expression (6) degenerates to the second invariant of the strain rate tensor, with the stress exponent $n = 3$, the widely used Glen's flow law [4] for polycrystalline ice is recovered.

The enhancement factor E usually is set to unity, but it can be used to model weaker layers caused by ice-impurities, by assigning a somewhat larger value.

The Arrhenius factor $A(T')$ in the flow law (5) introduces a pronounced dependency of viscosity upon the temperature, T. Consequently, this leads to the necessity of solving a heat transfer equation

$$\rho c(T) \left(\frac{\partial T}{\partial t} + \boldsymbol{v} \cdot \operatorname{grad} T \right) = -\operatorname{div} \boldsymbol{q} + \sigma, \tag{8}$$

along with the flow and pressure field. Expression (8) implicitly already contains the caloric equation of state, $u = \int_{T_0}^{T} c(\hat{T})\, d\hat{T}$, with the specific internal energy, u. Heat conduction is expressed by Fourier's law

$$\boldsymbol{q} = -\kappa(T) \operatorname{grad} T. \tag{9}$$

Non-linearity is imposed by temperature-dependence of the heat conductivity, $\kappa(T)$, as well as the heat capacity, $c(T)$ of ice [5]. The last contribution on the right-hand side of Eq. (8)

is a volumetric heat source that in Glaciological application usually can be identified with the production due to strain heating, i.e., $\sigma = \text{sym grad}\, \boldsymbol{v} : \boldsymbol{t}$. With the Clausius-Clapeyron constant, β, and the melting point at low pressure, T_0, the pressure melting point of ice is given by

$$T_{\mathrm{M}} = T_0 - \beta p. \tag{10}$$

It defines the (locally varying) upper limit of the temperature. Introducing the relative temperature, $T' = T - T_{\mathrm{M}}$, the constraint imposed on Eq. (8) reads as

$$T' \leq 0. \tag{11}$$

Ice at temperatures below T_{M} is called *cold ice*, whereas ice at the pressure melting point, i.e., $T' = 0$, is denominated as *temperate ice*.

A major application area of ice flow models is the support of dating of ice cores by providing an age-depth relation for the ice. To that end, the advection-reaction type of equation

$$\frac{\partial \mathcal{A}}{\partial t} + \boldsymbol{v} \cdot \text{grad}\, \mathcal{A} = 1, \tag{12}$$

has to be solved. If at the inflow parts at the free surface (indicated by the subscript "s") the boundary condition $\mathcal{A}|_{\mathrm{s}} = 0$ is set, the variable $\mathcal{A}$ then is equal to the age of the ice before present.

2 NUMERICAL METHODS

2.1 Diffusion Convection Equation

Diffusion-convection type of equations, such as the energy balance (8) take the general form

$$\nabla \cdot (a \nabla U) + \boldsymbol{v} \cdot \nabla U = \sigma. \tag{13}$$

We are interested in solutions for the unknown U on the closed domain $V \subset R^3$. Applying the standard Galerkin formulation with the test functions $\Phi \in H_0^1$ (e.g., the piecewise linear functions $\Phi_i \in P_1$) we obtain for the j-th node of the discretized domain

$$\sum_i U_i \int_V \underbrace{[a \nabla \Phi_i \cdot \nabla \Phi_j + \boldsymbol{v} \cdot \nabla \Phi_i \Phi_j]\, dV}_{=M_{ij}} = -\oint_{\partial V} q_n \Phi_j \, dA + \int_V \underbrace{\sigma \Phi_j \, dA}_{=F_i}. \tag{14}$$

The sum on the left-hand side represents the product of the solution vector, $\boldsymbol{U}$, and the system matrix $\boldsymbol{M}$. On the right-hand side the last term can be identified as the force vector $\boldsymbol{F}$. The surface integral of the outward pointing flux q_n, either contributes to the force vector, or – if there is a linear or even linearized – dependence upon the solution itself, also can have a contribution to the system matrix at the boundary nodes. As the resulting system

$$\boldsymbol{M} \cdot \boldsymbol{U} = \boldsymbol{F}, \tag{15}$$

is not stable for significant convection taking place, stabilization has to be obtained either by adding a term proportional to the residual (c.f., Stabilized method [6]) or by expansion of the numerical degrees of freedom inside an element (c.f., Residual Free Bubble method [7]).

If Eq. (14) has to comply with a local constraint of the form $U_i \leq U_i^{\mathrm{m}}$, then the problem turns into a variational inequality. At this point we would like to stress the fact that an algorithm applying an afterwards correction of values of the solution of the unconstrained system exceeding the imposed limit – as it has been applied to the heat transfer equation in several glaciological applications – is, neither in the physical nor in the numerical sense, an appropriate procedure.

2.1.1 Uzawa Method

The standard Uzawa method (see e.g., [8]) is – compared to for instance the Penalty method – a consistent way to introduce constraints to the solution of (15). The algorithm is straight forward implemented, simply by adding a vector $-\boldsymbol{\lambda}$ to the force vector on the right-hand side,

$$\boldsymbol{M} \cdot \boldsymbol{U} = \boldsymbol{F} - \boldsymbol{\lambda}. \tag{16}$$

Starting from an initial vector $\boldsymbol{\lambda} = \boldsymbol{0}$ the entry corresponding to the i-th node for the k-th step of the iteration is updated the following way

$$\lambda_i^{(k)} = \max\left[0,\, \lambda_i^{(k-1)} + \varrho \cdot (U_i - U_i^{\mathrm{m}})\right]. \tag{17}$$

The value of the Uzawa parameter, ϱ, has to be chosen with respect to convergence as well as stability of the algorithm. The iteration is stopped if convergence of the solution has been obtained. Then, λ_i represents the local sink (in case of heat transfer, the local cooling) needed to comply with the constraint $U_i \leq U_i^{\mathrm{m}}$.

2.2 A New Method for Introducing Limits of Variables

A very robust introduction of an upper constraint to a scalar variable can be achieved by direct manipulation of the system matrix and the force vector in (15). To this end, an outer iteration loop is implemented in the algorithm. For the k-th iteration the steps are then the following:

1. the system (15) is assembled: $\boldsymbol{M} \cdot \boldsymbol{U}^{(k)} = \boldsymbol{F}$.
2. if the i-th node is identified as *active*, an internal Dirichlet condition for the i-th entry in the solution vector is set by assigning the diagonal element the unity value (while nullifying the rest of the row), $M_{ii} \rightarrow \tilde{M}_{ii} \equiv 1$, and setting $F_i \rightarrow \tilde{F}_i = U_i^{\mathrm{m}}$.
3. a solution for the manipulated system

$$\tilde{\boldsymbol{M}} \cdot \boldsymbol{U}^{(k)} = \tilde{\boldsymbol{F}}, \tag{18}$$

   is computed.
4. the residual vector then is obtained by multiplication with the original system matrix and subtraction of the unmanipulated force vector

$$\boldsymbol{R} = \boldsymbol{M} \cdot \boldsymbol{U}^{(k)} - \boldsymbol{F}. \tag{19}$$

5. nodes are set to *active*, if $U_i^{(k)} \geq U_i^{\mathrm{m}}$
6. nodes are taken out of the *active*-node list only if $R_i > 0$

The loop given above is repeated, as long as convergence of $\boldsymbol{U}$ is obtained. In the case of pure diffusion (i.e., $\boldsymbol{v} \equiv \boldsymbol{0}$) in connection with an upper limit, the system (13) is equal to the contact problem of a membrane. Hence, the algorithm further shall be referred to as the *Contact method*. If $\boldsymbol{U}$ is interpreted as the solution for the dislocation of a membrane in the presence of a force, σ, the set of *active* nodes of the converged solution then defines the contact area.

2.3 Advection-reaction Equation

Advection-reaction type of equations, such as the dating problem (12) take the general form

$$a\,U + \boldsymbol{v} \cdot \nabla U = \Gamma. \tag{20}$$

The lack of any diffusion in the system requires a numerical approach somewhat different to the standard Galerkin formulation presented in Eq. (14). In our case a Discontinuous Galerkin method for advection-reaction equations [9] is applied.

3 APPLICATIONS

3.1 A Simple Test Case

A simple test case of pure heat diffusion (i.e., $\boldsymbol{v} \equiv \mathbf{0}$) shall provide the means to compare the performance of the both algorithms presented in section 2.1. A volume heat source of the roof-like form

$$\sigma = \sigma_{\max}\,(1 - |x|/50)\,, \tag{21}$$

is applied over a two-dimensional area of 100×100 meters ($x \in [-50, +50]$, $y \in [-50, +50]$) with constant temperature of $-10°$ Celsius at the boundaries. The maximum value of the heat source, $\sigma_{\max} = 0.1\,\mathrm{W\,m^{-2}\,m^{-1}}$, is chosen such that the upper limit of $0°$ Celsius in case of the solution of (8) clearly is exceeded (by about $13°$). The left part of Fig. 1 shows the unconstrained solution with values of the temperature larger than $0°$ being reset to that value, i.e., $T \rightarrow \max(T, 0)$. The numerical value of the heat conductivity is set to the one for pure ice at $-10°$ Celsius, $\kappa = 2.1929\,\mathrm{W\,m^{-1}\,K^{-1}}$, and for the sake of simplicity assumed to be independent of temperature. The result obtained with the contact

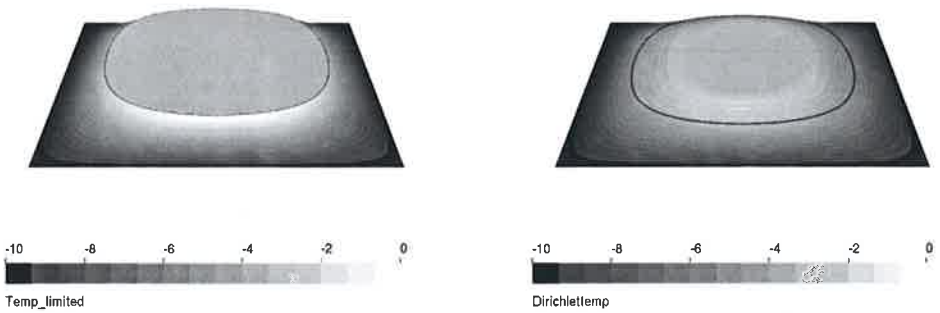


Figure 1: Distribution of the unconstrained solution (left) with a threshold of $0°$ Celsius applied as well as of the Dirichlet algorithm (right). The iso-line of zero temperature in the unconstrained run is shown in the latter picture. Scales are given in ° Celsius

method shows a smooth transition to the temperate area, whereas the posterior implied threshold $T \rightarrow \max(T, 0)$ to the unconstrained solution naturally – besides producing a too large temperate area – shows an unsteadiness of the solution at the cold-temperate ice transition line. The runs with the Uzawa and the contact method have been iterated with a

convergence criterion $\Delta = 2\,|N^{(k-1)} - N^{(k)}|\,(N^{(k-1)} + N^{(k)})^{-1} < 1 \times 10^{-6}$ for the iteration loop. Here, $N^{(k)}$ stands for the norm obtained for the solution of the k-th iteration step.

The convergence history diagram in Fig. 2 shows the clear advantage in terms of performance of the contact method in comparison to the the Uzawa algorithm. The converged solutions differ only in the range of numerical accuracy.

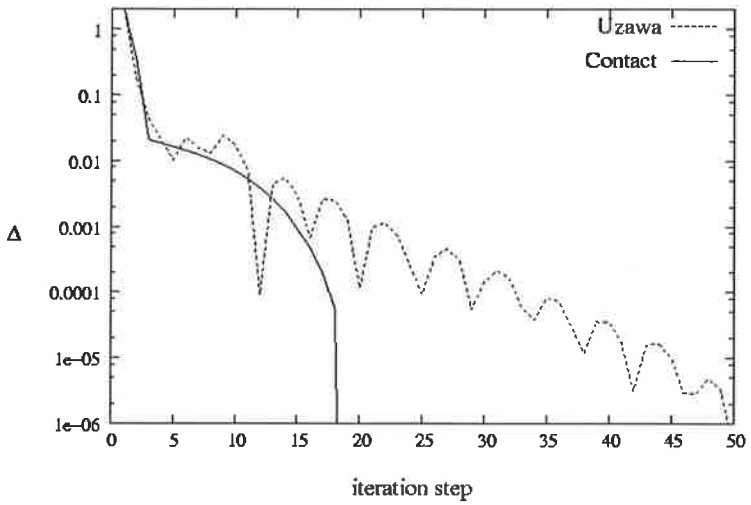


Figure 2: Comparison of the convergence history between the Uzawa- (dashed line) and the Contact method (solid line)

As mentioned before, for the converged solution of the contact algorithm, the distribution of the residual, $\boldsymbol{R}$, then is equivalent to the needed heat sink to comply with the imposed upper limit. The distribution of the residual is shown in Fig. 3.

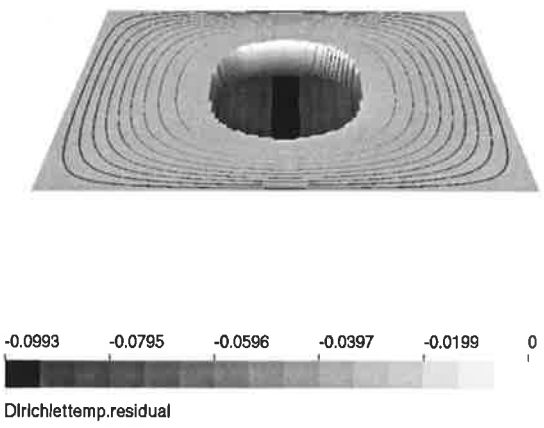


Figure 3: Distribution of the residual of the converged solution using the the Contact method. The residual $\boldsymbol{R}$ can be interpreted as the needed local cooling in order to keep the material at the imposed limit of the pressure melting point. The iso-lines show the temperature distribution beginning at $-10°$ Celsius in unit steps

3.2 Application to Crater Glacier Flow

We further present the steady state ($\partial/\partial t = 0$) solution of a diagnostic run (i.e., prescribed shape of free surface) of the Gorshkov crater glacier, at Kamchatka, Russia. The glacier fills the whole, approximately 800 meters in diameter wide crater and has an outflow region into the caldera at its northern face [10]. The assumed distribution of the (in Glaciological terms extremely high) geothermal heatflux, q_n, is shown in Fig. 4, where also the position of a bore-hole (K2) is indicated. The density profile obtained from this bore-hole is taken to be valid for the whole glacier [10].

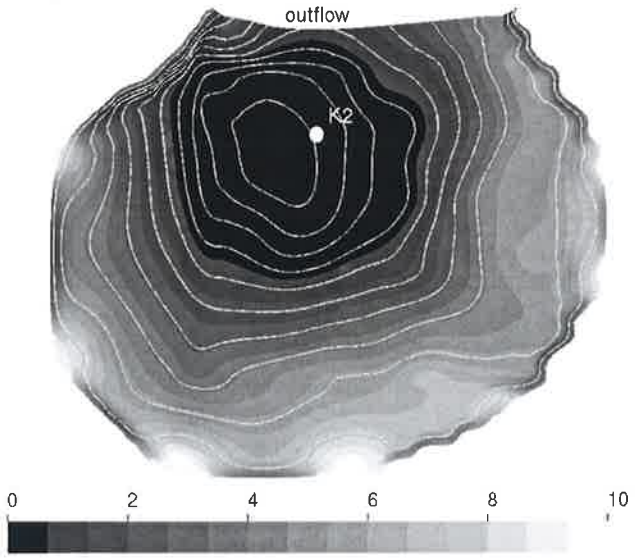


Figure 4: Distribution of the geothermal heatflux, q_n, along the bedrock of the Gorshkov crater. Values are given in $\mathrm{W\,m^{-2}}$. Iso-levels show the bedrock elevation starting from 3650 meters a.s.l. in 25 meter steps until 3900 meters a.s.l. The position of the bore-hole K2 is indicated

The obtained field for the relative temperature, T', as it is depicted in Fig. 5, shows a more or less linear distribution with flow depth from the minimum value of the average surface temperature ($-16.6°$ C) to the maximum value at the bedrock. We find that, apart

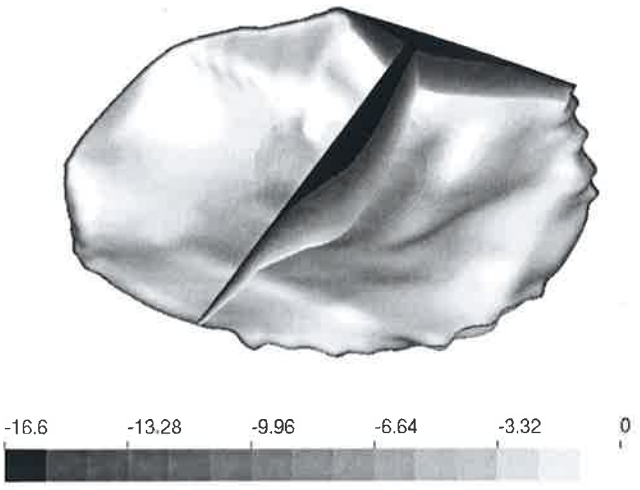


Figure 5: Relative temperature, T' [° C], distribution along the bedrock and at a south-north oriented cross section (including the bore-hole K2 position)

from an area of about 100 meters radius around the lowest part of the bedrock, the whole

bedrock temperature is at the pressure melting point.

The functions, $a(\varphi)$ and $b(\varphi)$, entering the constitutive relations in (1) and (3) are set according to the values given by Gagliardini and Meyssonier [3] obtained from shear/compression experiments of firn.

The distributions of the relative temperature, T', as well as the date, $\mathcal{D} = 2000 - \mathcal{A}$, along the K2 bore-hole position are depicted in Fig. 6.

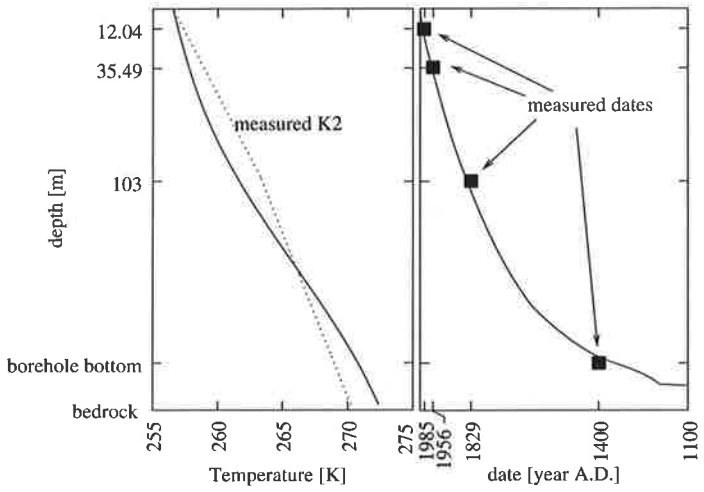


Figure 6: Profiles along the K2 bore-hole obtained from computer simulation

4 SUMMARY

A wide spectrum of numerical challenges is imposed by numerical Glaciology. CSC's indoor FEM code Elmer [11] has been adapted to a variety of these problems. In particular, methods for solving advection-reaction equations (tracer equation, hyperbolic nature) as well as diffusive systems including convection (e.g., heat transfer equation) have been implemented. Diagnostic simulation runs of existing ice masses show good agreement with field data.

ACKNOWLEDGMENTS

The studies of the crater glacier have been undertaken in close cooperation with Prof. R. Greve from the Institute of Low Temperature Science (ILTS), Hokkaido University, Japan, and Dr. T. Shiraiwa from the Research Institute for Humanity and Nature, Kyoto, Japan.

The work of Olivier Gagliardini at CSC has been supported by the *Association Franco-finlandaise pour la Recherche Scientifique et Technique* (AFFRST) as well as the *Vilho, Yrjö and Kalle Väisälä Foundation.*

REFERENCES

[1] E. Le Meur, O. Gagliardini, T. Zwinger and J. Ruokolainen. *Glacier flow modelling: a comparison of the Shallow Ice Approximation and the full-Stokes solution*, C. R. Physique **5** (7), 709-722, 2004.

[2] F. Gillet-Chaulet, O. Gagliardini, J. Meyssonnier, T. Zwinger and J. Ruokolainen. *Flow-induced anisotropy in polar ice and related ice-sheet flow modelling*, J. Non-Newtonian. Fluid Mech. **134**, 33-43, 2006

[3] O. Gagliardini, J. Meyssonnier. *Flow simulation of firn-covered cold glacier*, Annals Glaciol., **24** (158), 242-248, 1997

[4] J. W. Glen. *The creep of polycrystalline ice*, Proc. R. Soc. London, Ser. A, **228** (1175), 519-538, 1955

[5] W. S. B. Paterson. *The Physics of Glaciers*, Pergamon Press, Oxford etc., 3rd edition, 1994

[6] L.P. Franca, S.L. Frey and T.J.R. Hughes. *Stabilized finite element methods: II The incompressible Navier-Stokes equations*, Comp. Meths. Appl. Mech. **95**, 253-276, 1992

[7] C. Baiocchi, F. Brezzi and L.P. Franca. *Virtual bubbles and the Galerkin least square method*, Comp. Meths. Appl. Mech.Eng. **105**, 125-141, 1993

[8] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, Springer series in computational mathematics **15**, Springer Verlag, New York etc., 1991

[9] F. Brezzi, L. D. Marini, E. Süli. *Discontinuous Galerkin methods for first-order hyperbolic problems* Math. Models Methods Appl. Sci. **14** (12), 1893-1903, 2004

[10] T. Shiraiwa, Y.D. Murav'yev, T. Kameda, F. Nishio, Y. Toyama, A. Takahashi, A.A. Ovsyannikov, A.N. Salamatin and K. Yamagata. *Characteristics of a crater glacier at Ushkovsky volcano, Kamchatka, Russia, as revealed by the physical properties of ice cores and borehole thermometry*, J. Glac. **47** (158), 423-432, 2001

[11] *ELMER finite element software homepage*, `http://www.csc.fi/elmer`

NIPPIKONTAKTIN VÄRÄHTELYJEN SYNTYMEKANISMISTA KUMILLA PINNOITETUN TELAN YHTEYDESSÄ

Juha Toivola
Mekalyysi Oy
Myllymäentie 5
37960 Sotkia
toivolaj@surfeu.fi

TIIVISTELMÄ

Paperia päällystetään erilaisilla aineilla hiertämällä päällystysaine paperiin nippikontaktissa. Nippikontaktin pituuden lisäämiseksi nipin toinen tela pinnoitetaan kumilla. Kumi käyttäytyy viskoelastisesti, jolloin siihen syntynyt painuma ei välttämättä ehdi palautumaan telan kiertäessä yhden kierroksen. Kun kumipinnoitteen palautumaton painautuma tulee uudelleen nippiin, se herättää telojen jonkin tai jotkin ominaistaajuudet. Tästä aiheutuva telojen värähtely aiheuttaa uusia painaumia kumipinnoitteeseen, ja mikäli telan pyörimistaajuus tai joku sen kerrannainen osuu telojen värähtelyjen taajuudelle, painaumat syntyvät aina samaan kohtaan kumipinnoitetun telan kehällä ja ilmiö voimistuu itsestään.

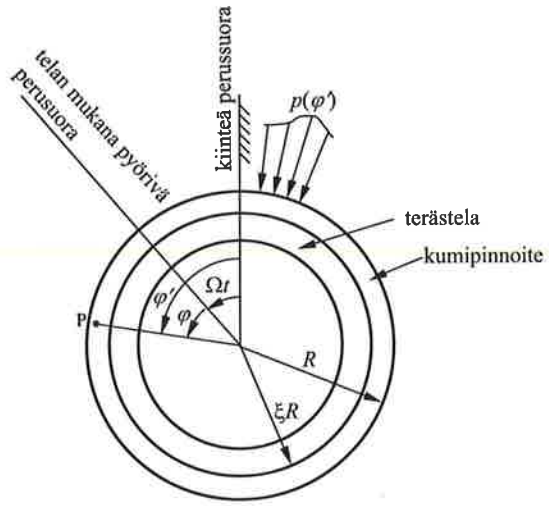
Edellä kuvattu ilmiö esiintyy vain joidenkin tiettyjen telan pyörimistaajuuden kerrannaisten osuessa telojen ominaistaajuuksille. Jotta voitaisiin selvittää, miksi vain tietyt telan pyörimistaajuuden kerrannaiset ovat kriittisiä, on tarkasteltava kumipinnoitteen dynaamista käyttäytymistä. Ilmiön ymmärtämistä voidaan syventää ratkaisemalla kumipinnoitteen siirtymäkentän osittaisdifferentiaaliyhtälöt analyyttis-numeerisesti sopivassa malliongelmassa.

Tässä tutkimuksessa kumipinnoitetta mallinnettiin lineaarisesti kimmoisella, isotrooppisella ja homogeenisella materiaalimallilla. Siirtymäkentän kinematiikka oletettiin lineaariseksi. Telan keskellä vallitsee tasomuodonmuutostila, jolloin taso-ongelman ratkaisu riittää. Kumipinnoitteen kehän suunnassa ratkaisuyritteeksi otettiin telan pyörimisen suuntaan etenevä aalto (co-rotating wave) ja vastakkaiseen suuntaan etenevä aalto (counter-rotating wave), mikä merkitsee myös sitä, että tarkastellaan vain steady-state-ratkaisua. Ratkaisun radiaalisuuntainen osuus saadaan tämän jälkeen tavallisen differentiaaliyhtälöryhmän ratkaisuna. Tässä tutkimuksessa tämä tavallinen differentiaaliyhtälöryhmä ratkaistiin numeerisesti differenssimenetelmällä.

MALLIONGELMA

Tarkasteltavan malliongelman geometria on esitetty kuvassa 1. Malliongelmassa vakiokulmanopeudella Ω pyörivä terästela oletetaan täysin jäykäksi. Lineaarisesti kimmoiseen, isotrooppiseen ja homogeeniseen kumipinnoitteeseen kohdistuu kiinteästä koordinaatistosta katsoen paikallaan pysyvä ja muuttumaton painekuormitus $p(\varphi')$. Telan pyörimisestä johtuen painekuormitus liikkuu telaan nähden, jolloin kyseessä on liikkuvan kuorman ongelma (moving load problem). Telan kes-

kialueella kumipinnoitteessa valitsee tasomuodonmuutostila, jolloin taso-ongelman tarkastelu riittää. Telaan kiinnitetään telan mukana pyörivä $r\varphi$ -napakoordinaatisto ja kiinteään koordinaatistoon kiinnitetään samankeskinen $r\varphi'$ -napakoordinaatisto.



Kuva 1. Malliongelman geometria ja koordinaatistot.

DEFORMAATIOKENTÄN MALLI

Esitetyillä oletuksilla kumipinnoitteen siirtymäkenttää kuvaavat yhtälöt (Lamen yhtälöt) telaan kiinnitetyssä sylinterikoordinaatistossa ovat

$$
\begin{aligned}
&e_{,\tilde{r}} - 2\varepsilon\left(\frac{1}{\tilde{r}}\omega_{z,\varphi} - \omega_{\varphi,\tilde{z}}\right) - \frac{\varepsilon\rho}{G}\tilde{a}_r = 0 \\
&\frac{1}{\tilde{r}}e_{,\varphi} - 2\varepsilon\left(\omega_{r,\tilde{z}} - \omega_{z,\tilde{r}}\right) - \frac{\varepsilon\rho}{G}\tilde{a}_\varphi = 0 \\
&e_{,\tilde{z}} - 2\varepsilon\frac{1}{\tilde{r}}\left[\left(\tilde{r}\omega_\varphi\right)_{,\tilde{r}} - \omega_{r,\varphi}\right] - \frac{\varepsilon\rho}{G}\tilde{a}_z = 0
\end{aligned}
\qquad (1)
$$

missä

$$
e = \frac{1}{\tilde{r}}\left(\tilde{r}\tilde{u}\right)_{,\tilde{r}} + \frac{1}{\tilde{r}}\tilde{v}_{,\varphi} + \tilde{w}_{,\tilde{z}} \qquad \varepsilon = \frac{1-2\nu}{2(1-\nu)}
$$

$$
\omega_r = \tfrac{1}{2}\left(\frac{1}{\tilde{r}}\tilde{w}_{,\varphi} - \tilde{v}_{,\tilde{z}}\right) \qquad \omega_\varphi = \tfrac{1}{2}\left(\tilde{u}_{,\tilde{z}} - \tilde{w}_{,\tilde{r}}\right) \qquad \omega_z = \frac{1}{2\tilde{r}}\left[\left(\tilde{r}\tilde{v}\right)_{,\tilde{r}} - \tilde{u}_{,\varphi}\right]
$$

(2)

Dimensiollisia muuttujia on merkitty tildellä (~). Telan keskialueella kumipinnoitteessa vallitsee tasomuodonmuutostila, jolloin $\widetilde{w} \equiv 0$, $(),_{\widetilde{z}} \equiv 0$ ja $\widetilde{a}_z = 0$. Tällöin yhtälöryhmän (1) viimeinen yhtälö toteutuu identtisesti.

Kun tela pyörii vakiokulmanopeudella $\widetilde{\Omega}$, nollasta eroavat kiihtyvyyskomponentit ovat

$$\begin{aligned} \widetilde{a}_r &= -\widetilde{\Omega}^2\widetilde{r} - 2\widetilde{\Omega}\widetilde{v},_{\widetilde{t}} + \widetilde{u},_{\widetilde{t}\widetilde{t}} \\ \widetilde{a}_\varphi &= 2\widetilde{\Omega}\widetilde{u},_{\widetilde{t}} + \widetilde{v},_{\widetilde{t}\widetilde{t}} \end{aligned} \tag{3}$$

Ottamalla käyttöön dimensiottomat muuttujat

$$r = \frac{\widetilde{r}}{R} \qquad u = \frac{\widetilde{u}}{R} \qquad v = \frac{\widetilde{v}}{R} \qquad t = \frac{\widetilde{t}}{\sqrt{\rho R^2 / G}} \qquad \Omega = \widetilde{\Omega}\sqrt{\rho R^2 / G} \tag{4}$$

saadaan yhtälöryhmän (1) kaksi ensimmäistä yhtälöä dimensiottomaan muotoon

$$\begin{aligned} &u,_{rr} + \frac{1}{r}u,_r - \frac{1}{r^2}u + \frac{\varepsilon}{r^2}u,_{\varphi\varphi} - (1+\varepsilon)\frac{1}{r^2}v,_\varphi + (1-\varepsilon)\frac{1}{r}v,_{r\varphi} - \varepsilon\ddot{u} + \varepsilon 2\Omega\dot{v} + \varepsilon\Omega^2 r = 0 \\ &\varepsilon v,_{rr} + \frac{\varepsilon}{r}v,_r - \frac{\varepsilon}{r^2}v + \frac{1}{r^2}v,_{\varphi\varphi} + (1+\varepsilon)\frac{1}{r^2}u,_\varphi + (1-\varepsilon)\frac{1}{r}u,_{r\varphi} - \varepsilon\ddot{v} - \varepsilon 2\Omega\dot{u} = 0 \end{aligned} \tag{5}$$

Aikaderivaattaa dimensiottoman ajan t suhteen on merkitty pisteellä suureen päällä.

Kumipinnoitteen sisäreunalla $\widetilde{r} = \xi R$ siirtymät ovat nollia (telan mukana pyörivässä koordinaatistossa), joten sisäreunalla reunaehdoiksi saadaan

$$\begin{aligned} u(\xi,\varphi,t) &= 0 \\ v(\xi,\varphi,t) &= 0 \end{aligned} \tag{6}$$

Dimensiottomat radiaalisuuntainen normaalijännitys ja leikkausjännitys ovat

$$\begin{aligned} \sigma_r &= \frac{\widetilde{\sigma}_r}{G} = \frac{2}{1-2\nu}\left[(1-\nu)u,_r + \frac{\nu}{r}u + \frac{\nu}{r}v,_\varphi\right] \\ \tau_{r\varphi} &= \frac{\widetilde{\tau}_{r\varphi}}{G} = \frac{1}{r}u,_\varphi + v,_r - \frac{1}{r}v \end{aligned} \tag{7}$$

Kumipinnoitteen ulkoreuna $\widetilde{r} = R$ on vapaa lukuunottamatta kuormitusaluetta, jossa vaikuttaa annettu, kiinteästä koordinaatistosta katsoen paikallaan pysyvä painejakauma. Ulkoreunan reunaehdot ovat siis

$$\begin{aligned} \sigma_r(1,\varphi,t) &= -p(\varphi,t) \\ \tau_{r\varphi}(1,\varphi,t) &= 0 \end{aligned} \tag{8}$$

Koska ulkoreunan kuormitus pysyy kiinteästä koordinaatistosta katsottuna paikallaan, se voidaan kehittää Fourier-sarjaksi käyttäen kiinteän koordinaatiston kulmamuuttujaa φ':

$$p(\varphi')=\frac{\tilde{p}(\varphi')}{p_{\max}}=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left[a_n\cos(n\varphi')+b_n\sin(n\varphi')\right]$$

$$a_n=\frac{1}{\pi}\int_0^{2\pi}p(\varphi')\cos(n\varphi')d\varphi',\quad n=0,1,2,...$$

$$b_n=\frac{1}{\pi}\int_0^{2\pi}p(\varphi')\sin(n\varphi')d\varphi',\quad n=1,2,3,... \tag{9}$$

Telan mukana pyörivän koordinaatiston kulmamuuttujan φ ja kiinteän koordinaatiston kulmamuuttujan φ' välillä on yhteys

$$\varphi'=\Omega t+\varphi \tag{10}$$

Tällöin Fourier-sarja (9) voidaan kirjoittaa telan mukana pyörivässä napakoordinaatissa muotoon

$$p(\varphi)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left[a_n\cos(n\Omega t+n\varphi)+b_n\sin(n\Omega t+n\varphi)\right] \tag{11}$$

DEFORMAATIOKENTÄN RATKAISU

Yhtälöille (5) haetaan ratkaisua muodossa

$$\begin{aligned}u(r,\varphi,t)=&\frac{U_0(r,t)}{2}+\sum_{n=1}^{\infty}\left[U_{n1}(r)\sin(n\Omega t-n\varphi)+U_{n2}(r)\cos(n\Omega t-n\varphi)+\right.\\&\left.+U_{n3}(r)\sin(n\Omega t+n\varphi)+U_{n4}(r)\cos(n\Omega t+n\varphi)\right]\\v(r,\varphi,t)=&\frac{V_0(r,t)}{2}+\sum_{n=1}^{\infty}\left[V_{n1}(r)\sin(n\Omega t-n\varphi)+V_{n2}(r)\cos(n\Omega t-n\varphi)+\right.\\&\left.+V_{n3}(r)\sin(n\Omega t+n\varphi)+V_{n4}(r)\cos(n\Omega t+n\varphi)\right]\end{aligned} \tag{12}$$

Ratkaisuyritteissä muotoa $\sin(n\Omega t-n\varphi)$ ja $\cos(n\Omega t-n\varphi)$ olevat termit esittävät pyörimisen suuntaan eteneviä deformaatioaaltoja ja muotoa $\sin(n\Omega t+n\varphi)$ ja $\cos(n\Omega t+n\varphi)$ vastakkaiseen suuntaan eteneviä deformaatioaaltoja. Ottamalla huomioon yhteys (10) nähdään, että pyörimisen suuntaan nähden vastakkaiseen suuntaan etenevät deformaatioaallot pysyvät kiinteästä koordinaatistosta katsottuna paikallaan.

Steady-state-ratkaisua haettaessa havaitaan, että ainoa heräte tulee reunaehtojen kautta. Reunaehtojen Fourier-sarjassa esiintyy ainoastaan vastakkaiseen suuntaan etenevien deformaatioaaltojen herätteitä, joten voidaan päätellä, että pyörimisen suuntaan etenevien termien steady-state-ratkaisuiksi tulee $U_{n1}=U_{n2}=V_{n1}=V_{n2}=0$. Samaan tulokseen päästään kirjoittamalla ko. termien yhtälöt reunaehtoineen. Tuloksena saadaan homogeenisia yhtälöitä homogeenisilla reunaehdoilla.

Tuntemattomat funktiot U_{n3}, U_{n4}, V_{n3} ja V_{n4} saadaan ratkaistua sijoittamalla yritteet (12) yhtälöihin (5) sekä reunaehtoihin (6 ja 8) ja vertaamalla sini- ja kosinitermien kertoimia, jolloin saadaan tuntemattomia funktiota hallitsevat differentiaaliyhtälöt reunaehtoineen.

Tapaukselle $n = 0$ saadaan:

$$U_{0,rr} + \frac{1}{r}U_{0,r} - \frac{1}{r^2}U_0 - \varepsilon\ddot{U}_0 + \varepsilon 2\Omega\dot{V}_0 + \varepsilon 2\Omega^2 r = 0$$

$$\varepsilon V_{0,rr} + \frac{\varepsilon}{r}V_{0,r} - \frac{\varepsilon}{r^2}V_0 - \varepsilon\ddot{V}_0 - \varepsilon 2\Omega\dot{U}_0 = 0$$

$$\begin{aligned} r = \xi: \quad & U_0 = 0 \\ & V_0 = 0 \\ r = 1: \quad & (1-\nu)U_{0,r} + \frac{\nu}{r}U_0 = -\frac{1-2\nu}{2}a_0 \\ & V_{0,r} - \frac{1}{r}V_0 = 0 \end{aligned} \tag{13}$$

Koska tapauksessa $n = 0$ kenttäyhtälöissä ja reunaehdoissa ei esiinny ajasta riippuvia kuormitustermejä, voidaan päätellä, että myöskään steady-state-ratkaisu ei voi riippua ajasta. Tällöin saadaan

$$U_0 = \frac{1}{1+(1-2\nu)\xi^2}\left\{\frac{(1-2\nu)a_0}{2}\left(\frac{\xi^2}{r} - r\right) + \frac{\varepsilon\Omega^2}{4}\left[-\left(1+(1-2\nu)\xi^2\right)r^3 + \right.\right.$$

$$\left.\left. + \left(3-2\nu+(1-2\nu)\xi^4\right)r - \left(3-2\nu-\xi^2\right)\frac{\xi^2}{r}\right]\right\}$$

$$V_0 = 0 \tag{14}$$

Jatkossa tässä käsitellään erityistapausta, jossa painekuormitus on symmetrinen kiinteän koordinaatiston kulmamuuttujan arvon $\varphi' = 0$ suhteen. Tällöin Fourier-sarjan (9 tai 11) kertoimet $b_n = 0$ ja mikäli ao. yhtälöt kirjoitetaan, saadaan steady-state-ratkaisuksi $U_{n3} = V_{n4} = 0$. Funktioille U_{n4} ja V_{n3} saadaan:

$$U_{n4,rr} + \frac{1}{r}U_{n4,r} - \left(\frac{1+\varepsilon n^2}{r^2} - \varepsilon n^2\Omega^2\right)U_{n4} + \frac{(1-\varepsilon)n}{r}V_{n3,r} - \left[\frac{(1+\varepsilon)n}{r^2} - \varepsilon n 2\Omega^2\right]V_{n3} = 0$$

$$\varepsilon V_{n3,rr} + \frac{\varepsilon}{r}V_{n3,r} - \left[\frac{\varepsilon+n^2}{r^2} - \varepsilon n^2\Omega^2\right]V_{n3} - \frac{(1-\varepsilon)n}{r}U_{n4,r} - \left[\frac{(1+\varepsilon)n}{r^2} - \varepsilon n 2\Omega^2\right]U_{n4} = 0$$

$$\begin{aligned} r = \xi: \quad & U_{n4} = 0 \\ & V_{n3} = 0 \\ r = 1: \quad & (1-\nu)U_{n4,r} + \frac{\nu}{r}U_{n4} + \frac{\nu n}{r}V_{n3} = -\frac{1-2\nu}{2}a_n \\ & -\frac{n}{r}U_{n4} + V_{n3,r} - \frac{1}{r}V_{n3} = 0 \end{aligned} \tag{15}$$

Edellä esitetyt tavalliset differentiaaliyhtälöt reunaehtoineen ratkaistiin numeerisesti differenssimenetelmällä.

LASKENTATULOKSIA

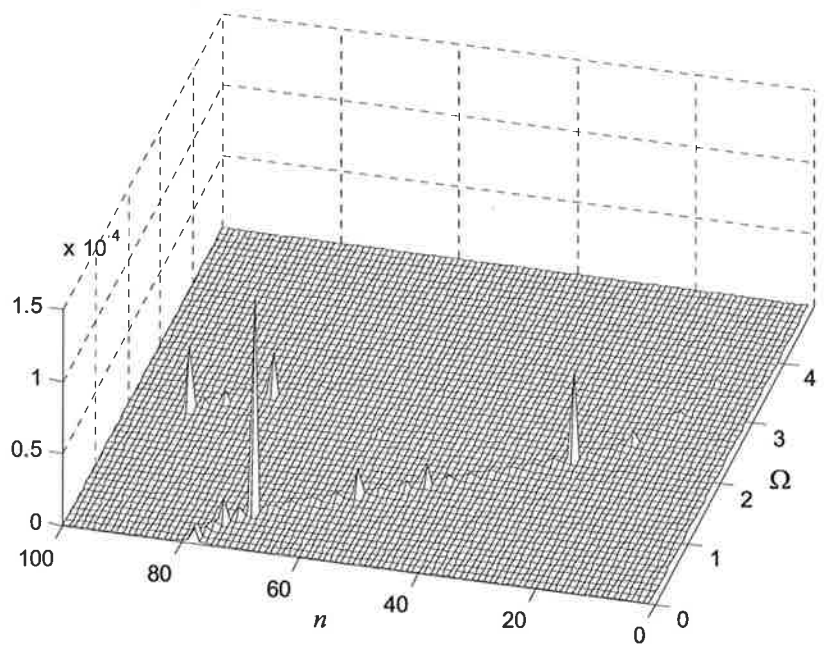
Kuormitus annetaan laskentaan painejakauman Fourier-sarjan kertoimien avulla. Tällöin voidaan helposti tarkastella erilaisia painejakaumia. Painejakauma voitaisiin myös määrittää iteratiivisesti sellaiseksi, että kumipinnoitteeseen syntyvä painauma seuraa pinnoitetta painavan telan muotoa. Tämän tutkimuksen yhteydessä siihen ei kuitenkaan menty, vaan esimerkkituloksien laskentaan käytettiin parabolista painejakaumaa:

$$p(\varphi') = \begin{cases} -\frac{1}{\alpha_0^2}\varphi'^2 + 1 & \varphi' \in [-\alpha_0, \alpha_0] \\ 0 & \varphi' \notin [-\alpha_0, \alpha_0] \end{cases} \tag{16}$$

Kuormituksen Fourier-sarjan kertoimet ovat tällöin

$$a_n = -\frac{4}{\pi}\frac{n\alpha_0\cos(n\alpha_0) - \sin(n\alpha_0)}{\alpha_0^2 n^3}$$
$$b_n = 0 \tag{17}$$

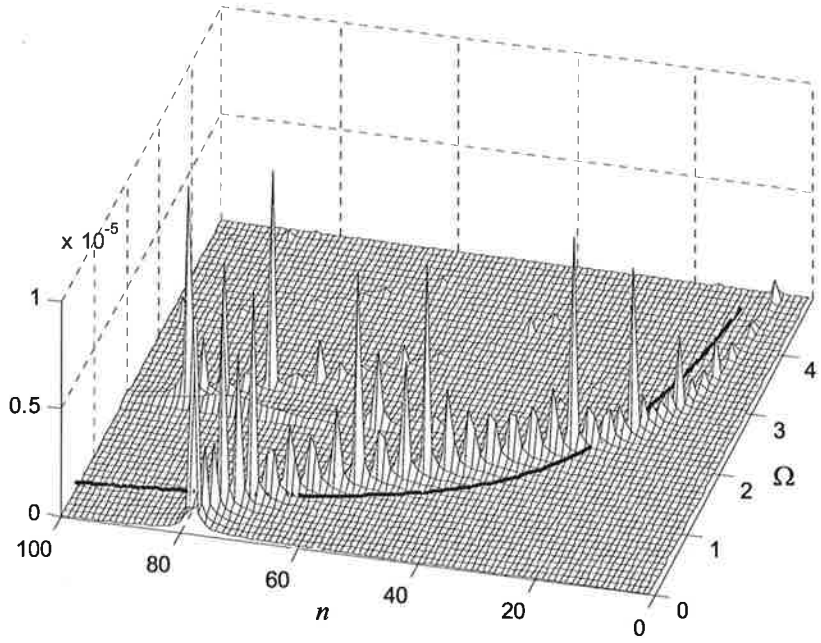
Kuvassa 2 on esitetty funktioiden U_{n4} itseisarvo telan pinnalla ($r = 1$) dimensiottoman kulmanopeuden Ω ja kertaluvun n funktiona. Tämä osoittaa, miten voimakkaasti kyseinen kertaluku herää. Kuten kuvasta 2 nähdään voimakkaimmin heräävän deformaatioaallon kertaluku riippuu telan pyörimisnopeudesta.



Kuva 2. Radiaalisiirtymän itseisarvon maksimi telan pinnalla. Radiaalisuunnassa käytetty 10 hilapistettä. $\nu = 0{,}49$, $\xi = 0{,}94$.

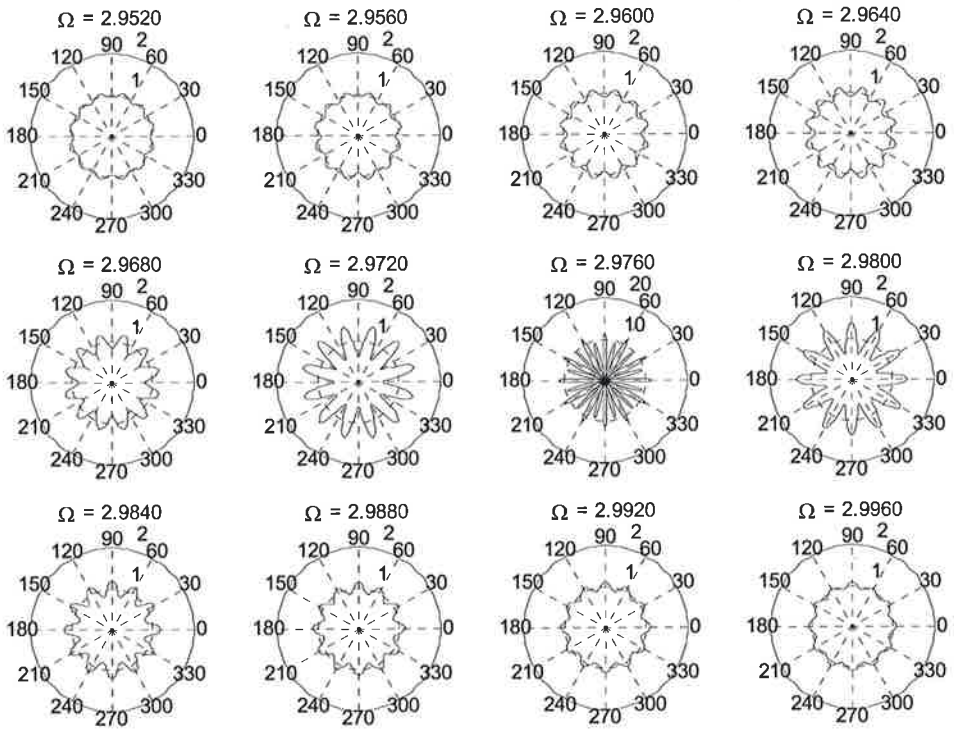
Kun oletetaan, että telan ominaistaajuudet eivät riipu (merkittävästi) pyörimisnopeudesta, ominaistaajuuksia kuvaavat käyrät ovat $\Omega - n$-koordinaatistossa hyperbelejä ($n\Omega = \omega_n$, missä ω_n on telan taivutusmuodon dimensioton ominaiskulmataajuus). Kuvassa 3 on muutettu pystyakselin skaalausta kuvaan 2 nähden ja lisätty telan taivutusmuodon erään ominaistaajuuden kuvaaja. Selvyyden vuoksi esitetty maksimi on rajattu arvoon 10^{-5} . Kun ominaistaajuuden käyrä leikkaa kuvassa näkyvän "vuorijonon", kumipinnoitteeseen herää deformaatioaalto, jonka kertaluku kerrottuna telan kulmanopeudella on ominaiskulmataajuuden suuruinen. Kumin viskoelastisen käyttäytymisen vuoksi on mahdollista, että tämä deformaatioaalto ei ehdi palautumaan, vaan lähtee pyörimään telan mukana saapuen nippiin ja herättäen telan taivutusominaismuodon juuri "oikealla" taajuudella. Tämän mahdollisuuden selvittämiseksi tarkastelu olisi tehtävä viskoelastisella materiaalimallilla.

On syytä huomata, että kuvissa 2 ja 3 näkyvän "vuorijonon" piikkien korkeus ei välttämättä anna oikeaa kuvaa deformaatioaallon voimakkuudesta laskennassa käytetystä Ω -resoluutiosta johtuen. Laskennassa käytetyt Ω -arvot eivät välttämättä osu "vuorijonon" harjanteen kohdalle kyseeseen tulevalla kertaluvulla. Ω -resoluutiota suurentamalla kuva tarkentuisi tältä osin.

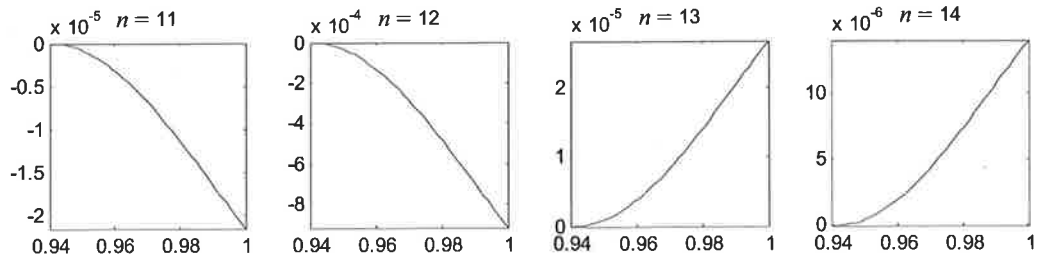


Kuva 3. Radiaalisiirtymän itseisarvon maksimi telan pinnalla, maksimi rajoitettu arvoon 10^{-5} . Radiaalisuunnassa käytetty 10 hilapistettä. $\nu = 0{,}49$, $\xi = 0{,}94$.

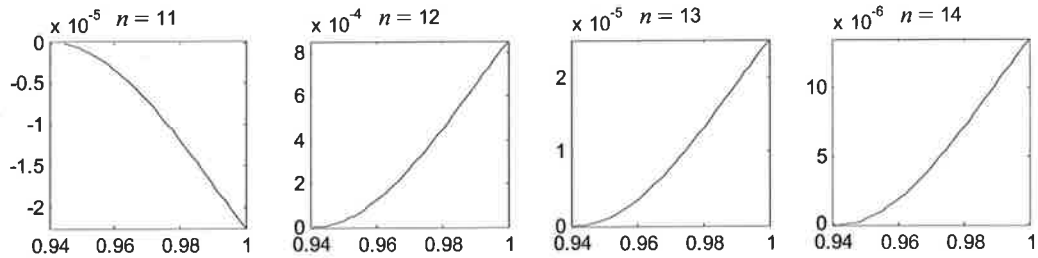
Kuvassa 4 on esitetty kertaluvun 12 deformaatioaallon voimistuminen pyörimisnopeuden kasvaessa. Kuvan 4 toisen rivin toinen ja neljäs tulostus osoittavat, että pyörimisnopeuden ohittaessa deformaatioaallon voimakkuuden kannalta kriittisimmän pyörimisnopeuden, deformaatioaallon vaihe kääntyy vastakkaiseksi. Tämä johtuu mallissa käytetystä painekuormituksesta. Mikäli käytettäisiin realistisempaa siirtymäkuormitusta, kriittisen pyörimisnopeuden ohittamisen jälkeen teloja toisiaan vasten puristava voima kasvaisi, koska nipin kohdalla kumipinnoitteessa ei voi esiintyä ulospäin suuntautuvaa radiaalisiirtymää. Tämä ilmiö erityisesti transienttitilanteessa voi toimia alussa kuvatun värähtelyn syntymekanismina. Toisaalta tämä ilmiö voi liittyä mallissa käytettyyn materiaalimalliin. Tarkastelu olisi syytä tehdä viskoelastisella materiaalimallilla.



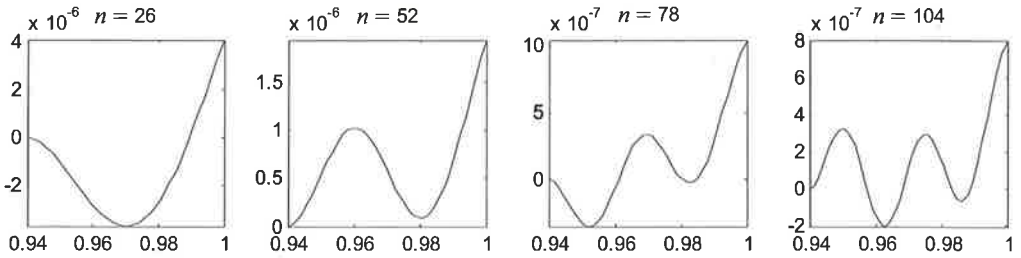
Kuva 4. Kumipinnoitteen ulkopinnan siirtymä eri pyörimisnopeuksilla. Painekuormitus (nippi) kohdassa 90 astetta (ylin piste). Radiaalisuunnassa käytetty 40 hilapistettä. $\nu = 0{,}49$, $\xi = 0{,}94$.



Kuva 5. Radiaalifunktioita $U_{n4}(r)$, radiaalisuunnassa hilapisteitä 40. $\Omega = 2{,}972$, $\nu = 0{,}49$, $\xi = 0{,}94$.



Kuva 6. Radiaalifunktioita $U_{n4}(r)$, radiaalisuunnassa hilapisteitä 40. $\Omega = 2{,}980$, $\nu = 0{,}49$, $\xi = 0{,}94$.



Kuva 7. Radiaalifunktioita $U_{n4}(r)$, radiaalisuunnassa hilapisteitä 40. $\Omega = 2{,}980$, $\nu = 0{,}49$, $\xi = 0{,}94$.

Kuvissa 5 ja 6 on esitetty funtion $U_{n4}(r)$ kuvaajat kertaluvuille 11-14, kun dimensioton pyörimisnopeus saa arvot 2,972 ja 2,980. Tällöin ylitetään kuvassa 4 näkyvä kertaluvun 12 deformaatioaallon kriittinen pyörimisnopeus. Kuten kuvista 5 ja 6 havaitaan, funktio $U_{12,4}(r)$ vaihtaa etumerkkiä kriittisen pyörimisnopeuden yli mentäessä.

Kuvissa 5 ja 6 esitetyt radiaalisiirtymäfunktiot ovat sen verran yksinkertaisesti käyttäytyviä, että esimerkiksi kumipinnoitteen fem-mallinnusta ajatellen muutama elementti riittää kumipinnoitteen paksuussuunnassa. Kuvassa 7 on esitetty radiaalisiirtymäfunktioita suuremmilla kertaluvuilla. Mikäli kriittiset kertaluvut ovat suuria, kumipinnoitteen paksuussuuntaan pitää laittaa useampia elementtejä.

NIPPIVÄRÄHTELYJEN MAHDOLLISIA SYNTYMEKANISMEJA

Laskentatulosten perusteella voidaan tunnistaa kaksi mahdollista nippivärähtelyjen syntymekanismia, kuten edellä on jo mainittu. Ensinnäkin telan vakiokulmanopeudella nippikontakti synnyttää pyörimissuuntaa vastaan etenevän deformaatioaallon, jossa tietyt kehän suuntaiset kertaluvut ovat tietyillä pyörimisnopeuksilla voimakkaita. Mikäli tällä tietyllä kertaluvulla kerrottu telan kulmanopeus osuu telan taivutusominaiskulmataajuudelle, voidaan olettaa että kyseinen taivutusominaismuoto herää. Näin käy mikäli kumin viskoelastisen käyttäytymisen vuoksi deformaatioaalto lähtee pyörimään telan mukana. Toiseksi, ylitettäessä tiettyjä kriittisiä pyörimisnopeuksia vasteen vaihe kääntyy vastakkaiseksi. Tällöin käytännössä lähinnä siirtymäohjatun nippikontaktin nippivoima kasvaa mikäli telojen välinen etäisyys ei muutu. Nippivoiman muutokset voivat herättää telan taivutusominaismuodot.

On huomattava, että käytetty malli ei osoita, että kumpikaan edellä ehdotetuista nippivärähtelyjen syntymekanismeista todella herättäisi värähtelyn. Mallin avulla saadaan kuitenkin esille ilmiöitä, jotka voivat esim. transienttitilanteessa toimia nippivärähtelyjen herättäjinä.

KÄYTETYN MALLIN PUUTTEITA JA KEHITYSKOHTEITA

Laskentamallissa kumia on mallinnettu lineaarisesti kimmoisella, homogeenisella ja isotrooppisella materiaalimallilla vaikka kumin tiedetään käyttäytyvän viskoelastisesti. Viskoelastisuudesta aiheutuva vaimennus voitaisiin lisätä laskentamalliin (yhtälöryhmään (5)) suoraan nopeuteen verrannollisina termeinä tai sitten johtaa yhtälöt käyttäen viskoelastista materiaalimallia. Vaimennusparametrit voitaisiin määrittää esim. vierimisvastuksen perusteella. Vaimennus ei kuitenkaan ilmeisesti poista systeemistä lineaarisesti kimmoisalla materiaalimallilla löydettyjä ilmiötä. Kuvissa 2 ja 3 esitetty "vuorijono" ainoastaan madaltuu ja levenee. Sen sijaan vaimennus voi aiheuttaa sen, että

deformaatioaalto lähtee pyörimään telan mukana (esim. transienttitilanteessa). Voimakkaasti epälineaarinen materiaalimalli voi tuoda esiin myös muita ilmiöitä.

Mikäli siirtymät nippikontaktissa ovat suuria, käytetyn lineaarisen kinematiikan sijasta olisi käytettävä epälineaarisia kinemaattisia yhteyksiä. Mikäli näiden mukanaan tuoma epälineaarisuus on heikkoa, mallissa käytetyllä lineaarisella kinematiikalla löydetyt ilmiöt eivät häviä, mutta epälineaarisuus voi tuoda esiin uusia ilmiöitä. Epälineaarisen kinematiikan tarve voidaan todeta mittauksin: jos esitetty malli selittää tai ennustaa mittauksilla todettavia ilmiöitä huonosti ja jos kaikki muut tekijät on otettu mahdollisimman hyvin huomioon, syynä voi olla lineaarisen kinemaattisen mallin soveltumattomuus. Toinen mahdollisuus on tehdä analyysi epälineaarisella kinemaattisella mallilla (esim. fem) ja verrata tuloksia lineaarisen mallin vastaaviin. Jos oleellisia eroja ei ole, epälineaaristen kinemaattisten yhteyksien käyttö on tarpeetonta.

Ratkaisuyritteenä on käytetty steady-state-tilan mukaista yritettä. Tällöin voidaan löytää vain steady-state-tilan ilmiöitä. Transientti-ilmiöiden tutkimiseksi ratkaisuyritteestä olisi jätettävä aikariippuvuus pois, mikä sinällään on aivan mahdollista. Saatavat osittaisdifferentiaaliyhtälöt on semidiskretoitava paikan suhteen, minkä jälkeen aikaintegrointi voidaan suorittaa numeerisesti.

Kuormitus on kuvattu mallissa kiinteästä koordinaatistosta katsottuna muuttumattomana painekuormituksena. Nippikontaktin luonteen ja deformaatioaallon kannalta kriittisen pyörimisnopeuden ylittämisen aiheuttaman vaiheen kääntymisen vuoksi realistisempaa olisi käyttää päällystettyä telaa puristavan toisen telan aiheuttamaa siirtymäkuormitusta. Koska käytetyllä mallilla voidaan laskea mitä tahansa painejakaumaa vastaava siirtymätila nipissä, käytetyn laskentamallin avulla voidaan helposti iteroimalla ratkaista siirtymäkuormitusta vastaava kosketusongelma. Kosketusongelman epälineaarisuuden vuoksi tämä voi tuoda esiin uusia ilmiöitä.

Mallista puuttuu telojen dynamiikka. Tämä voitaisiin ottaa huomioon likimääräisesti sallimalla telan/telojen liikkeet tasossa. Telan/telojen jäykkyys voitaisiin ottaa huomioon jousilla. Mikä nippivärähtelyn synnyn selittäminen vaatii transientti-ilmiöiden tarkastelua, on oletettavaa että telojen dynamiikka vaikuttaa niihin oleellisesti.

Lähestyttäessä deformaatioaallon kannalta kriittistä pyörimisnopeutta deformaatioaallon voimakkuus kasvaa. Vaimennuksen vuoksi lämmönkehitys kumipinnoitteessa voimistuu, ja lämpötilan nousu muuttaa kumin viskoelastisia ominaisuuksia, mitkä puolestaan vaikuttavat kriittiseen pyörimisnopeuteen. Tällaisen kytketyn ongelman ratkaisemiseksi materiaalimallissa pitäisi olla lämpötilariippuvuus mukana ja lämpötilakenttä pitäisi ratkaista samanaikaisesti deformaatiokentän kanssa. Kynnyskysymyksenä tässä suhteessa on realistisen lämpötilariippuvan viskoelastisen materiaalimallin muodostaminen ja sen parametrien mittaaminen.

Mallin kehittämisen kannalta helppoja kehityskohteita ovat viskoelastinen materiaalimalli, siirtymäkuormitus ja transienttitilan tarkastelu, vaativampia telojen dynamiikan huomioiminen sekä kytketyn ongelman formulointi ja ratkaisu. Epälineaarisen kinematiikan tarpeellisuutta voi testata vertaamalla esitetyn mallin tuloksia epälineaarisen kinematiikan huomioonottaviin numeerisiin tuloksiin.

Jatkotutkimuksia voisi tehdä puhtaasti numeerisilla menetelmillä. Esitetty analyyttis-numeerinen menettelytapa on kuitenkin laskenta-ajan suhteen ylivoimainen, mikä mahdollistaa huomion keskittämisen nippikontaktin yhteydessä esiintyviin fysikaalisiin ilmiöihin. Puhtaasti numeerisilla menetelmillä voidaan päätyä tutkimaan, pystytäänkö ratkaisemaan miljoona vapausastetta sisältävä malli, jossa tutkittavien ilmiöiden kannalta on 999 900 ylimääräistä vapausastetta (ja itse ilmiöihin ei edes päästä käsiksi). Analyyttisiä menetelmiä ei pitäisi hylätä myöskään siksi, että jotkut pitävät niitä vaikeina ja niiden oppimiseksi on tehtävä työtä.

YHTEENVETO

Kumipinnoitteen siirtymäkentän analyyttis-numeerinen ratkaisu käytetyllä fysikaalisella mallilla steady-state-tilanteessa osoittaa että:

1. Nippikontakti ei lähetä deformaatioaaltoja telan pyörimisen suuntaan.
2. Telan pyörimissuuntaan nähden vastakkaiseen suuntaan etenevä deformaatioaalto pysyy paikallaan kiinteästä koordinaatistosta käsin tarkasteltuna.
3. Nippikontaktin lähettämän voimakkaimman deformaatioaallon kertaluku kehän suunnassa riippuu telan pyörimisnopeudesta.
4. Joillakin pyörimisnopeuksilla nippikontaktin lähettämä deformaatioaalto vahvistuu selvästi.
5. Analyyttiset ratkaisumenetelmät eivät ole vanhentuneet, vaan antavat mahdollisuuden syventää erilaisten ilmiöiden ymmärtämistä.

Tämä tutkimus on tehty keväällä 2003. Lamen yhtälöiden napakoordinaatistomuoto on peräisin prof. Hannu Outisen luentomonisteista. Suoranaisesti mitään muita lähteitä ei käytetty. Ajatus paikallaan olevan kuormituksen alla pyörivän telan pinnoitteen dynaamisen deformaatiokentän ratkaisemisesta lienee peräisin Journal of Sound and Vibration –lehdessä julkaistuista artikkeleista, joissa on käsitelty ympyrärenkaan dynamiikkaa kaarevan palkin teorialla, kun kuormituksena on renkaalla liikkuva radiaalikuormitus.

Tämän tutkimuksen valmistumisen jälkeen muissa yhteyksissä havaittiin, että samaa ongelmaa on käsitelty numeerisilla menetelmillä jo 70-luvun lopussa ja 80-luvun alussa, joista mainittakoon /1, 2/. Koska näissä artikkeleissa esimerkkilaskelmien mallit edellyttävät kaksikerroksista pinnoitetta, tulosten kvantitatiivinen vertailu ei ole mahdollista, mutta kvalitatiivisella tasolla tarkasteltuna artikkeleissa /1, 2/ ja tässä tutkimuksessa löydetyt ilmiöt täsmäävät. Tulosten saamiseen tarvittavaa työmäärää verrattaessa tässä tutkimuksessa esitetty analyyttis-numeerinen menettelytapa on ylivertainen.

VIITTEET

1. Zeid, I., Padovan, J. *Finite element modeling of rolling contact*, Computers & Structures, Vol. 14, No. 1-2, pp. 163-170, 1981
2. Padovan, J., Paramadilok, O. *Transient and steady state viscoelastic rolling contact*, Computers & Structures, Vol. 20, No. 1-3, pp. 545-553, 1985

VIBRATION ANALYSIS OF NONLINEAR MODEL FOR NIP CONTACT

L. Yuan

Tampere University of Technology, Laboratory of Machine Dynamics
lihong.yuan@tut.fi

ABSTRACT

This paper presents modelling and analysis of a nip contact between two paper machine rolls as two degrees of freedom lumped mass systems including quadratic and cubic non-linear contact stiffness terms and a time delay. The two rolls studied here are made of cast iron. One roll has polymer cover on its surface. Due to the finite relaxation characteristics from the polymer viscoelastic material the time delay effect has been involved in the model. The nonlinear terms originate from the hyperelastic stiffness of the polymer material. Therefore numerical methods such as Poincare maps have been applied for of analysing the nonlinear system equations. Numerical results in case studies will illustrate the dynamic characteristics of the nip contact.

1. INTRODUCTION

A roll system with nip contact is studied in this work. The system is related to paper machine which has multiple roll nips used for finishing the surface of the paper during the paper manufacturing process in paper mills. A test nip installation in Tampere University of Technology (TUT) laboratory environment has been built for research purposes and it is half scale of a real sized industrial nip unit. The vibration behaviours of the nip contact have not been understood completely and the complex solutions are still under investigations. The main elements of the system are one polymer cover roll and one pure metal roll in the nip contact. The spindles of both rolls will be running in speeds about 500 rpm (in real machine, the speed can reach 1800rpm). The soft polymer cover enables smooth processing of the paper. But the consequence is that it may cause the time delay effect due to the non-recovered polymer material deformations in the rolling contact. The stability of time delay system is more complex and interesting especially when it involves nonlinear elements [Stepan 1989].

This aim of this work is mainly to investigate the nonlinear dynamic behaviour in the rotor system [Yamamoto 2001, Ehehalt 2002, Genta 2005]. The model of the system will be given with considering the time delay effect. The rolls are considered as analytical rotors with contact stiffness in the nip. The nonlinear terms come from the hyperelastic material behaviour of the polymer. The numerical simulations have been carried out with lumped parameters. The bifurcation and chaos behaviours of the system may be illustrated by Poincare maps [Jordan 1987].

In the results, firstly, the transient time domain response of the system illustrates the time delay effect due to the feedback deformation. The rotational speeds of rolls are key parameters for the vibration analysis because the time delay changes with the speed. The vibration responses of the

system show non-classical resonances due to the delay effect. Secondly, the time delay and contact stiffness will give influences on the bifurcation behaviour.

2. MODELLING OF NIP CONTACT

The main components of the TUT nip installation are two metallic rolls (Figure 1) manufactured of cast iron. Both rolls have been designed identical with two exceptions: the upper one has soft polymer cover (11 mm thickness) on its surface and the lower one has just a hard metal surface. This kind of combination is typical because of the process quality requirement of the paper manufactured. The contact length of one roll is about 4.4 m, the diameter is 500 mm and the weight about 3 tons. The dynamic time delay effect originates from the polymer material of the soft roll. At the beginning, the two-roll are driven to the desired web speed and then the contact is closed. If the polymer cover cannot recover from the contact deformation completely after each round contact undesired dynamic behaviour can occur. The delay effect will be carried into the next round and provide an additional source of vibrations, which cannot be ignored when the dynamical behaviour of the installation is considered.

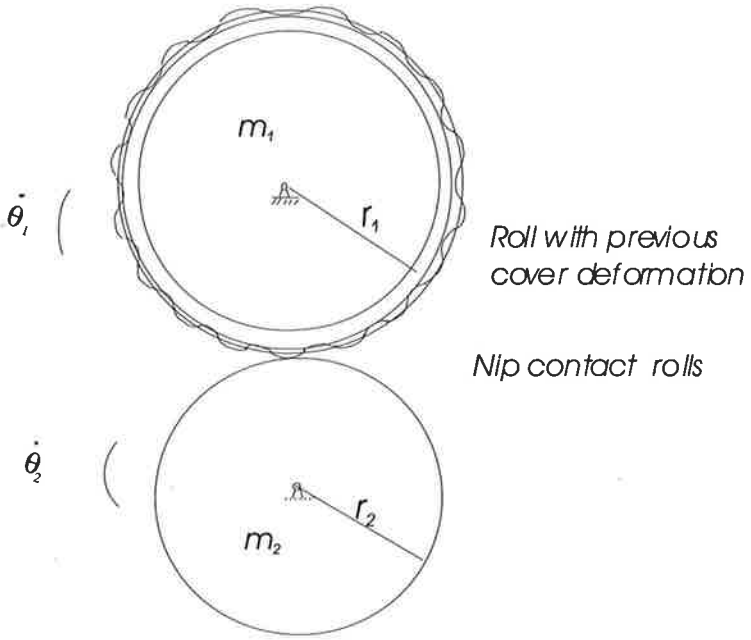


Figure 1. The roll installation system with nip contact with internal excitation from the previous rotation.

Dynamics of the two rolls and their interaction can be modeled by an equivalent spring-mass system (Figure 2).

The vertical motion of the system is considered here, thus the displacments of the centre line of the two rolls are denoted as x_1, x_2. Applying the Newton's second law to each mass in Figure 2, then the governing two-degree-of-freedom equations of motion can be written as

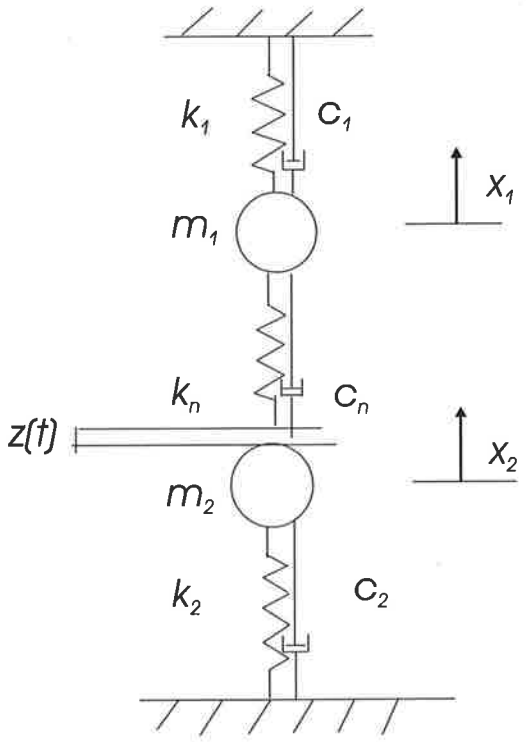


Figure 2. The equivalent spring-mass system.

$$\ddot{x}_1 + 2\zeta_1\dot{x}_1 + {p_1}^2(x_1 + \beta_{12}{x_1}^2 + \beta_{13}{x_1}^3) = -N(x_1, x_{1\tau})/m_1 \tag{1}$$

$$\ddot{x}_2 + 2\zeta_2\dot{x}_2 + {p_2}^2(x_2 + \beta_{22}{x_2}^2 + \beta_{23}{x_2}^3) = N(x_1, x_{1\tau})/m_2 \tag{2}$$

where $2\zeta_i = c_i/m_i$, ${p_i}^2 = k_i/m_i$, in which paramters m_i, c_i, k_i for $i=1, 2$, refer to the roll mass, damping and stiffness, respectively; β_{i2} , β_{i3} are constants describing the nonlinear stiffness of the nip system. The note $x_{1\tau} = x_1(t-T)$ refers to the term with time delay effect. The contact force takes the form

$$N = k_n\left[\varepsilon(t) - \gamma_e\varepsilon(t-T) + \alpha_2(\varepsilon(t) - \gamma_e\varepsilon(t-T))^2 + \alpha_3(\varepsilon(t) - \gamma_e\varepsilon(t-T))^3\right] + c_n\dot{\varepsilon}(t) \tag{3}$$

Here the nip force accounts for quadratic and cubic nonlinearities [as in Hanna & Tobias, 1974]. And α_2 , α_3 are nonlinear constant coefficients. The k_n is the spring coefficient and c_n is the damping coefficient for a Kelvin-like polymer roll covering material [Sperling,1985] for the nip contact. The decay factor duo to the elastic effect is discribed by an exponent function like

$$\gamma_e = e^{-Tk_n/c_n} \tag{4}$$

where the note T is the revolution time of the polymer roll, which is assumed as the constant time delay term of the response. The nip penetration on the surface of the polymer covered roll includes three components

$$\varepsilon(t) = x_1(t) - x_2(t) + z(t) \tag{5}$$

The source term $z(t)$ in (5) accounts for the shape profile of the material entering the nip; in real machine, this source also includes the paper web transtered through the nip. However, for the sake of testing simplication, this factor is not consider in this study. This excitation could be described by a periodic function

$$z(t) = Z\sin(\Omega t) \tag{6}$$

With the excitation frequency $\Omega = 2\pi j_p / T$, in which the note j_p is the number of wave resulted from the error shape of the polymer roll surface.

3. ANALYSIS METHODS

In the equations (1)-(2), there are time delay and nonlinear terms. The delay differential equations are more complicated than the ordinary ones but the methods for analysis the general nonlinear still can apply. For instance, the Lyapunov exponent (quantity index) and Poincare map (quality way) might be proper methods to be used.

Firstly, the Lyapunov exponent λ is defined as [Hilborn1994]

$$\lambda = \frac{1}{n}(\ln|f'(x_0)| + \ln|f'(x_1)| + \ldots + \ln|f'(x_{n-1})|) \tag{7}$$

where n is the iteration number. In the equation (7), the Lyapunov exponent λ is the average of the natural logarithm of the absolute value of the derivatives of the map function evaluated at the trajectory points. According to the nonlinear theory, the chaos occurs when the Lyapunov exponent is positive. From the time series in (7), the plots of the Lyapunov exponent can be obtained for the nip system. However, for the time delay system, it is much complicated to solve the Lyapunov exponent numerically.

Secondly, the Poincare section is a device to simplify phase space diagrams of complicated systems. It is constructed by viewing the phase space diagram stroboscopically in such a way that the motion is observed periodically. For the time delay system, having chaotic or limit cycle behaviour does not violate the Poicare theory. Based on this, Poincare maps could be plotted for the nip delay system.

4. SIMULATION AND NUMERICAL RESULTS

By using the Simulink in MATLAB® environment the simulation model of the nip installation can be set up for the time domain simulations (Figure 3). Introducing again the real-scale system parameters, the time domain simulations in different cases have been carried out. In the nip installation, the nip stiffness and the time delay are important parameters. So the effects of nip

stiffness and effects of the time delay on the chaos behaviour of the nip system have been demonstrated as below.

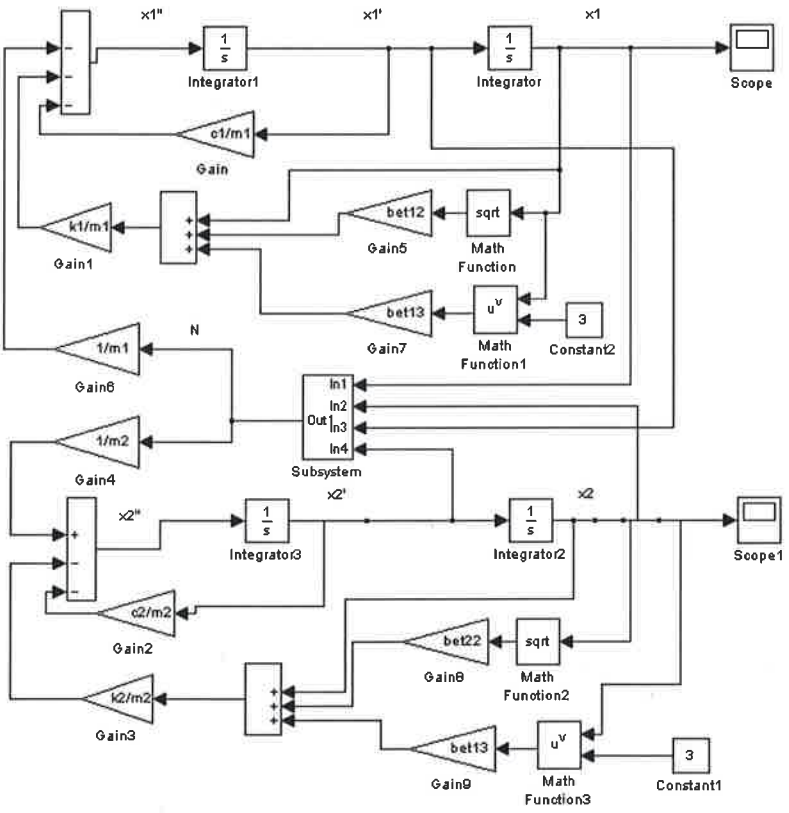


Figure 3. Simulink block diagram of the main system.

In Figures 4 to 5, the time response of x_1, phase trajectory and Poincare map have been plotted. In Figure 4 to 5 the nip stiffness keeps a same value k_n=*1e+7 N/m*, but the time delay decreases from *T=1/3.95 s* ,to *1/4 s*, then the Poincare sections illustrate in different fractal ways.

In Figure 6 to 7 , the time delay has a same value *T=1/4 s*, but the stiffness increases from k_n=*1e+7 N/m,* to *3e+7 N/m*, then the Poincare sections show again different fractal images. In Figure 6, a chaotic attractor emerges in the Poincare sections.

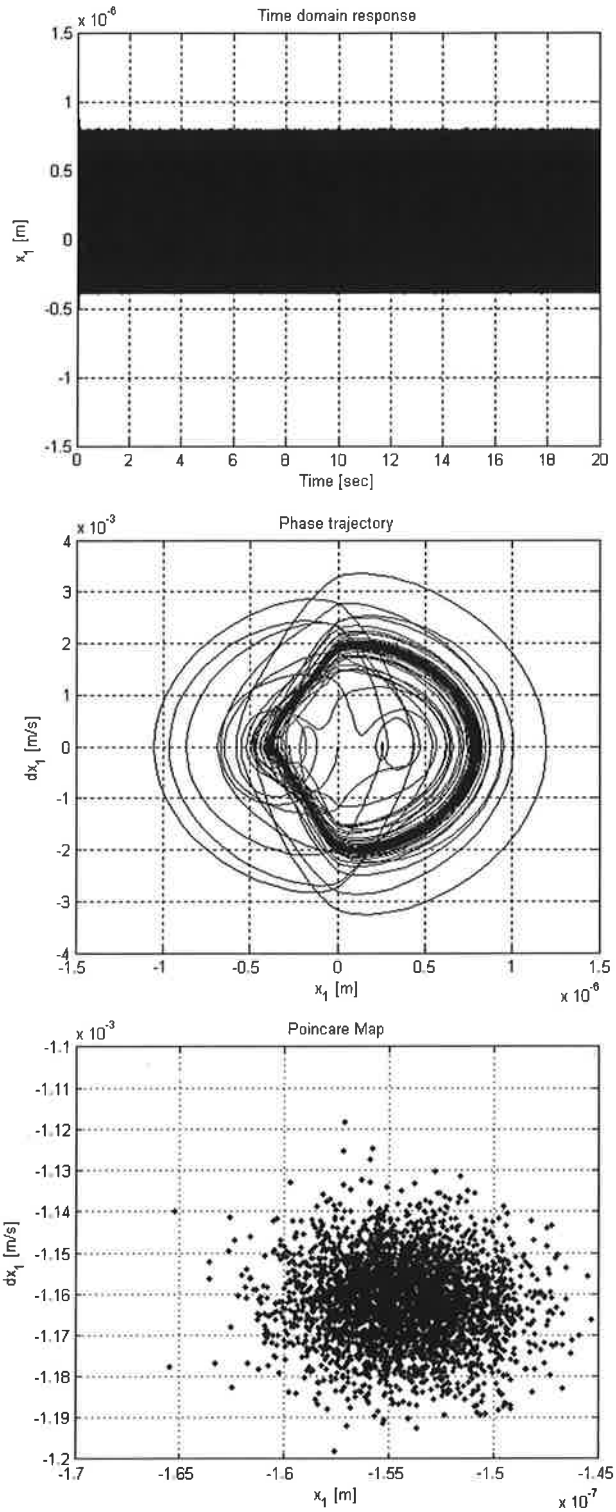


Figure 4. *k_n=1e+7 N/m, T=1/3.95 s:* (a) Time response; (b) Phase trajectory; (c) Poincare map.

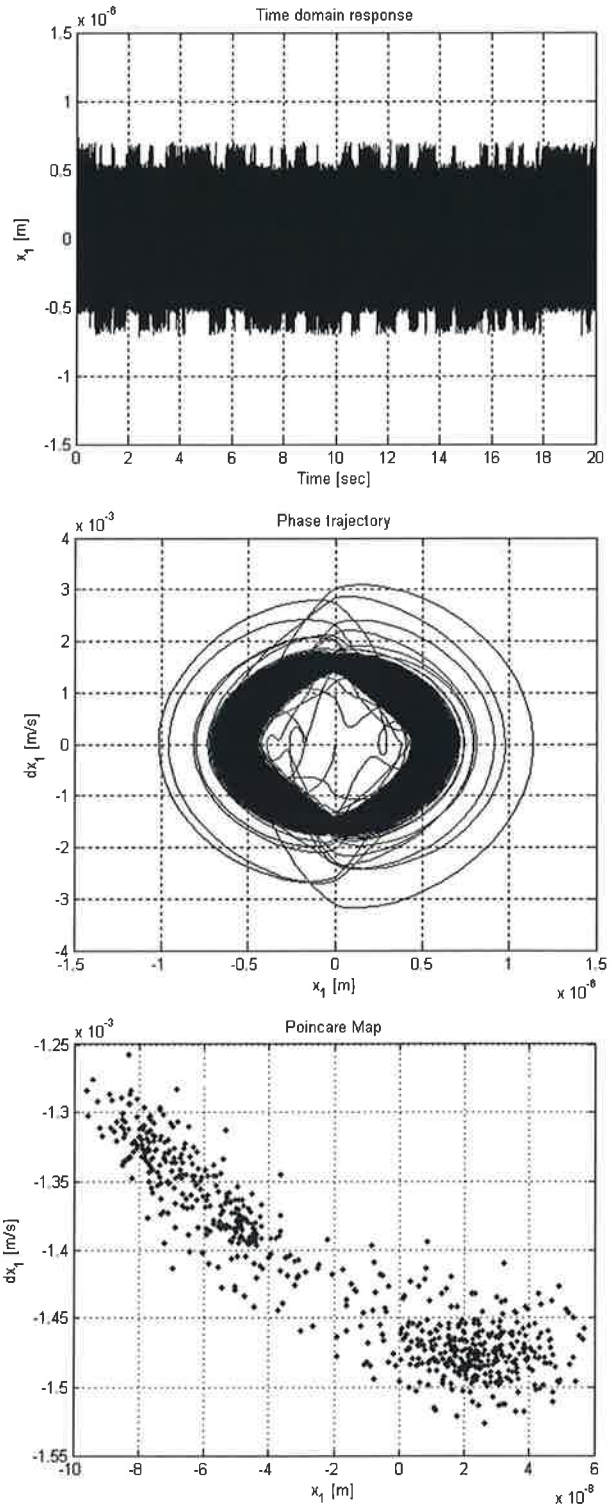


Figure 5. *k_n=1e+7 N/m, T=1/4 s* (a) Time response; (b) Phase trajectory; (c) Poincare map.

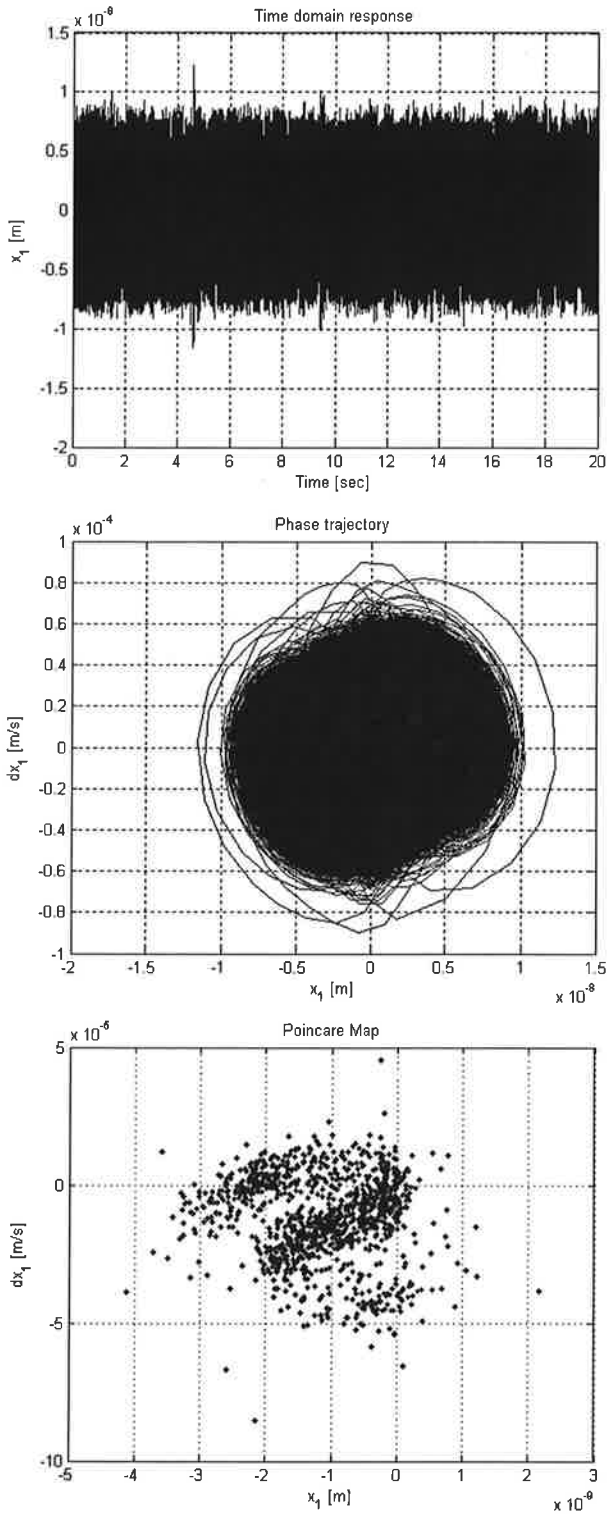


Figure 6. k_n=*1e+6 N/m, T=1/4 s* (a) Time response; (b) Phase trajectory; (c) Poincare map.

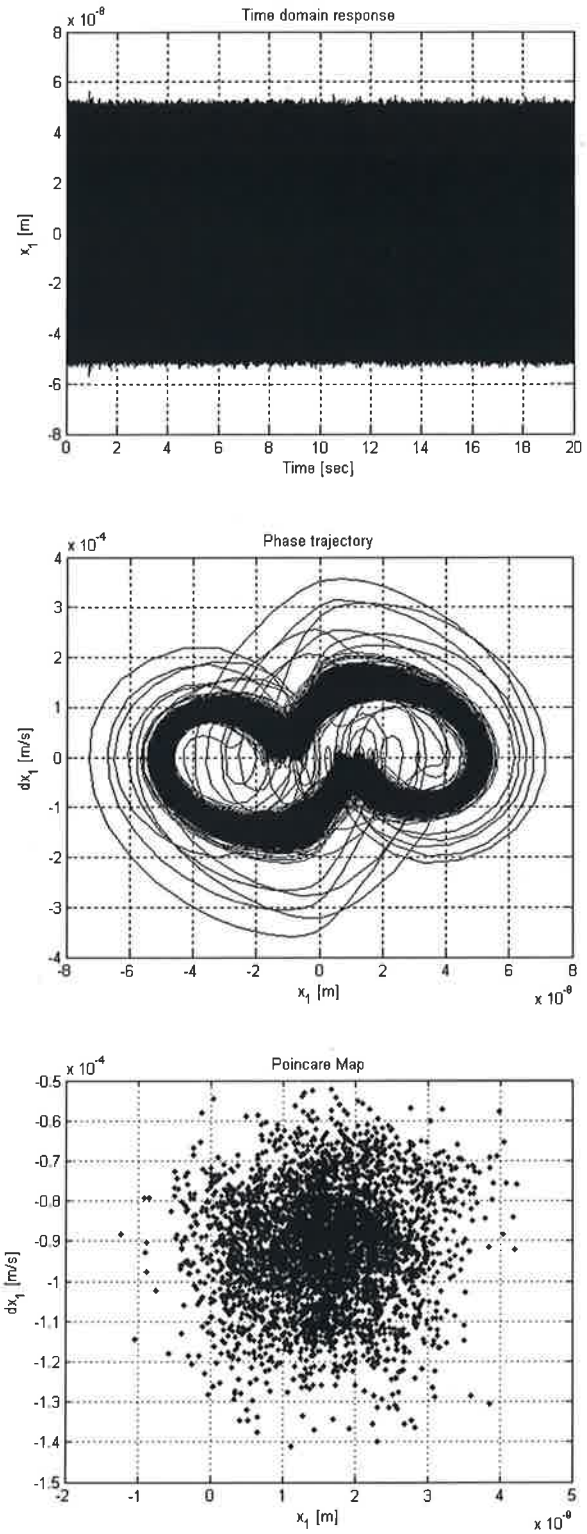


Figure 7. *k_n=3e+6 N/m, T=1/4 s* (a) Time response; (b) Phase trajectory; (c) Poincare map.

5. CONCLUSION

The model of the nip installation has been established including time delay and nonlinear contact stiffness terms. The nonlinear methods have been applied for analysing the nip system. The numerical time domain simulations show chaotic behaviours. The results give some theoretical evidences, which can be used in the development of strategies to control the vibrations of the system

6. ACKNOWLEDGEMENT

The author would like to express gratitude to the Academy of Finland, which has provided funding for this project.

REFERENCES

Genta G., *Dynamics of Rotating Systems*, *Mechanical Engineering Series*, Springer, 2005.
Jordan D. W. & Smith P., *Nonlinear Ordinary Differential Equations*, Oxford Univ. Press, 1987.
Ehehalt U. & Markert R, *Rotor Motion During Stator contact*, 6th Int. Conf. on Rotor Dynamics proceedings, 2002.
Stepan G., *Retared Dynamicl Systems: Stability and Characteistic Functions*, John Wiley & Sons, Inc., 1989.
Yamamoto T. & Ishida Y., *Linear and Nonlinear Rotordyanamics*, Wiley&Sons, Inc., 2001.
Hanna N.H. & Tobias S.A., *A Theory of Nonlinear Regenerative Chatter*, ASME J.Engng Industry 96, 1974.
Hilborn R.C., *Chaos and Nonlinear Dynamics, An Introduction for Scientists and Engineers*, Oxford University Press, 1994.
Sperling L.H., *Introduction to Physical Polymer Science*, Wiley-Interscience, New-York, 1985.

COUPLING BETWEEN LONGITUDINAL VIBRATIONS OF AN AXIALLY MOVING WEB AND ROTATIONAL VIBRATIONS OF ROLLS WITH TIME DEPENDENT PARAMETERS

TERO FRONDELIUS
University of Oulu
Department of Mechanical Engineering
Engineering Mechanics Laboratory
P.O. Box 4200, FIN-90014 University of Oulu
e-mail: Tero.Frondelius@oulu.fi

ABSTRACT

The aim of this study is to model the coupling between rotational vibrations of the rolls and longitudinal vibrations of an axially moving web. These vibrations are also known as changes in stress distribution. The web is treated as an axially moving rod. At the boundaries rolls are modelled such that the change in radius is taken into account, which makes it weakly non-linear. The mass flow out of the first roll and into the last one is correspondingly considered. In paper industry this is called unwinding and winding, respectively. A magnetic tape, paper and aluminium webs are chosen for this study, because they are widely used and are therefore good and economically interesting materials to study for. The equation of motion and the boundary conditions are derived using the extended Hamilton's principle. The longitudinal frequencies of the axially moving bar is calculated numerically. Moreover, the forced vibrations of the time dependent model is solved by perturbation method and by time integration.

NAVIGOINTIA PALLOLLA

Matti A Ranta
Teknillinen korkeakoulu, Matematiikan laitos,
PL 1100, 02015 Espoo
matti.ranta@tkk.fi

Raimo von Hertzen
Lappeenrannan Teknillinen Yliopisto, Rakenne-ja lujuustekniikan laboratorio,
PL 20, 53851 Lappeenranta
rhertzen@lut.fi

TIIVISTELMÄ

Artikkelissa käsitellään loksodromi- ja isoympyräpurjehduksen kinematiikkaa. Lähtien aluksen kinemaattisista differentiaaliyhtälöistä johdetaan loksodromi keskilatitudimenetelmällä, tarkasti integroitu loksodromi sekä analyyttiset lausekkeet aluksen latitudille ja longitudille kuljetun matkan funktiona isoympyräpurjehduksessa. Viimeksi mainituista johdetaan edelleen aluksen kulkusuunnan ja kääntymisnopeuden lausekkeet kuljetun matkan funktiona isoympyräpurjehduksen jatkuvaa ohjausta varten.

JOHDANTO

Tunnetusti jokaisella riittävän sileällä pinnalla lyhin tie pinnan kahden pisteen välillä kulkee pitkin geodeettista viivaa. Pallon pinnalla geodeettiset viivat ovat isoympyröitä ja lyhin tie kahden pisteen välillä on siis näiden pisteiden kautta kulkeva isoympyrän kaari. Käytännössä aluksen navigointi suoritetaan ohjaamalla alusta joka hetki tiettyyn suuntaan. Suunta mitataan yleensä kullakin hetkellä aluksen kautta kulkevaan meridiaaniin nähden. Koska kurssin jatkuva muuttaminen varsinkin manuaaliohjauksessa on työlästä, on isoympyräpurjehduksen sijasta jo pitkään käytetty yksinkertaisempaa loksodromipurjehdusta, jossa aluksen suunta meridiaaniin nähden säilyy vakiona. Kahden pisteen välisessä loksodromipurjehduksessa hävitään kuitenkin aina matkassa verrattuna vastaavaan isoympyräpurjehdukseen. Kompassi on ollut navigoinnissa jo yli 1000 vuotta suunnan mittauksen tavallisin apuväline. Sen avulla mitataan kulmaa aluksen sijaintipaikan kautta kulkevan meridiaanin ja aluksen kulkusuunnan välillä. Nykyään voidaan suunta määrittää myös GPS-laitteiden avulla. Pelkistetysti voidaan sanoa, että purjehdus suoritetaan seuraavasti:

- Lyhyillä matkoilla kuljetaan vakiosuuntaan eli pitkin *loksodromia*. Tällöin navigointi on helppoa eikä matkan pitenemisestä ole käytännössä haittaa. Loksodromipurjehdusta voidaan tarkentaa suorittamalla sarja peräkkäisiä lyhyempiä loksodromipurjehduksia, joiden alku- ja loppupisteet (ns. reittipisteet) sijaitsevat alkuperäisellä isoympyrällä.

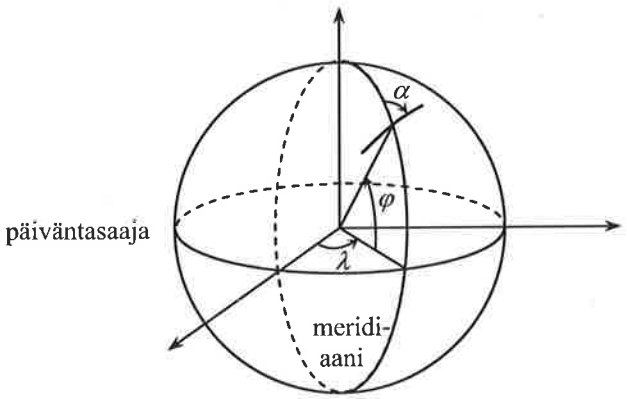
- Pitkillä matkoilla kannattaa pyrkiä seuraamaan lyhintä reittiä eli kulkemaan pitkin *isoympyrää*. Tällöin suuntakulma muuttuu jatkuvasti, mikä on ollut vaikea käytännössä toteuttaa.

Tässä työssä esitetään kinemaattisten differentiaaliyhtälöiden integrointi sekä loksodromi- että isoympyräpurjehduksen tapauksessa. Molemmissa tapauksissa yhtälöt onnistutaan integroimaan suljetussa muodossa ja ratkaisuille saadaan parametriesitykset kuljetun matkan s funktiona.

Eliminoimalla parametri s päädytään latitudin ja longitudin väliseen integraalikäyrään, joka määrittää aluksen reitin maapallon pinnalla. Isompyräpurjehdukselle on johdettu lisäksi analyyttinen lauseke aluksen kulkusuunnan jatkuvalle muuttamiselle ja aluksen kääntymisnopeudelle kuljetun matkan funktiona. Tässä artikkelissa johdettuja isompyräpurjehdusta koskevia analyyttisiä tuloksia ei kirjallisuudessa esiinny. Huolimatta siitä, että loksodromipurjehduksen kaavat sekä isompyräpurjehduksen pallotrigonometriset kaavat ovat olleet tunnettuja jo pitkään, niiden tiimoilta käydään jatkuvaa keskustelua maailmalla. Sitä osoittavat artikkelien runsaus vuosina 1980-2004 julkaisussa "*The Journal of Navigation*": loksodromia koskevia 3kpl, isompyrää koskevia 6kpl sekä loksodromi/isompyrä-vertailuja 2kpl. Myös maapallon litistyneisyys on otettu huomioon monessa artikkelissa.

ALUKSEN LIIKEYHTÄLÖT PALLOLLA

Maapallolla navigoitaessa pallokoordinaatisto on luonnollinen valinta. Siinä yksi koordinaatti, *latitudi* φ, mitataan päiväntasaajasta pohjoiseen, jolloin $-90° \le \varphi \le +90°$ ja toinen koordinaatti, *longitudi* λ, lännestä itään, jolloin $-180° \le \lambda < +180°$. Pallon pinnalla liikkuvan aluksen kinemaattiset differentiaaliyhtälöt voidaan johtaa helposti tarkastelemalla pallolla äärettömän pientä kolmiota, jota voidaan pitää tasokolmiona. Sen yksi kateetti yhtyy etelä-pohjois-suuntaan eli meridiaanin $\lambda = vakio$ suuntaan ja toinen kateetti länsi-itä-suuntaan eli latitudin $\varphi = vakio$ suuntaan. Itse alus liikkuu pitkin hypotenuusaa suuntaan α meridiaaniin nähden. Kulma mitataan pohjoisesta idän kautta aluksen kulkusuuntaan ($0° \le \alpha < 360°$) (ks. kuva 1.). Kaikki suureet on parasta esittää dimensiottomassa muodossa eli radiaaneina. Todelliset matkat saadaan kertomalla dimensiottomat matkat maapallon säteellä R.



Kuva 1. Pallokoordinaattien käyttö navigoinnissa. Kulma α ilmaisee aluksen kulkusuunnan.

Aluksen ollessa pisteessä $P(\varphi,\lambda)$ ja sen kulkiessa matkan ds suuntaan α muuttuvat sen koordinaatit, latitudi φ ja longitudi λ, infinitesimaalisilla määrillä. Kinemaattiset differentiaaliyhtälöt kuuluvat

$$\frac{d\varphi}{ds} = \cos\alpha \qquad (1)$$

$$\frac{d\lambda}{ds} = \frac{\sin\alpha}{\cos\varphi} \qquad (2)$$

Nämä ovat navigoinnin merkintälaskun perustana olevat yhtälöt. Jakamalla yhtälöt (1) ja (2) keskenään saadaan

$$\frac{d\lambda}{d\varphi} = \frac{\tan\alpha}{\cos\varphi} \tag{3}$$

Jotta nämä differentiaaliyhtälöt voitaisiin integroida, täytyy tietää miten suuntakulma α muuttuu matkan edistyessä.

PURJEHDUS PITKIN LOKSODROMIA (α = vakio)

Loksodromi keskilatitudimenetelmällä

Jos suuntakulma $\alpha \equiv k$ on vakio, kulkee alus pitkin loksodromia ja sen koordinaattien muuttuminen voidaan helposti laskea. Kaava (1) voidaan integroida suoraan ja kaava (3) integraalilaskennan väliarvolausetta soveltaen likimäärin, pitämällä $\varphi = \Phi$ vakiona. Kun matka lasketaan päiväntasaajalta eli $s = 0$ ekvaattorilla, saadaan kaavoista (1) ja (2) latitudi ja longitudi

$$\varphi = s\cos k \tag{4}$$

$$\lambda = s\frac{\sin k}{\cos\Phi} + \Lambda_k \tag{5}$$

Integroimisvakio Λ_k on loksodromikäyrän ja päiväntasaajan leikkauspisteen longitudi. Pisteiden $P(\varphi_1, \lambda_1)$ ja $P(\varphi_2, \lambda_2)$ väliselle loksodromille pätee

$$\varphi_1 - s_1\cos k = \varphi_2 - s_2\cos k$$

$$\lambda_1 - s_1\frac{\sin k}{\cos\Phi} = \lambda_2 - s_2\frac{\sin k}{\cos\Phi} = \Lambda_k$$

Näistä kaavoista saadaan tulos

$$\frac{\tan k}{\cos\Phi} = \frac{\lambda_2 - \lambda_1}{\varphi_2 - \varphi_1} \tag{6}$$

Suuntaa k ja välilatitudia Φ ei saada ratkaistua ilman lisäehtoa, joksi yleisesti valitaan ns. keskilatitudi (päiväntasaaja ei saa olla välillä (φ_1, φ_2) eikä erotus $|\varphi_2 - \varphi_1|$ liian suuri) eli

$$\Phi = (\varphi_2 + \varphi_1)/2 \tag{7}$$

Tällöin etsitty loksodromisuunta saadaan kaavasta (6)

$$k = \arctan\left[\frac{(\lambda_2 - \lambda_1)\cos\Phi}{\varphi_2 - \varphi_1}\right] \tag{8}$$

Loksodromimatka saadaan kaavojen (4) ja (5) avulla

$$d = s_2 - s_1 = \sqrt{(\varphi_2 - \varphi_1)^2 + \left[(\lambda_2 - \lambda_1)\cos\Phi\right]^2} \equiv \frac{\varphi_2 - \varphi_1}{\cos k} \tag{9}$$

Eliminoimalla matka s kaavoista (4) ja (5) voidaan ratkaista joko latitudi longitudin funktiona tai päinvastoin, jolloin saadaan

$$\varphi = (\lambda - \Lambda_k)\frac{\cos\Phi}{\tan k} \tag{10}$$

$$\lambda = \Lambda_k + \frac{\tan k}{\cos\Phi}\varphi \tag{11}$$

Keskilatitudiloksodromi on täten täysin määritetty. Vertaa [1] kappale Merkintälasku.

Tarkasti integroitu loksodromi

Koska $\alpha \equiv K$ on vakio, saadaan kaavat (1) ja (3) integroimalla tulokset

$$\varphi = s\cos K \tag{12}$$

$$\lambda = \tan K \, Ar\sinh\left[\tan(s\cos K)\right] + \Lambda_l \tag{13}$$

Integroimisvakio Λ_l on tarkan loksodromikäyrän ja päiväntasaajan leikkauspisteen longitudi. Integrointi $\int d\varphi/\cos\varphi$ kaavassa (3) on lähes standarditehtävä, katso [5] sivu 166 kohta 9. Mutta sen edelleen sieventämisestä katso [2] sivu 256 tai [4] sivu 54. Pisteiden $P(\varphi_1, \lambda_1)$ ja $P(\varphi_2, \lambda_2)$ väliselle tarkalle loksodromille pätee

$$\varphi_1 - s_1\cos K = \varphi_2 - s_2\cos K$$

$$\lambda_1 - \tan K \, Ar\sinh(\tan\varphi_1) = \lambda_2 - \tan K \, Ar\sinh(\tan\varphi_2) = \Lambda_l$$

Näistä voidaan ratkaista matka d

$$d = s_2 - s_1 = \frac{\varphi_2 - \varphi_1}{\cos K} \tag{14}$$

ja suunta K

$$K = \arctan\left[\frac{\lambda_2 - \lambda_1}{Ar\sinh(\tan\varphi_2) - Ar\sinh(\tan\varphi_1)}\right] \tag{15}$$

Kaavat (12) ja (13) voidaan ratkaista joko latitudin tai longitudin suhteen, jolloin saadaan vaihtoehtoiset yhtälöt pisteen $P(\varphi, \lambda)$ kautta suuntaan K kulkevalle tarkalle loksodromille

$$\varphi = \overline{arc}\tan\left\{\sinh\left[(\lambda - \Lambda_l)\cot K\right]\right\} \tag{16}$$

$$\lambda = \Lambda_l + \tan K \, Ar\sinh(\tan\varphi) \tag{17}$$

Jos $\cos K = 0$, on keskilatitudi-loksodromi tarkka. Loksodromi on täten täysin määritetty. Kaavasta (16) nähdään, että jos alus kiertäisi maapalloa vakiosuuntaan $K \neq 0°$, jäisi tarkka loksodromi kiertämään napaa saavuttamatta sitä koskaan.

ISOYMPYRÄPURJEHDUS

Isoympyrä on lyhin tie kahden pisteen välillä pallolla. Sillä on seuraava ominaisuus, katso [3] sivu 72

$$\sin\alpha\cos\varphi = \cos\phi_v \quad (= \text{vakio}) \tag{18}$$

Tässä ϕ_v on niin sanotun vertex-pisteen latitudi. Vertex-pisteessä on voimassa $\alpha = 90°$ tai $270°$.

Koska suuntakulma α muuttuu latitudin funktiona kaavan (18) mukaan, tulee tästä ehdosta ensin

$$\cos\alpha = \pm\sqrt{1-\sin^2\alpha} = \pm\sqrt{\cos^2\varphi - \cos^2\phi_v}\,/\cos\varphi \equiv \pm\sqrt{\sin^2\phi_v - \sin^2\varphi}\,/\cos\varphi \tag{19}$$

ja sitten

$$\tan\alpha = \frac{\sin\alpha}{\pm\sqrt{1-\sin^2\alpha}} \equiv \frac{\cos\phi_v}{\pm\sqrt{\cos^2\varphi - \cos^2\phi_v}} \equiv \frac{\cos\phi_v}{\pm\sqrt{\sin^2\phi_v - \sin^2\varphi}} \tag{20}$$

Neliöjuuren etumerkin valinta on aina erikseen tutkittava samoin kuin arcus-funktioiden haara. Jos vertex-piste on lähtö- ja tulopisteiden välissä, kannattaa vertex-pistettä lähestyä sekä vasemmalta että oikealta ja valita kaavoissa (19) ja (20) oikeat etumerkit. Jos merkki ei selviä, sitä ei saa arvata vaan on käytettävä jotain sopivaa arctan-kaavaa.

Sijoittamalla lausekkeet (19) ja (20) kaavoihin (1) ja (2) sekä muuttujat erottamalla saadaan

$$ds = \frac{\cos\varphi d\varphi}{\sqrt{\sin^2\phi_v - \sin^2\varphi}}$$

$$d\lambda = \frac{\cos\phi_v \, d\varphi}{\cos\varphi\sqrt{\sin^2\phi_v - \sin^2\varphi}} \tag{21}$$

Suoritetaan uusi sijoitus

$$\sin\varphi = \sin\phi_v \sin\psi \qquad \text{eli} \qquad \psi = \arcsin(\sin\varphi/\sin\phi_v) \tag{22}$$

Differentioimalla tästä seuraa

$$\cos\varphi d\varphi = \sin\phi_v \cos\psi d\psi \tag{23}$$

Yhtälöistä (21) tulee sijoituksen ja sievennysten jälkeen

$$ds = d\psi$$
$$d\lambda = \frac{\cos\phi_v \, d\psi}{1-\sin^2\phi_v \sin^2\psi} \tag{24}$$

Ensimmäisen integrointi on helppo. Lasketaan kaarenpituus s päiväntasaajasta, jolloin integroimisvakio häviää. Kaavan (22) avulla tulos on

$$s = \psi = \overline{\text{arc}}\sin(\sin\varphi/\sin\phi_v) \tag{25}$$

Lähteen [5] sivun 185 kohdan 2.562 kaavan 1 mukaan tulee toisesta integroitavasta (24) tulos

$$\lambda - \arctan(\cos\phi_v \tan\psi) = \Lambda_i \tag{26}$$

Tässä vakio Λ_i on isoympyrän ja päiväntasaajan leikkauspisteen longitudi.

Koordinaatit matkan funktiona

Kaavasta (25) voidaan ratkaista latitudi, ja koska $\psi = s$, saadaan kaavasta (26) longitudi eli molemmat koordinaatit matkan (päiväntasaajasta) funktiona ovat

$$\varphi = \overline{\text{arc}}\sin(\sin\phi_v \sin s) \tag{27}$$

$$\lambda = \arctan(\cos\phi_v \tan s) + \Lambda_i \tag{28}$$

Integraalikäyrä

Koska kaavan (22) perusteella on $\tan\psi = \sin\varphi \Big/ \left(\pm\sqrt{\cos^2\varphi - \cos^2\phi_v}\right)$, saadaan kaavasta (26)

$$\lambda - \arctan\left[\frac{\cos\phi_v \sin\varphi}{\pm\sqrt{\cos^2\varphi - \cos^2\phi_v}}\right] = \Lambda_i \tag{29}$$

Tästä kaavasta seuraa pisteen $P(\varphi,\lambda)$ kautta kulkevalle isoympyrälle sekä latitudin

$$\varphi = \overline{\text{arc}}\sin\left[\frac{\sin\phi_v \tan(\lambda-\Lambda_i)}{\pm\sqrt{\cos^2\phi_v + \tan^2(\lambda-\Lambda_i)}}\right] \tag{30}$$

että longitudin suhteen ratkaistut lausekkeet

$$\lambda = \Lambda_i + \arctan\left[\frac{\cos\phi_v \sin\varphi}{\pm\sqrt{\cos^2\varphi - \cos^2\phi_v}}\right] \tag{31}$$

Kulkusuunta aluksen matkan ja nopeuden funktiona

Kaavasta (27) seuraa $\sin\varphi = \sin\phi_v \sin s$ ja derivoidaan tämä muuttujan s suhteen. Käytetään apuna kaavaa (1), jolloin saadaan $\cos\alpha = \cos s \sin\phi_v / \cos\varphi$. Kirjoitetaan kaava (18) muotoon $\sin\alpha = \cos\phi_v / \cos\varphi$. Jaetaan se edellisellä $\cos\alpha$ -kaavalla. Tuloksesta voidaan ratkaista aluksen suunta matkan funktiona

$$\alpha(s) = \arctan(\cot\phi_v \sec s) \tag{32}$$

Kun tämä derivoidaan ajan suhteen, saadaan aluksen kääntymisnopeudelle lauseke

$$\frac{d\alpha(s)}{dt} = \frac{d\alpha(s)}{ds}\frac{ds}{dt} = \frac{\cot\phi_v \sec s \tan s}{1 + (\cot\phi_v \sec s)^2}\frac{ds}{dt} \tag{33}$$

Jos aluksen nopeus V on annettu solmuina, saadaan derivaatta kaavasta

$$\frac{ds}{dt} = \pm\frac{V/60'}{3600}\frac{\pi}{180°}\,\mathrm{rad\,s^{-1}} \tag{34}$$

Nykyisenä elektroniikan aikakautena tuskin on vaikea rakentaa näihin kaavoihin perustuvaa isoympyräpurjehduksen ohjausta.

Vertex-piste

Jos isoympyrä kulkee pisteiden $P(\varphi_1, \lambda_1)$ ja $P(\varphi_2, \lambda_2)$ kautta, voidaan vertex-latitudi ϕ_v ratkaista numeerisesti suoraan kaavasta

$$\lambda_1 - \arctan\left[\frac{\cos\phi_v \sin\varphi_1}{\pm\sqrt{\cos^2\varphi_1 - \cos^2\phi_v}}\right] = \lambda_2 - \arctan\left[\frac{\cos\phi_v \sin\varphi_2}{\pm\sqrt{\cos^2\varphi_2 - \cos^2\phi_v}}\right] = \Lambda_i \tag{35}$$

Käytännössä latitudin ϕ_v numeerinen ratkaisu tästä kaavasta voi osoittautua hankalaksi. Kaavoissa (25)-(33) tarvittava vertex-latitudi ϕ_v voidaan haluttaessa ratkaista myös alla annettuja perinteisiä pallotrigonometrian kaavoja käyttäen.

Pallotrigonometrian kaavoja

Pisteiden $P(\varphi_1, \lambda_1)$ ja $P(\varphi_2, \lambda_2)$ väliseen isoympyrään liittyviä pallotrigonometriasta tunnettuja kaavoja kannattaa tarvittaessa käyttää seuraavasti:

Isoympyrämatka d ([3] s. 34)

$$d = \overline{\text{arc}}\cos\left[\sin\varphi_1 \sin\varphi_2 + \cos\varphi_1 \cos\varphi_2 \cos(\lambda_2 - \lambda_1)\right]$$

Lähtösuunta ([3] s. 52)

$$\alpha_1 = \arctan\left\{\frac{\sin(\lambda_2 - \lambda_1)}{\cos\varphi_1 \tan\varphi_2 - \sin\varphi_1 \cos(\lambda_2 - \lambda_1)}\right\}$$

Tulosuunta ([3] s. 52)

$$\alpha_2 = \arctan\left\{\frac{\sin(\lambda_2 - \lambda_1)}{-\left[\cos\varphi_2 \tan\varphi_1 - \sin\varphi_2 \cos(\lambda_2 - \lambda_1)\right]}\right\}$$

Nämä määrittävät myös kaavassa (32) suuntakulman arctan-funktion haaran.

Vertex-pisteen latitudi ([3] s. 72)

$$\phi_v = \pm\arccos(\sin\alpha_1 \cos\varphi_1) = \pm\arccos(\sin\alpha_2 \cos\varphi_2)$$

Vertex-pisteen longitudierot ([3] s. 72)

$$\tan(\lambda_v - \lambda_1) = \frac{1}{\tan\alpha_1 \sin\varphi_1} \qquad \text{ja} \qquad \tan(\lambda_v - \lambda_2) = \frac{1}{\tan\alpha_2 \sin\varphi_2}$$

ESIMERKKI

Atlantin ylitys Irlannista Pohjois-Amerikkaan

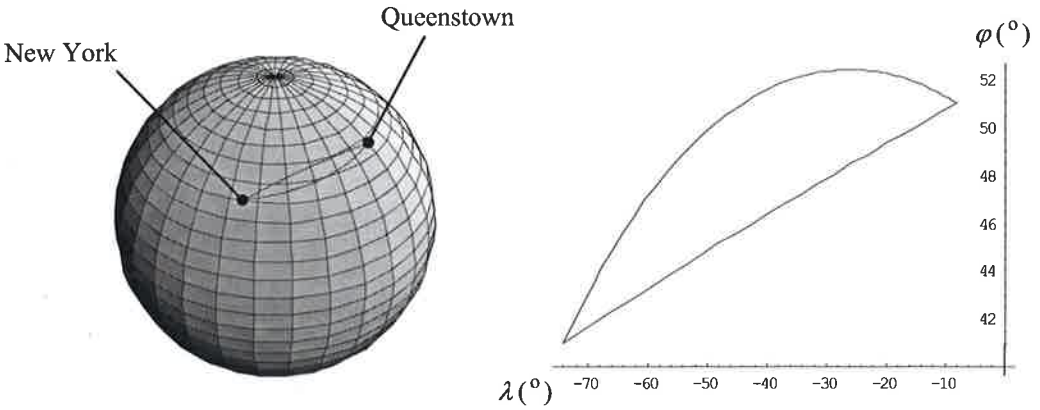
Lähtö Queenstownista ($\varphi_1 = +51°$, $\lambda_1 = -8°$), tulo New Yorkiin ($\varphi_2 = +41°$, $\lambda_2 = -74°$). Tarkalle loksodromille, loksodromille keskilatitudimenetelmällä sekä isoympyrälle saadaan alla olevat tulokset. Etäisyydet on annettu meripeninkulmina (M).

	Tarkka loksodromipurjehdus	Keskilatitudi-loksodromi
Etäisyys	2804,76 M	2815,52M
Suunta	257,65°	257,70°

Isoympyräpurjehdus

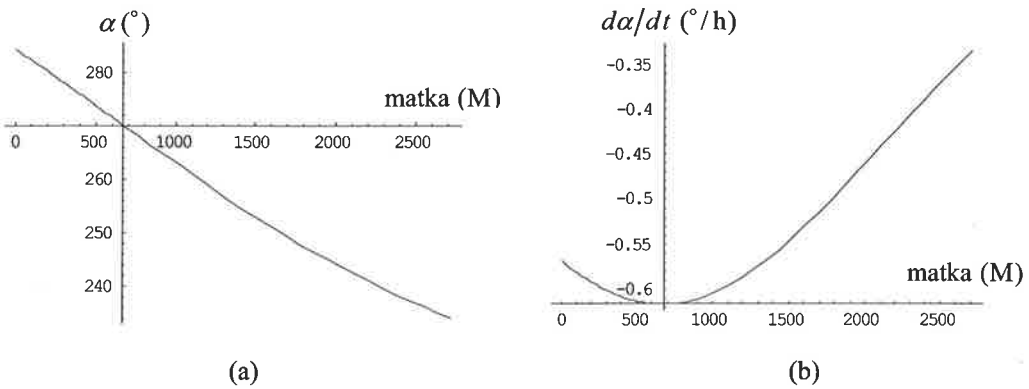
Lähtösuunta	$\alpha_1 = 284{,}19°$
Tulosuunta	$\alpha_2 = 233{,}94°$
Vertex	$\phi_v = +52°24{,}1'$, $\lambda_v = -26°01{,}6'$ (lähtö- ja tulopisteiden välissä)
Etäisyys	$d = (s_{v1} - s_1) + (s_{v2} - s_2) = 673{,}59\text{M} + 2046{,}15\text{M} = 2719{,}74\text{M}$ (kaava (25))

Kuvassa 2 on esitetty isoympyräpurjehdukseen ja tarkkaan loksodromipurjehdukseen liittyvät reitit maapallon pinnalla.



Kuva 2. Purjehdus Irlannista Pohjois-Amerikkaan. (a) Loksodromin ja isoympyrän vertailu. Isoympyrä kulkee pohjoisempana, (b) Latitudi φ longitudin λ funktiona isoympyrällä ja loksodromilla. Mercatorin projektiossa loksodromi on suora.

Isoympyrän ja tarkan loksodromin latitudiero vertexissä on 237,12 M. Maksimaalinen latitudiero 314,27M saavutetaan longitudilla $\lambda = -43°21,0'$. Kuvassa 3 on esitetty suuntakulma α ja suuntakulman muutosnopeus $d\alpha/dt$ kuljetun matkan funktiona kuvan 2 isoympyräpurjehduksessa.



Kuva 3. Suuntakulma (a) ja suuntakulman muutosnopeus (b) kuljetun matkan funktiona kuvan 2 isoympyrällä. Aluksen nopeus on 20,5 solmua. Origo on vertexissä.

TULOSTEN TARKASTELUA JA YHTEENVETO

Koska maapallon litistyneisyys on vain suuruusluokkaa 0,335%, ovat edellä esitetyt pallolle johdetut navigointikaavat monissa tapauksissa riittävän tarkat. Usein selitetään onnettomuuden vesillä johtuneen "navigointivirheestä", mikä tavallisesti tarkoittaa, että karttaa on luettu väärin tai loksodromin suuntakulma on asetettu väärin.

Esimerkkinä laskettu Atlantin ylitys Irlannista Pohjois-Amerikkaan noudattaa suurin piirtein Titanicin muinaista kohtalokasta matkaa. Koska lähtöpisteen ja tulopisteen latitudiero on vain $10°$, ero keskilatitudiloksodromin ja tarkan loksodromin kaavoilla laskien on pieni sekä suuntakulmassa että matkassa. Isompyräetäisyys on 85,0M lyhyempi kuin loksodromietäisyys, mikä oletetulla vauhdilla 20,5 solmua merkitsee ajassa noin neljää tuntia. Ylitysennätykseen pyrittäessä tällä pienellä aikaerolla oli tietenkin merkitystä. Kohtalokkainta oli se, että isoympyrä loksodromiin verrattuna kulki sen verran pohjoisempana, maksimissaan 314,3M longitudilla $\lambda = -43°21,0'$, että alus joutui rannikkoa lähestyessään ajelehtivien jäävuorien alueelle tunnetuin seurauksin. Uppoamispaikka [19] oli suunnilleen $\varphi = +42°$, $\lambda = -50°$ eli jo lähes tulolatitudilla. Tässä vaiheessa Titanic oli poikennut isoympyrältä selvästi etelään lukuisten jäävuorivaroitusten takia. Jos Titanic olisi käyttänyt suuntakulman jatkuvaa säätöä, havaitaan kuvista 3(a) ja (b), että suuntakulman muutosnopeus olisi suurimmillaan vertexissä ja muutenkin matkan kuluessa suuruusluokkaa $-0,62°\,\mathrm{h}^{-1} \cdots -0,30°\,\mathrm{h}^{-1}$. Vertexin osuessa matkalle oli isoympyrä viisainta jakaa itäiseen ja läntiseen haaraan ja laskea ne erikseen etumerkkeihin erityistä huomiota kiinnittäen. Päiväntasaajasta laskettu etäisyys s osoittautui käyttökelpoiseksi parametriksi myös kuvia piirrettäessä. Varsinkin isoympyrän kohdalla uudet tulokset, kaavat (27), (28) ja (32), toimivat hyvin kunhan arcus-funktioiden haara on ensin saatu selville.

Tässä työssä johdetut jatkuvan ohjauksen kaavat soveltuvat erityisen hyvin elektronisesti ohjattuun avomeripurjehdukseen. Aluksen mitatusta nopeudesta voidaan integroimalla laskea aluksen kulkema matka, jonka perusteella voidaan edelleen laskea tässä työssä johdettuja analyyttisiä (suljetun muodon) kaavoja käyttäen aluksen kulkusuunta ja kulkusuunnan muutosnopeus. Nykyaikaisen säätötekniikan avulla tämä informaatio voidaan suoraan muuttaa aluksen peräsimen liikkeitä ohjaaviksi käskyiksi.

KIRJALLISUUTTA

[1] Veneilijän merenkulkuoppi II, *Rannikkonavigointi* 7. painos, Suomen Navigaatioliiton julkaisusarja, Yliopistopaino, Helsinki 2004

[2] Hannu Lappalainen, *Merenkulku 2, Avomeripurjehdus*, Ammattikasvatushallitus, Valtion painatuskeskus, 1990

[3] Hannu Lappalainen, *Merenkulun pallotrigonometria*, Ammattikasvatushallitus, Valtion painatuskeskus, 1982

[4] Hannu Lappalainen, *Laskimen käytöstä navigoinnissa*, Kotkan Merenkulkuopiston oppilaskunta ry, Teho Testi, Kotka 1985

[5] I. S. Gradshteyn & I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Fifth Edition, Academic Press, Inc. 1980

[6] R. Williams and J. E. Phythian, *Navigation Along Geodesic Path of a Spheroid*, The Journal of Navigation 1989, Vol. 42 NO.1, s 129-136

[7] R. Bourbon, *Geodesic Line on the Surface of a Spheroid*, The Journal of Navigation 1990, Vol. 43 NO.1, pp 132-135

[8] Ivica Tijardovic, *Trans-oceanic Passages by Rhumbline Sailing*, The Journal of Navigation 1990, Vol. 43 NO.2, pp 292-293

[9] Matti A Ranta, *Trans-oceanic Passages by Rhumbline Sailing*, The Journal of Navigation 1991, Vol. 44 NO.1, pp 129-130

[10] Han-Wei Lu, Hsin-Hsiung and Chung-Hsing Chiang, *Trans-oceanic Passages by Rhumbline Sailing*, The Journal of Navigation 1991, Vol. 44 NO.3, pp 423-428

[11] Kitt C. Carlton-Wippern, *On Loxodromic Navigation*, The Journal of Navigation 1992, Vol. 45 NO.2, pp 292-297

[12] Tim Zukas, *Shortest Spheroidal Distance*, The Journal of Navigation 1994, Vol. 47 NO.1, pp 107-108

[13] Roger Prince and Roy Williams, *Sailing in Ever Decreasing Circles*, The Journal of Navigation 1995, Vol. 48 NO.2, pp 307-313

[14] G. G. Bennett, *Practical Rhumb Line Calculations on the Spheroid*, The Journal of Navigation 1996, Vol. 49 NO.1, pp 112-119

[15] Roy Williams, *The Great Ellipse on the Surface of the Spheroid*, The Journal of Navigation 1996, Vol. 49 NO.2, pp 229-234

[16] Chich-Li Chen and Tien-Pen Hsu and Jiang-Ren Chang, *A Novel Approach to Great Circle Sailings: The Great Circle Equation*, The Journal of Navigation 2004, Vol. 57 NO.2, pp 311-320

[17] Paul Hickley, *Great Cirle Versus Rhumb Line Cross-track Distance at Mid-Longitude*, The Journal of Navigation 2004, Vol. 57 NO.2, pp 320-325

[18] John Ponsonby, *Concerning Cross Track Distance at Mid-Longitude (1)*
Peter Hoare, *Concerning Cross Track Distance at Mid-Longitude (2)*
Paul Hickley, *Concerning Cross Track Distance at Mid-Longitude (3)*
The Journal of Navigation 2005, Vol. 57 NO.2, pp 149-153

[19] The Grave of the Titanic
http://www.gma.org/space1/titanic.html

AltaVista hakukoneella poimittuja hakusanalla

Loxodrome:

http://eom.springer.de/L/l060970.htm (ei tavallisesti aukea)

http://www.mat.uc.pt/~jfqueiro/Britannica_errors.html

Loksodromin esittäjä löytynyt: Pedro Nunes, *Tratado da Sphera*, Lisbon, 1537
Great Circle Route:

http://www.islandwaterworld.com/greatcircle.htm

http://www.geocities.com/CapeCanaveral/2265/gcsail.htm

Näissä artikkeleissa on kaava $\tan\varphi = \sin(\lambda - \Lambda_i)\cdot\tan\phi_v$ (miten saatu?)
Siinä longitudit ovat välillä 0-360

http://www.optimumroute.com/tutorial/tutorial_1.php

Tuulioptimoitu reitti

http://www.iho.shom.fr/publicat/free/files/S-51_Ed4-EN.pdf

Mielenkiintoinen julkaisu maapallosta

Google hakukoneella poimittuja hakusanalla

Rhumb line:

http://jacq.istos.com.au/sundry/navrhumb.html

http://www.mathpages.com/home/kmath502/kmath502.htm

http://en.wikipedia.org/wiki/Rhumb_line

http://www.geocities.com/capecanaveral/2265/msail.htm

http://williams.best.vwh.net/ellipsoid/node3.html

http://www.ams.org/featurecolumn/archive/navigation5.html

Titanic:

http://www.gma.org/space1/titanic.html

THE FORCE CAPACITY AND FATIGUE OF MUSCLE

J. FREUND
Department of Mechanical Engineering
Helsinki University of Technology
P.O.Box 4300, FIN-02015 HUT Espoo
e-mail:jouni.freund@tkk.fi

ABSTRACT

A phenomenological model for the evolution of the maximal muscle force capacity is described. The starting point is the assumption that fatigue is best modeled by a differential equation. The model reproduces the experimental results for muscle recovery and fatigue. However, the onset and development of muscle fatigue can also be predicted without simplifying assumptions concerning the loading, which is important in practical applications.

1. INTRODUCTION

Everyday experience indicates that the time a muscle is able to maintain a fixed force level, exceeding about 20% of the maximal muscle force, is rather limited. After the so-called endurance time, which decreases rapidly in the required force magnitude, the force level cannot be maintained any longer and the muscle is said to suffer from fatigue. A relaxed muscle is also known to retain its original capacity finally after an exhausting exercise. Couple of minutes may be enough for a short-term near-maximal loading but, after total exhaustion with a prolonged lower level dynamic loading, the recovery may take even hours or days (Byström, 1991). Several explanations for the mechanism of fatigue have been suggested in the literature (see e.g. (Byström, 1991)). The aerobic and anaerobic energy and force production mechanisms clearly affect the setting. Blood flow characteristics such as occlusion, the lack of motivation (Asmussen, 1979; Kent-Braun, 1999), central and peripheral factors (Kent-Braun, 1999), having to do with the central nervous system and muscle itself in the same order, may also be important.

Unfortunately, there seem not to be any simple explanation based e.g. on the accumulation of the fatigue substance or inhibition by the central nervous system (Chaffin, 1973; Kirkendall, 1990) and the entire chain of the complex chemical processes of the central nervous system, muscle, etc., may have to be taken into account for the complete picture. A rough idea of the effect of the various parameters and loading conditions can be obtained from experimental data in the literature. The relationship between fatigue and blood circulation has been discussed e.g., by Barnes (1980), Morton (1987), Stephens and Taylor (1972), and Wilson (1996). The effect of age, etc., has been discussed e.g., by Deeb et al. (1992) and Rich (1960). The effects of temperature on the fatigue and recovery curves are discussed by Clarke and Stelmach (1966). Recovery times have been studied, e.g. by Byström (1991), Caldwell (1970), and Clarke (1962). Fatigue during intermittent loading conditions has been studied by Björksten and Jonsson (1977), Byström (1991), Clarke (1962), and Rich (1960). Fatigue data and models have been presented e.g., by Clarke (1962), Deeb et al. (1992), Hawkins and Hull (1993), Kothiyal and Ibramsha (1986), Morton (1987), and Wood et al. (1997).

In this study, we propose a relatively simple model for muscle fatigue and recovery and show that the model predicts the observed qualitative behavior of muscle correctly. A refined model is needed, as the existing models of literature cannot be used with general loading profiles of the muscle. Also, the picture given by them about recovery is not satisfactory. This is somewhat problematic, as maximal muscle force, fatigue, and recovery are important elements e.g. in biomechanical modeling of musculoskeletal system (Dul et al., 1984; Epstein and Herzog, 1998; Freund and Takala, 2001; Prilutsky et al., 1998), in various optimization tasks concerning man-machine-work system, and in ergonomic studies (Wood et al., 1997), as a reduction of muscle fatigue may reduce the perceived fatigue or discomfort (Dul et al., 1994) and the criterion for the acceptable load level may partly be based on a fatigue criterion. Both the simplified version used already by Freund and Takala (2001) and a version aiming at quantitatively correct results are presented. The main assumption is that evolution of the force capacity of a muscle is best modeled by a differential equation, as most physical phenomena.

2. FATIGUE MODEL

The main idea is as follows: a quantity $S_{\max}$, called as the muscle force capacity and representing the maximal available muscle force, is associated to each muscle or muscle group. This quantity decreases when the muscle is used and it increases in rest. As muscle force S cannot exceed capacity $S_{\max}$, condition $S = S_{\max}$ means that the muscle suffers from fatigue. The starting point is the assumption that the rate of change $dS_{\max}/dt$ of the muscle force capacity $S_{\max}$ depends somehow on the actual muscle force S, the blood flow rate through the muscle Q and $S_{\max}$ itself. This general idea is not new, but it has not been used for modeling purposes.

If for simplicity, the blood flow rate Q and the muscle force S are assumed to be related algebraically, the mathematical representation becomes

$$\frac{d\bar{s}}{dt} = F(\bar{s}, s, q), \tag{1}$$

$$G(s, q) = 0, \tag{2}$$

in which the forms F and G are to be found experimentally. Above and in what follows, a dimensionless approach is used with dimensionless muscle force, muscle force capacity and blood flow rate defined by $s = S/\bar{S}_{\max}$, $\bar{s} = S_{\max}/\bar{S}_{\max}$ and $q = Q/\bar{Q}$. The reference values $\bar{S}_{\max}$ and $\bar{Q}$ denote the muscle force capacity after a long rest, and the blood flow rate when $S = 0$, respectively.

When equation (2) is used to eliminate q from (1), the initial value problem to be studied can be written as (with some other F)

$$\frac{d\bar{s}}{dt} = F(\bar{s}, s) \quad t > t_0, \tag{3}$$

$$\bar{s} = \bar{s}_0 \quad t = t_0, \tag{4}$$

$$0 \le s \le \min\{1, \bar{s}\} \quad t \ge t_0. \tag{5}$$

The inequality means that the actual muscle force must always be non-negative and smaller than the available force capacity. It is noteworthy that the capacity $\bar{s}$ is allowed to exceed the value

one (to keep the model flexible enough), whereas the dimensionless muscle force s always satisfies the condition $s \le 1$. Knowing $F(\bar{s},s)$ and s as function of time, one may answer questions concerning the muscle force capacity and fatigue phenomenon, defined by condition $s = \bar{s}$, by solving the ordinary differential equation (3) containing $\bar{s}$ and s only.

To make things more concrete, one may consider the muscle as a reservoir for the force capacity. The reservoir is filled by the 'surrounding' systems at a certain rate and, when the muscle is activated, the capacity decreases at some other rate. What actually happens, depends on the relative strengths of these two processes in $F(\bar{s},s)$ of equation (3).

3. RATE EXPRESSION

The remaining task in the modeling is to find the expression for $F(\bar{s},s)$ on a region of possible values of $\bar{s}$ and s. This is a matter of experimentation, although some general principles and prior knowledge about fatigue can be used to restrict the set of possible expressions. The fact that $\bar{s}$ is not an observable 'as is' except at the endurance limit, at which the force capacity coincides with the actual force, has also to be accounted for. Let us note that a good approximation combines properly accuracy and simplicity, and therefore we aim at an expression predicting only the main features observed.

The linear approximation

$$F(\bar{s},s) = \alpha(\bar{s}_{\max} - \bar{s}) - \beta s \tag{6}$$

used already in Freund and Takala (2001) is the simplest possible choice (Figure 1). Above α, β and $\bar{s}_{\max}$ are parameters to be chosen so that the predicted and measured results match. It is plausible that the reduction rate of the force capacity is approximately proportional to the actual force but linearity assumption with respect to $\bar{s}$ is rather strong.

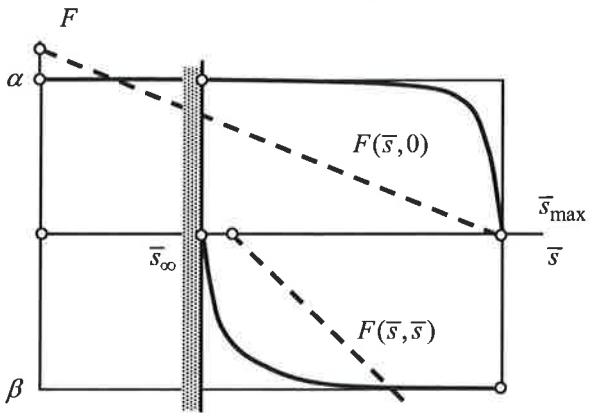


Figure 1: Schematic illustration of $F(\bar{s},\bar{s})$ and $F(\bar{s},0)$. The linear approximation is illustrated by a broken line. The solid line shows the refined approximation. The values $0 \le \bar{s} < \bar{s}_{\infty}$ are unattainable under normal blood flow conditions. The actual muscle force is restricted to $0 \le s \le \min\{\bar{s},1\}$.

For a refined expression, it is assumed that experimental data about $F(\bar{s},\bar{s})$ and $F(\bar{s},0)$ describing the force capacity reduction during fatigue ($s = \bar{s}$) and recovery ($s = 0$), are available. Then, the overall expression can be approximated by linear interpolation (note that $s/\bar{s} \le 1$ and hence $s/\bar{s}$ is bounded)

$$F(\bar{s},s) \approx F(\bar{s},0)\frac{\bar{s}-s}{\bar{s}} + F(\bar{s},\bar{s})\frac{s}{\bar{s}} \tag{7}$$

giving the correct behavior at least during fatigue and recovery. As some approximation is already introduced in (7), the use of simplified parameterized forms

$$F(\bar{s},0) = \alpha\frac{\tanh m(\bar{s}_{\max}-\bar{s})}{\tanh m\bar{s}_{\max}}, \tag{8}$$

$$F(\bar{s},\bar{s}) = \beta\frac{\tanh n(\bar{s}_{\infty}-\bar{s})}{\tanh n(\bar{s}_{\max}-\bar{s}_{\infty})}, \tag{9}$$

with positive parameters α, β, $\bar{s}_{\max}$, $\bar{s}_{\infty}$, m and n, may also be justified (Figure 1). The drawback of the refined expression is that a numerical method is needed in predicting the evolution of the force capacity even with a simple loading.

4. RECOVERY OF FORCE CAPACITY

In what follows, we compare the predictions by the model to the experimental data from various references (Figure 2). The analytic solution to (3) with $s=0$ and the approximate expression (8) is given by

$$\alpha(t-t_0) = -\frac{\tanh m\bar{s}_{\max}}{m}\log\frac{\sinh m(\bar{s}_{\max}-\bar{s})}{\sinh m(\bar{s}_{\max}-\bar{s}_0)}. \tag{10}$$

Above $\bar{s}_0$ is the value of the force capacity at $t=t_0$. The exponential expression for $\bar{s}$, suggested in some references, is obtained at the limit $m\to 0$.

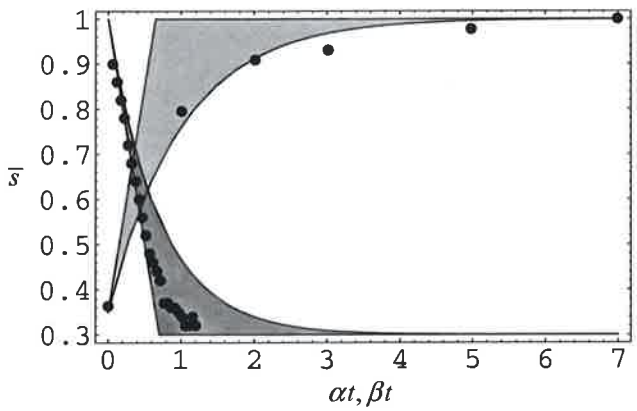


Figure 2: Recovery (light gray) and reduction (dark gray) of the force capacity $\bar{s}$ as a function of the dimensionless times αt and βt and smoothness parameters m and n ($\bar{s}_{\infty}=0.2$, $\bar{s}_{\max}=1$). The shaded areas illustrate the curves obtained by the different selections of the smoothness parameters $m,n\in]0,\infty[$. The value 0 corresponds to the exponential behavior and the value ∞ the piecewise linear behavior. The experimental data of Clarke (1962) (with appropriate scaling) are shown as dots.

Experimentation in recovery of the muscle force capacity is difficult, as the condition $\bar{s} = s$ has to be used, which inevitably affects the results. In fact, the experimental results of the literature concern intermittent loading to be discussed later. Strength recovery of the finger flexor muscles has been studied experimentally in Clarke (1962), Clarke and Stelmach (1966), Stephens and Taylor (1972) after static and dynamic loading. The quite fast initial recovery of these results implies that $\alpha \approx 1$ and m is small. Since the loading is actually intermittent, these values are naturally just rough estimates. Negligible recovery under external occlusion of the blood circulation has been verified experimentally e.g. in Stephens and Taylor (1972). This implies that parameter α reduces rapidly under external occlusion of the blood circulation.

Recovery time in the force magnitude is a useful concept, say, in design of experiments concerning the maximal muscle force, as repeated measurements should not be affected by fatigue too much. The expressions can be obtained from equation (10) by using condition $\bar{s} = 1$. If $\bar{s}_{\max} = 1$, the total recovery in the force magnitude takes infinitely long time with the present $F(\bar{s}, 0)$ vanishing at $\bar{s} = \bar{s}_{\max}$. However, any fixed level below that (99 %, say) is attained in finite time. It is noteworthy, that the recovery times in force magnitude and endurance time are quite different concepts as the latter means the rest time needed the endurance time to take its original value. According to the experimental data, the recovery in the endurance time for a totally fatigued muscle can take 60 min or even several days after a very exhaustive dynamic exercise (Byström, 1991), whereas recovery in force magnitude is much faster.

5. REDUCTION OF FORCE CAPACITY

The force capacity reduction in fatigue is described by $F(\bar{s}, \bar{s})$. The analytic solution to (3) with expression (9) is given by (Figure 2)

$$t - t_0 = -\frac{\tanh n(\bar{s}_{\max} - \bar{s}_\infty)}{n\beta} \log \frac{\sinh n(\bar{s}_\infty - \bar{s})}{\sinh n(\bar{s}_\infty - \bar{s}_0)}, \tag{11}$$

in which $\bar{s}_0$ is the initial value of $\bar{s}$ at t_0. Expression (9) is chosen to vanish at $\bar{s} = \bar{s}_\infty$, as the experiments show that the recovery and reduction processes are in balance when $s = \bar{s} = \bar{s}_\infty \approx 0.10...0.30$, and hence a constant force $s < \bar{s}_\infty$ can be held for a very long time.

Experimental results on the reduction of force capacity have been described by Clarke (1962), Heyward (1975), and Stephens and Taylor (1972). According to Stephens and Taylor (1972), the reduction of the force capacity in time is nearby exponential, leveling off to the limit value $\bar{s}_\infty$. This observed behavior results with a small n, $\beta \approx 0.5\,\mathrm{min}^{-1}$ and $\bar{s}_\infty \approx 0.20$. The same experiment with blood flow occlusion gave an almost piecewise linear behavior with a very small limit value $\bar{s}_\infty$ meaning a large n, $\beta \approx 0.5\,\mathrm{min}^{-1}$ and $\bar{s}_\infty \approx 0.05$. In conclusion, both n and $\bar{s}_\infty$ are affected by the external flow conditions, but β is not.

According to the present study, evolution of the maximal force capacity during fatigue does not depend on the force history. This is confirmed by the result of Stephens and Taylor (1972), as fatigue from the initial submaximal force level gave the same β and $\bar{s}_\infty$. Also Clarke (1962) observed an almost exponential behavior with the limit value $\bar{s}_\infty < 0.3$ and $\beta \approx 0.5...0.7\,\mathrm{min}^{-1}$.

6. ENDURANCE TIME

Endurance time as a function of the loading level is often used to quantify the fatigue phenomenon. This integral quantity contains information about the behavior of $F(\bar{s}, s)$ between fatigue with $s = \bar{s}$ and recovery with $s = 0$. In principle, the quantity can be obtained by solving the initial value problem and finding the time instant at which $\bar{s} = s$.

We discuss next only the case of a constant s, although the definition holds true also with arbitrary time-dependent loading. The integral solution for a constant s can be written as

$$t - t_0 = -\int_{\bar{s}}^{\bar{s}_0} \frac{1}{F(\bar{s},s)} d\bar{s} \tag{12}$$

in which $\bar{s}_0$ is the initial force capacity at $t = t_0$ and t is the endurance time. In the case of the linear rate expression (6), equation (12) gives

$$t - t_0 = -\frac{1}{\alpha} \log \frac{s + \varepsilon s - \bar{s}_{\max}}{\bar{s}_0 + \varepsilon s - \bar{s}_{\max}} \tag{13}$$

in which $\varepsilon = \beta / \alpha$ and $\bar{s}_0$ is the value of force capacity at $t = t_0$. Despite the rather rough rate expression, a close fit to the experimental results is obtained (Figure 3). With the refined expression of equations (7), (8), (9) and calculation based on (12), a close fit was obtained with numerous different combinations of the parameters, of which two are shown in Figure 3. As an integral quantity, the endurance time is not sensitive to the details of $F(\bar{s},s)$ and therefore the parameters of $F(\bar{s},s)$ cannot be determined reliably from the endurance data. Insensitivity explains the co-existence of the various exponential and rational expressions for endurance time suggested in the references and the remarkably coherent experimental results of Figure 3.

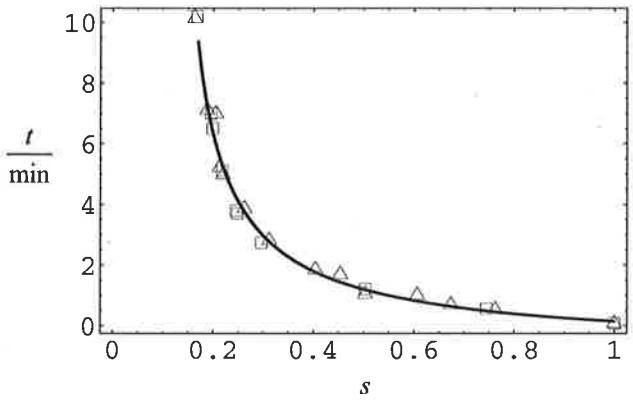


Figure 3: Experimental results for endurance time t as function of the relative magnitude of the muscle force taken from various resources according to Morton (1987). With the linear model, a close fit is obtained with $\alpha = 0.17$, $\beta = 1.26$ and $\bar{s}_{\max} = 1.18$. With non-linear model, a close fit is obtained with numerous combinations of which $\alpha \approx 0.1\,\mathrm{min}^{-1}$, $\beta \approx 0.6\,\mathrm{min}^{-1}$, $\bar{s}_\infty \approx 0.2$, $n \approx 5$, $m \approx 0$, $\bar{s}_{\max} \approx 1$ and $\alpha \approx 0.4\,\mathrm{min}^{-1}$, $\beta \approx 0.9\,\mathrm{min}^{-1}$, $\bar{s}_\infty \approx 0.05$, $n \approx 10$, $m \approx 0$, $\bar{s}_{\max} \approx 1.2$ are just two examples.

The limit force $\bar{s}_\infty$, at which the endurance time goes to infinity, is given by (argument of the logarithm in (13) goes to zero)

$$\bar{s}_\infty = \frac{\bar{s}_{\max}}{1 + \varepsilon}. \tag{14}$$

The value $\bar{s}_\infty \approx 0.15...0.20$ (giving $\varepsilon \approx 4$) has been suggested by Byström (1991), Hagberg (1981), and Rohmert (1960). However, these data do not extend beyond 15 min endurance times. A lower value $\bar{s}_\infty \approx 0.08$ was found by Björksten and Jonsson (1977) for the flexor muscles of the forearm with an experiment lasting 60 min.

7. INTERMITTENT LOADING

Experimentation on force capacity recovery and effects of repetitive work (Björksten and Jonsson, 1977; Byström and Kilbom, 1990; Rich, 1960) are based on loading with period $\Delta t = \Delta t_w + \Delta t_r$, in which Δt_w is the length of the work phase with assumedly constant s (before fatigue) and Δt_r is the length of the rest phase with $s = 0$. Intermittent loading brings the additional parameters Δt and the work ratio $k = \Delta t_w / \Delta t$ to play, which complicates the experimentation. The pure recovery and reduction of the force capacity discussed earlier correspond to the selections $k = 0$ and $k = 1$, respectively.

In what follows, we use double overbars to indicate intermittent loading. The linear approximation (6) is convenient in assessing the roles of the work ratio and the length of the cycle qualitatively, as closed form expressions for the force capacity are available during the rest and work periods. The endurance time for fatiguing work periods, so that $\bar{\bar{s}} = s$, can be shown to be

$$t - t_0 = -\frac{1}{\alpha} \log[\frac{(1+k\varepsilon)s - (1+\varepsilon)\bar{s}_\infty}{\varepsilon k s}] \tag{15}$$

in which $\bar{s}_\infty$ is defined in equation (14). With $k = 1$ the expression reduces into equation (13) of the static case. The limit force for the intermittent loading is (the argument of log function of (15) goes to zero)

$$\bar{\bar{s}}_\infty = \bar{s}_\infty \frac{1+\varepsilon}{1+\varepsilon k} = \bar{s}_{\max} \frac{1}{1+\varepsilon k} . \tag{16}$$

The latter form of (16) is a consequence of the relationship $\bar{s}_\infty = \bar{s}_{\max} / (1+\varepsilon)$ of equation (14). The limit force exceeds that of the static case, as the multiplier of $\bar{s}_\infty$ in (16) is larger than one if $k < 1$.

These results predict well the finding (see Byström and Kilbom (1990) and references therein) that in the intermittent loading case the decisive factors of fatigue are k and s, as the length of the period Δt does not appear in (15) at all. The reduction of $\bar{\bar{s}}_\infty$ in k displayed by (16) has been verified experimentally by Caldwell (1970) and Hagberg (1981). In the force recovery experiment conducted by Clarke (1962), the force stabilized to 0.9 level. This value follows from (16) with the parameters of the linear model (Figure 3), assuming that $\Delta t_w = 2...3$ s. In general, the linear model with representative values of the parameters seems to overestimate $\bar{\bar{s}}_\infty$ somewhat.

Assuming that the mean loading s/k is constant, equation (15) shows that the endurance time increases in k. The results of Björksten and Jonsson (1977) for $k \in \{9,...,1\}/10$ $\Delta t = 1/6$ min contradict this prediction, as the static endurance curve is approached when $k = 9/10$ and also when $k = 1/10$. Hence, assuming continuity, the endurance time should take an extremum value for some value of the work ratio. To add to the confusion, Björksten and Jonsson (1977) found that the endurance time for a very short resting period approaches that of continuous work, whereas Mathiassen and Winkel (1992) reported the longest mean endurance time of all cases studied with a very short resting time of 1.7 s. Clearly, more experimentation on the effect of k would be needed, as knowing the value of k minimizing the fatigue has significance in the planning of repetitive work (Janero, 1985; Wood et al., 1997).

8. DISCUSSION

Considerable experimental effort has been directed to understanding the effects of various factors on muscle fatigue. Arriving at a coherent picture is, however, complicated by the different experimental settings and definitions, and sometimes contradicting results. A mathematical model for muscle force capacity based on observed behavior was introduced here. With this aid, discussion about the effects of age, etc., on fatigue boils down to discussion about the change of the parameters of the model in age etc., which is a great asset in statistical studies. Preliminary values of the parameters of the rate expression were found to be $\alpha \approx 1\,\mathrm{min}^{-1}$, $\beta \approx 0.5\,\mathrm{min}^{-1}$, $m \approx 1$, $n \approx 10$, $\bar{s}_\infty \approx 0.2$ and $\bar{s}_{\max} \approx 1.05$. Since the data had to be extracted from various references, and concerned different muscles or muscle groups, these figures can be used only for qualitative predictions.

The causes of fatigue were chosen to be the muscle force and the local blood flow rate. However, the local blood flow rate was eliminated from the rate expression $F(\bar{s}, s)$. The introduction of this variable may be advantageous from the modeling viewpoint: posture and simultaneous activation of other muscle groups affect the blood pressure and the local blood flow rates, and thereby also fatigue. The effect of the local blood flow rate has been demonstrated in various experiments with venous and arterial occlusions (Asmussen, 1979; Barnes, 1980; Morton, 1987; Stephens and Taylor, 1972). According to Barnes (1980), the local blood flow rate increases rapidly up to $\bar{s}_\infty$ and decreases thereafter to total occlusion. The total occlusion takes place at a lower force value in high strength subjects than in low strength subjects, possibly due to the differences in the organization and distribution of the vascular tree inside the muscle (Kahn et al., 1988). This may explain why high absolute strength implies reduced endurance time (Carlson, 1969) and why the endurance times of female subjects consistently seem to exceed those of male subjects (Kahn et al., 1988).

In the biomechanical model of Freund and Takala (2001), the muscle force capacities define the set of feasible solutions in which the likely solution is sought. Hence, knowing the muscle force capacity is essential in these calculations. Muscle length does not affect the dimensionless muscle force capacity $\bar{s}$ or the quantities discussed in this study but, as known, $S_{\max}$ and $\bar{S}_{\max}$ change (see Chaffin and Andersson (1991) or Epstein and Herzog (1998)). Therefore, when predicting the fatigue of a given muscle under the known S (perhaps obtained by using a biomechanical model with known posture and external forces), also the length of the muscle has to be known.

ACKNOWLEDGEMENT

This work is partly based on fatigue studies and experiments at the Finnish Institute of Occupational Health. The author would like to thank MD, DrMed Esa-Pekka Takala and other colleaques of Department of Physiology for useful discussions concerning the subject and co-operation in experimentation.

REFERENCES

Asmussen, E. (1979) *Muscle fatigue*. Medicine and Science in Sports 11, 313-321.

Barnes, W.S. (1980) *The relationship between maximum isometric strength and intramuscular circulatory occlusion*. Ergonomics 23, 351-357.

Björksten, M., Jonsson, B. (1977) *Endurance limit of force in long-term intermittent static contractions*. Scand. j. work environ. & health 3, 23-27.

Byström, S. (1991) *Physiological response and acceptability of isometric intermittent handgrip contractions*, Karolinska Institute, Stockholm, 174 pp.

Byström, S., Kilbom, Å. (1990) *Physiological response in the forearm during and after isometric intermittent handgrip*. European Journal of Applied Physiology 60, 457-466.

Caldwell, L.S. (1970) *Decrement and recovery with repetitive maximal muscular exertions.* Human Factors 12, 547-552.

Carlson, B.R. (1969) *Level of Maximum Isometric Strength and Relative Load Isometric Endurance.* Ergonomics 12, 429-435.

Chaffin, D. (1973) *Localized Muscle Fatigue - Definition and Measurement.* Journal of Occupational Medicine 15, 346-354.

Chaffin, D.B., Andersson, G.B.J. (1991) *Occupational Biomechanics.* John Wiley & Sons.

Clarke, D.H. (1962) *Strength recovery from static and dynamic muscular fatigue.* Research quarterly 33, 349-355.

Clarke, D.M., Stelmach, G.E. (1966) *Muscular fatigue and recovery curve parameters at various temperatures.* Research quarterly 37, 468-479.

Deeb, J.M., Drury, C.G., Pendergast, D.R. (1992) *An exponential model of isometric fatigue as a function of age and muscle groups.* Ergonomics 35, 899-918.

Dul, J., Douwes, M., Smitt, P. (1994) *Ergonomic guidelines for the prevention of discomfort of static postures based on endurance data.* Ergonomics 37, 807-815.

Dul, J., Johnson, G.E., Shiavi, R., Townsend, M.A. (1984) *Muscular synergism -II. A minimum-fatigue criterion for load sharing between synergistic muscles.* Journal of Biomechanics 17, 675-684.

Epstein, M., Herzog, W. (1998) *Theoretical Models of Skeletal Muscle.* John Wiley & Sons, Chichester.

Freund, J., Takala, E.-P. (2001) *A dynamic model of the forearm including fatigue.* Journal of Biomechanics 34, 597-605.

Hagberg, M. (1981) *Muscular endurance and surface electromyogram in isometric and dynamic exercise.* Journal of Applied Physiology 51, 1-7.

Hawkins, D., Hull, M.L. (1993) *Muscle force as affected by fatigue: mathematical model and experimental verification.* Journal of Biomechanics 26, 1117-1128.

Heyward, V.H. (1975) *Influence of static strength and intramuscular occlusion on submaximal static muscle endurance.* Research quarterly 46, 393-402.

Janero, R.E. (1985) *A study of the reduction of fatigue impact on productivity through optimal rest break scheduling.* Human Factors 27, 459-466.

Kahn, J.F., Huart, F., Monod, H. (1988) *Variation of the maximum maintenance time of an isometric contraction in the presence of a second contraction.* Ergonomics 31, 1287-1298.

Kent-Braun, J.A. (1999) *Central and peripheral contributions to muscle fatigue in humans during sustained maximal effort.* Eur J Appl Physiol 80, 57-63.

Kirkendall, D.T. (1990) *Mechanisms of peripheral fatigue.* Medicine and Science in Sports and Exercise 22, 444-449.

Kothiyal, K.P., Ibramsha, M. (1986) *Fatigue in isometric contraction in a single muscle fibre: a compartmental calcium ion flow model.* Journal of Biomechanics 19, 943-950.

Mathiassen, S.E., Winkel, J. (1992) *Can occupational guidelines for work-rest schedules be based on endurance time data* ? Ergonomics 35, 253-259.

Morton, R.H. (1987) *A simple model to link hemodynamics, fatigue and endurance in static work.* Journal of Biomechanics 20, 641-644.

Prilutsky, B.I., Isaka, T., Albrecht, A.M., Gregor, R.J. (1998) *Is coordination of two-joint leg muscles during load lifting consistent with the strategy of minimum fatigue.* Journal of Biomechanics 31, 1025-1034.

Rich, G.Q. (1960) *Muscular Fatigue Curves of Boys and Girls.* Research quarterly 31, 485-498.

Rohmert, W. (1960) *Ermittlung von Erholengspausen für statische Arbeit des Menschen.* Int. Z. angew Physiol. einschl. Arbeitsphysiol. 18, 123-164.

Stephens, J.A., Taylor, A. (1972) *Fatigue of maintained voluntary muscle contraction in man.* Journal of Physiology 220, 1-18.

Wilson, J.R. (1996) *Evaluation of skeletal muscle fatigue in patients with heart failure.* Journal of Molecular and Cellular Cardiology 28, 2287-2292.

Wood, D.W., Fisher, D.L., Andres, R.O. (1997) *Minimizing Fatigue during Repetitive Jobs: Optimal Work-Rest Schedules*. Human Factors 39, 83-101.

BUCKLING ANALYSIS THROUGH EQUILIBRIUM EQUATIONS OF CURVED STRUCTURES

J. PAAVOLA, E.-M. SALONEN

Helsinki University of Technology,
Laboratory of Structural Mechanics
P.O.Box 2100, FIN-02015 TKK
email: Eero-Matti.Salonen@tkk.fi

ABSTRACT

In the applications of the fundamental courses of structural mechanics, the derivation of the equilibrium equations is based on the initial geometry of the structure, in other words, on the so-called small displacement theory. Buckling analysis is probably the first contact for students where the deformed geometry of the structure must be taken in some measure into account when striving for equilibrium equations. The figures utilized in this context in the literature are often of so schematic nature that the assumptions employed in the derivations may remain obscure. It is not satisfactory if the students find this important step towards more sophisticated formulations of more general problems more or less vague. Here, a rather systematic approach where the assumptions used are clearly expressed will be employed. Two main points will be emphasized: the Lagrangian description — as is conventional in structural mechanics — and the use of the equilibrium equations of initially curved rods.

1. INTRODUCTION

It is a helpful practice to draw pictures so that the displacement of the deformed geometry is temporarily strongly exaggerated. Quite often in the literature, a similar picture is drawn not showing the real displacement but just, for instance a suitable vertical or horizontal measure. The analysis then continues by using this measure as a tool. Anyone who is not very familiar with this practice, may not be quite sure what this means. It looks more or less like an Eulerian description. One may have this far only experience of the small displacement theory where the equilibrium equations have been derived with no consideration of the displacement field and no real reason to discuss the differences between the Eulerian and Lagrangian descriptions has emerged. The buckling case is thus a suitable place to speak of the Lagrangian description that is normally applied in solid mechanics and of the difference between the material and spatial coordinates.

2. KINEMATICS

Let us consider an initially straight slender rod in its initial, and due to buckling deformed configurations (Figures 1 (a) and (b)).

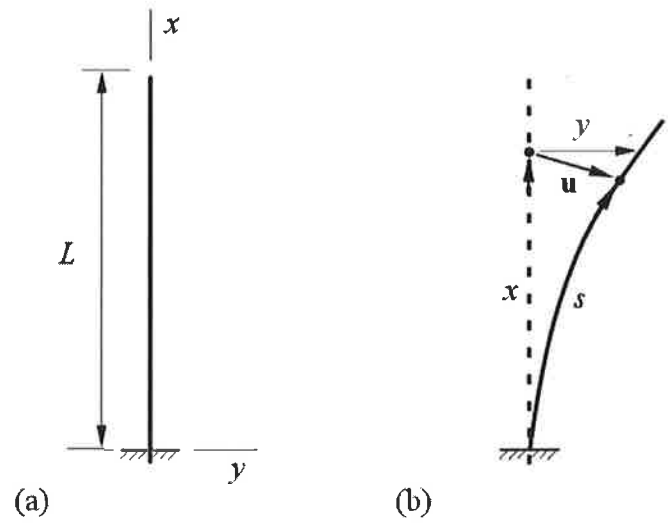


Figure 1 (a) Initial configuration. **(b)** Exaggerated deformed configuration.

The value of the axial coordinate x fixes a material point or more generally a cross-section in the initial state. A great simplification is achieved by assuming as usual that the displacement to the buckled state happens so that the rod axis remains inextensible. Thus if the arclength coordinate along the curved deformed axis is denoted by s, the value $s = x$ fixes the cross-section initially at x. The rod has been depicted with the boundary conditions according to the cantilever case but this is not essential for the general formulae to follow. The possible loading of the rod has not been shown here to keep the figures simple. We consider buckling in a plane so that the rod axis remains in the xy-plane. The displacement of a generic point on the beam axis initially at x is denoted by $\mathbf{u}$ and its components by u and v so that in more detail

$$\mathbf{u} = u\mathbf{i} + v\mathbf{j} \tag{1}$$

where $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors in x- and y-coordinate axis directions.

It is a helpful practice to draw pictures like Figure 1 (b) so that the displacement $\mathbf{u}$ is temporarily strongly exaggerated. Quite often in the literature, a similar picture is drawn not showing the displacement but just a measure like y in Figure 1 (b). The analysis then continues with this y. Now the student may not be quite sure what this means. It looks more or less like an Eulerian description $y = y(x)$ and thus further, for example, at $x = L$, problems arise. The student has probably had this far only experience of the small displacement theory where the equilibrium equations have been derived with no consideration of the displacement field and no real reason to discuss the differences between the Eulerian and Lagrangian descriptions has emerged. (Of course, if the principle of virtual work is used to derive the equilibrium equations, the displacement field comes into consideration.) The buckling case is thus a suitable place to speak seriously about the Lagrangian description that is normally applied in solid mechanics and that x is here a material (not spatial as normally in fluid mechanics) coordinate.

Continue with some geometric considerations. The position vector of a generic point in the deformed configuration is seen to be

$$\mathbf{r} = x\mathbf{i} + u(x)\mathbf{i} + v(x)\mathbf{j} = s\mathbf{i} + u(s)\mathbf{i} + v(s)\mathbf{j}. \tag{2}$$

These two forms are presented to emphasize that either x or s can be used as the independent variable as they refer to same material point and have the same value at each point. Usually we are going to use at least finally the symbol x as the independent variable.

The unit vector **t**, tangent to the curved rod axis and pointing in the increasing direction of s, is obtained as

$$\mathbf{t} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}x} = (1+u')\mathbf{i} + v'\mathbf{j} \tag{3}$$

where now and later on $(\cdot)' = \mathrm{d}(\cdot)/\mathrm{d}x = \mathrm{d}(\cdot)/\mathrm{d}s$. The scalar product

$$\mathbf{t}\ \mathbf{t} \equiv (1+u')^2 + (v')^2 = 1 \tag{4}$$

and from this,

$$u' = \sqrt{1-(v')^2} - 1. \tag{5}$$

The transverse displacement component $v(x)$ will be as the basic dependent variable. We now simplify by remembering that the displacements in Figure 1 (b) are strongly exaggerated and in reality we are considering an arbitrary small neighborhood of the straight state so that $v \quad L$ and $v' \quad 1$. The latter relation is made use of in many formulas to follow.

The formula (5) shows how u depends on v. This relation is simplified by using the truncated series approximation

$$\sqrt{1-(v')^2} = 1 - \frac{1}{2}(v')^2 + O(v')^4 \tag{6}$$

giving when applied in (5)

$$u' = -\frac{1}{2}(v')^2. \tag{7}$$

Integration on both sides from 0 to x (and taking the boundary condition $u(0) = 0$ into account) gives the expression

$$u(x) = -\frac{1}{2}\int_0^x (v')^2\,\mathrm{d}\xi. \tag{8}$$

Formula (8) is needed for example if the energy approach is applied for expressing the potential energy of axial conservative loads.

The simplified expression for the unit tangent vector from (3) by using (7) becomes

$$\mathbf{t} = \left(1 - \frac{1}{2}(v')^2\right)\mathbf{i} + v'\mathbf{j} \approx \mathbf{i} + v'\mathbf{j}. \tag{9}$$

The unit normal vector **n** to curved rod axis will be needed as well. Defining it here to be directed 90 in the clockwise direction from **t** gives based on (9)

$$\mathbf{n} \approx -v'\mathbf{i} + \mathbf{j}. \tag{10}$$

Of course, (9) and (10) are no more exactly of unit length due to the approximations.

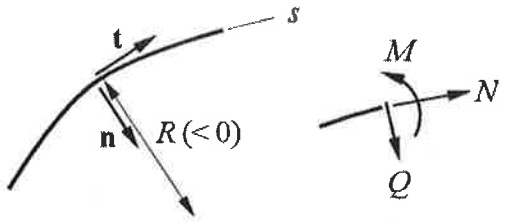
Finally, we will need also the curvature of the axis:

$$\frac{1}{R} \equiv \kappa = -\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s}\,\mathbf{n} \approx -v''\mathbf{j}\,\left(-v'\mathbf{i}+\mathbf{j}\right) = -v''. \tag{11}$$

The curvature is defined here to be positive when the center of curvature is on the opposite side of the direction of **n**, following the traditional beam theory. Formula (11) is obtained by recalling that $\mathrm{d}\mathbf{t}/\mathrm{d}s = -1/R\,\mathbf{n}$.

3. EQUILIBRIUM

It is advantageous in this connection if the student is familiar with the equilibrium equations for a curved plane rod. Following the notation in Figure 2 these equations are



(a) (b)

Figure 2 (a) Curved rod. **(b)** Stress resultants.

$$N' + \frac{Q}{R} + p_t = 0, \tag{12}$$

$$Q' - \frac{N}{R} + p_n = 0, \tag{13}$$

$$M' - Q = 0. \tag{14}$$

Here, N is the normal force, Q the shearing force, M the bending moment and R the radius of curvature of the rod axis. The quantities are drawn in the figure with their positive signs. The derivatives are with respect to the arclength s. The distributed load per unit arclength **p** is represented as

$$\mathbf{p} = p_t\,\mathbf{t} + p_n\,\mathbf{n}. \tag{15}$$

It is to be noted that equations (12) to (14) are exact. The equations have probably been derived in an earlier course in the case of an initially curved rod in connection with the small displacement theory. However, the equations are obviously valid also in the case of arbitrary large displacements. The current geometry must be used and if the axis is deforming in length it must be remembered that the independent variable in the equations is the current arclength s (which here however happens to coincide with x due to the inextensibility condition).

Further, if a point load **F** acts on the axis having the representation

$$\mathbf{P} = P_t\,\mathbf{t} + P_n\,\mathbf{n}, \tag{16}$$

the normal force and the shearing force obtain the jumps

$$N^+ - N^- = -P_t \tag{17}$$

and

$$Q^+ - Q^- = -P_n \tag{18}$$

in their values at that point. The plus- and minus superscripts refer to the limiting values on the positive and negative sides (with respect to s or x) at the point in question.

For slender rods perceptible for buckling it is customary to neglect the deformations due to shearing forces. Then the shearing force is a constraint force and it is useful to eliminate it. Differentiating (14) with respect to s and substituting Q' from (13) gives

$$M'' - \frac{N}{R} + p_n = 0. \tag{19}$$

This is one basic expression in buckling analysis of straight rods.

Equation (12) is approximated in the form

$$N' + p_t \approx 0, \tag{20}$$

that is, the term Q/R is neglected. The distribution of N can then be obtained from (20) if necessary in a piecewise manner by applying say the formula

$$N_2 - N_1 = -\int_{x_1}^{x_2} p_t \, \mathrm{d}\xi, \tag{21}$$

by making use of (17) at the pointloads and of the given value of N at one end of the rod.

4. ELASTIC MATERIAL LAW

For an elastic rod, following the Euler-Bernoulli beam theory, the bending moment is related to the curvature by the well-known expression

$$M = EI\kappa \tag{22}$$

where EI is the bending stiffness of the rod. Making use of (11), this becomes here

$$M \approx -EI\, v''. \tag{23}$$

From (14), the shearing force

$$Q \approx -\left(EI\, v''\right)'. \tag{24}$$

5. FINAL EQUATIONS

The basic differential equation (19) becomes now with (11) and (23) approximately as

$$-\left(EI\, v''\right)'' + N v'' + p_n = 0. \tag{25}$$

The distribution of the loading demands a detailed specification of the manner the loading is assumed to behave in the process of buckling. Two examples called "gravity loading" and "follower force loading" will be considered.

6. EXAMPLES

Gravity loading. If the distributed load **p** is due to the weight of the rod mass its direction and magnitude does not change during buckling. Thus we have the expression

$$\mathbf{p} = -f\mathbf{i} \tag{26}$$

where the intensity is denoted f. The components p_t and p_n in (15) are obtained by taking the scalar products of **p** with **t** and **n**. Thus, employing familiar approximations

$$p_t = \mathbf{p}\ \mathbf{t} \approx \mathbf{p}\ (\mathbf{i} + v'\mathbf{j}) = -f\mathbf{i}\ (\mathbf{i} + v'\mathbf{j}) = -f \tag{27}$$

and

$$p_n = \mathbf{p}\ \mathbf{n} \approx \mathbf{p}\ (-v'\mathbf{i} + \mathbf{j}) = -f\mathbf{i}\ (-v'\mathbf{i} + \mathbf{j}) = fv'. \tag{28}$$

Similarly, if a point load say due the weight G of a point mass is acting, we have the formula

$$\mathbf{P} = -G\mathbf{i}. \tag{29}$$

Proceeding as above, the components in (16) become

$$P_t \approx -G \tag{30}$$

and

$$P_n \approx Gv' \tag{31}$$

Consider a cantilever bar with constant bending stiffness as a simple example under the distributed own weight loading. Thus according to (27) and (28), the loading components are $p_t \approx -f$ and $p_n \approx fv'$. The normal force disappears at the free end, and is obtained by using (21)

$$N_2 - N_1 = -\int_{x_1}^{x_2} p_t \, \mathrm{d}\xi \ \Rightarrow\ 0 - N(x) = -\int_x^L -f \, \mathrm{d}\xi \tag{32}$$

resulting in

$$N = -f(L - x). \tag{33}$$

The field equation takes the form

$$-EIv'''' - f(L - x)v'' + fv' = 0 \tag{34}$$

which together with the boundary conditions $v(0) = v'(0) = v''(L) = v'''(L) = 0$ gives the final formulation for the eigenvalue problem, [1].

Follower force loading. By the follower force (also slave force) concept is normally used in connection with forces, which change their directions with the displacements of the system according to some rule. Here we consider the distributed loading

$$\mathbf{p} = -f\mathbf{t} \tag{35}$$

and the point loading

$$\mathbf{P} = -G\mathbf{t}. \tag{36}$$

These forces remain always tangential to the rod axis. Comparison of (35) with (15) gives simply

$$p_t = -f \tag{37}$$

and

$$p_n = 0 \tag{38}$$

and similarly, the point load has components

$$P_t = -G \tag{39}$$

and

$$P_n = 0. \tag{40}$$

When considering the same cantilever bar with constant bending stiffness as above, this under the distributed follower load, the loading components are $p_t = -f$ and $p_n = 0$. The normal force is the same as above (33). The final formulation of the eigenvalue problem consists of the field equation

$$-EIv'''' - f(L-x)v'' = 0 \tag{41}$$

and of the boundary conditions given above.

7. AXIAL DEFORMATION

If the axial deformation of the rod is taken into account, the equilibrium equations (12) to (14) of an initially curved rod can still be made use of. It is worth keeping in mind however, that in these equations the variable s has the role of the arclength and the differentiation has to be performed with respect to it.

Emphasizing again, the axial coordinate x (Figure 1 (a)) fixes a material point or more generally a cross-section in the initial state for the rod. It is the material or Lagrangian coordinate of the problem and when using the Lagrangian description, all quantities have to refer to it. If the axial strain ε is taken into account, the material fiber of length $\mathrm{d}x$ in the initial state becomes a fiber of the length $\mathrm{d}s = (1+\varepsilon)\mathrm{d}x$ in the deformed state. Thus we loose the direct connection for the arclength coordinate value s corresponding the value of x.

The differentiation with respect to x is denoted still by $(\ )'$ but differentiation with respect to s now by $(\ )^+$, see [2]. Chain differentiation gives

$$\frac{\mathrm{d}(\ )}{\mathrm{d}s}=\frac{\mathrm{d}(\ )}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}s}=\frac{1}{1+\varepsilon}\frac{\mathrm{d}(\ )}{\mathrm{d}x} \tag{42}$$

or

$$(\ )^{+}=\frac{1}{1+\varepsilon}(\ )'. \tag{43}$$

The equilibrium equations (12) to (14) are now

$$N^{+}+\frac{Q}{R}+\overline{p}_t=0, \tag{44}$$

$$Q^{+}-\frac{N}{R}+\overline{p}_n=0, \tag{45}$$

$$M^{+}-Q=0. \tag{46}$$

The distributed load per current unit arclength $\overline{\mathbf{p}}$ is represented as

$$\overline{\mathbf{p}}=\overline{p}_t\,\mathbf{t}+\overline{p}_n\,\mathbf{n}. \tag{47}$$

When the loading is referred per unit initial rod length, correspondingly

$$\mathbf{p}=p_t\,\mathbf{t}+p_n\,\mathbf{n}, \tag{48}$$

where

$$\overline{p}_t=\frac{1}{1+\varepsilon}p_t,\qquad \overline{p}_n=\frac{1}{1+\varepsilon}p_n. \tag{49}$$

When rule (43) is applied in (44) to (46) and when relations (49) are taken into account it will be obtained

$$\frac{1}{1+\varepsilon}N'+\frac{Q}{R}+\frac{1}{1+\varepsilon}p_t=0, \tag{50}$$

$$\frac{1}{1+\varepsilon}Q'-\frac{N}{R}+\frac{1}{1+\varepsilon}p_n=0, \tag{51}$$

$$\frac{1}{1+\varepsilon}M'-Q=0 \tag{52}$$

or

$$N'+(1+\varepsilon)\frac{Q}{R}+p_t=0, \tag{53}$$

$$Q'-(1+\varepsilon)\frac{N}{R}+p_n=0, \tag{54}$$

$$M'-(1+\varepsilon)Q=0. \tag{55}$$

These equilibrium equations can be now used for example in buckling analyses similarly as above in the case the axial deformation is taken into account. Of course, the kinematics must then also be altered appropriately but this will not be considered in this paper.

8. CONCLUSIONS

Two main points, when considering stability problems, will be emphasized in the present text: the Lagrangian description — as is conventional in structural mechanics — and the use of the equilibrium equations of initially curved rods. They will serve as an alternative view point to apply some basic principles for stability analyses.

REFERENCES

[1] Timoshenko, S.P., Gere, J.M., *Theory of Elastic Stability*, McGraw-Hill, 1961.

[2] Hodges, D. H.,1984, "Proper Definition of Curvature in Nonlinear Beam Kinematics", *AIAA Journal*, **22**, pp. 1825-1827.

STRESS AND DISPLACEMENT ANALYSIS OF A MONORAIL BEAM INCLUDING THE DISTORTION OF THE CROSS-SECTION

T. HALME
Laboratory of Fatigue and Strength
Lappeenranta University of Technology
Box 20, FIN-53851 Lappeenranta
e-mail: tapani.halme@lut.fi

S. HEINILÄ
Laboratory of Fatigue and Strength
Lappeenranta University of Technology
Box 20, FIN-53851 Lappeenranta
e-mail: sami.heinila@lut.fi

ABSTRACT

The basic assumption in the classical Euler-Bernoulli-Timoshenko beam theory is the non-deformable cross-section of the beam in structural analysis. This assumption is usually not a problem in basic structural members consisting of fairly thick walls as in hot-rolled beams. However, modern structural members are increasingly made of cold-formed thin-walled profiles. In these structures the deformation of the cross-section or distortion cannot be excluded in the analysis. The first-order Generalised Beam Theory (GBT) offers the possibility to evaluate the distortional effects of different loading combinations in thin-walled structural members. In GBT the total deformation of the structure is separated into cross-sectional and longitudinal deformation modes, which are then combined in the final analysis. The basic transverse deformation modes, i.e. bending about principal axes and torsion, are the same as in classical beam analysis. In GBT the distortional modes are added to these basic modes. In the first stage of the analysis the basic cross-sectional properties are calculated for all modes using warping of the cross-section as the unifying quantity. In the second stage a prismatic beam with arbitrary loading and end constraints can be analysed by the superposition of the different deformation modes arising from given loading. In this paper a stress and displacement analysis of a beam used in so-called monorail system is analysed in the context of the first-order GBT. The analysis is verified by comparing the results obtained from finite element analysis using thin shell elements.

1. INTRODUCTION

Cold-formed structural members have become more and more attractive as basic structural elements for designers and end users of buildings and machines for a number of reasons. Thin-walled sections can be produced from steel sheet less than 1 mm thick, producing very lightweight structures with almost unlimited scope of cross-sectional geometries. The introduction of high strength steels has increased the efficiency of these structures considerably. These improvements in manufacturing and metallurgical processes has generated a need for predictive and reliable design with regard to the various behaviour and failure modes inherent in these structures, e.g.

stresses due to distortion, buckling behaviour, vibrations, and non-linear material and geometrical behaviour. Generalised Beam Theory (GBT), developed by Prof. Schardt and his associates in Darmstadt, Schardt (1989), is a tool for the analysis of prismatic thin-walled structures based in the theory of Vlasov (1961). Vlasov used warping to define the cross sectional properties in non-uniform or warping torsion. GBT unifies the concept of warping in all deformation modes of structural members from axial deflection and beam bending to torsion and distortion. GBT handles the different degrees of freedom, i.e. extension, bending about the principal axes, torsion and the distortional degrees of freedom as deformation modes, which according to first-order theory are orthogonal. This means that they are uncoupled and can be analysed separately before their effects are combined with the simple procedure of superposition, as in the classical beam bending analysis.

In this paper a stress and displacement analysis of a monorail beam is performed as a case study using the first-order GBT.

2. THEORETICAL BACKGROUND OF GBT

The generalised displacement $V(x)$, i.e. transverse deformation along the length of a prismatic beam, can be solved from the basic equation of GBT, which is written in a form

$$E\,{}^{k}C\,{}^{k}V'''' - G\,{}^{k}D\,{}^{k}V'' + {}^{k}B\,{}^{k}V = {}^{k}q\,. \tag{2.1}$$

The index k refers to the deformation modes or eigenmodes of a prismatic cross-section which are independent of each other. The first two cross-section properties above are analogous the well known cross-section properties of the classical beam theory. The warping constant ${}^{k}C$ is analogous to the flexural moment of inertia or sectorial moment of inertia in torsion and ${}^{k}D$ is the torsion constant. The third property, ${}^{k}B$, is the transverse bending stiffness, which defines the stiffness of the cross-section in distortion. The cross-section deformation modes can be divided into two basic modes. The first is rigid section modes, which are extension, bending about principal axes, and torsion. The flexible section modes are the distortional modes starting from mode $k = 5$.

2.1 Basic idea of GBT

Prismatic structures have two main directions: the longitudinal direction, x, and the transverse direction, s. Thus the total deformation of the structure can be expressed as the sum of product functions $F(s)\cdot V(x)$ where $F(s)$ expresses the relative cross-sectional deformation, which also includes distortion, and $V(x)$ (generalised displacement), which is the amplitude function along the x-axis. In Fig. 2.1 a prismatic structure presents the division of displacement functions in the two main directions.

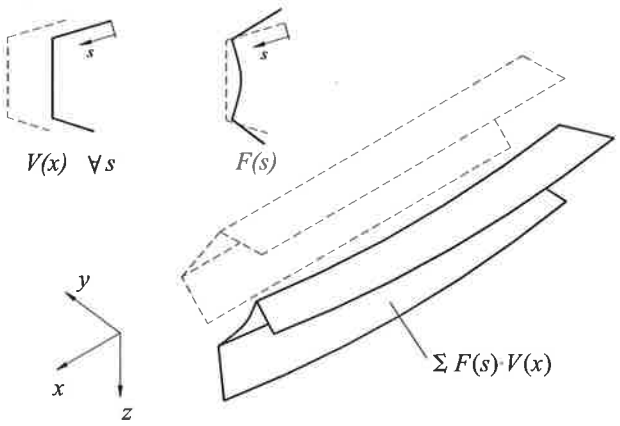


Figure 2.1 The total deformation of a prismatic structure can be divided into two displacements functions, i.e. longitudinal displacement $V(x)$ and cross-sectional deformation $F(s)$.

Both functions must fulfil the boundary conditions of the structure, $F(s)$ transversally and $V(x)$ longitudinally. This is the fundamental idea in the context of GBT. The transverse function $F(s)$ is always a relative quantity, which depends on the cross-section geometry, but the longitudinal displacement function is either absolute, as in the first-order analysis (stress and displacement analysis), or relative as in the second-order analysis (stability or vibration analysis).

2.2 First-order GBT

The basic procedure before any analysis type is used in GBT is the calculation of the orthogonal deformation modes of a given cross-section. As an example the first eight deformation modes of a lipped channel profile is presented in Fig. 2.2:

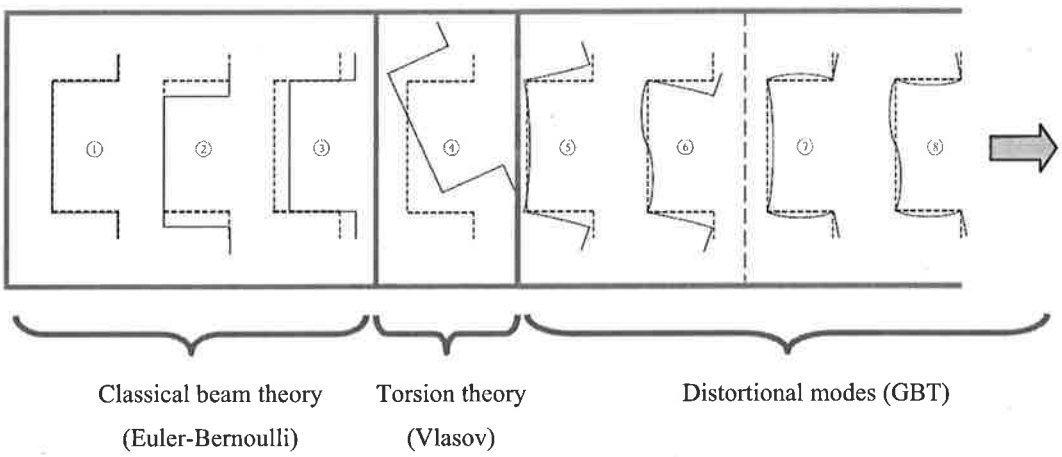


Figure 2.2 The first eight deformation modes of a lipped channel profile.

The maximum number of deformation modes k of an open thin-walled cross-section is $l + 1$ where l is the number of plate strips. The lipped channel section shown in Fig. 2.2 therefore has six 'natural' deformation modes. However, as the total number of deformation modes is infinite and

when analysis requires more than $l + 1$ deformation modes, the solution is to divide the strips using intermediate nodes between the corner points so that one or more plate strips are created in the cross-section. A typical cross-section of a thin-walled profile, as is shown in Fig. 2.2, consists of folded plates. In GBT these plates or strips are the basic structural elements from which the governing differential equation and the corresponding cross-sectional properties are derived. The global and local co-ordinate system of a folded plate structure is shown in Fig. 2.3:

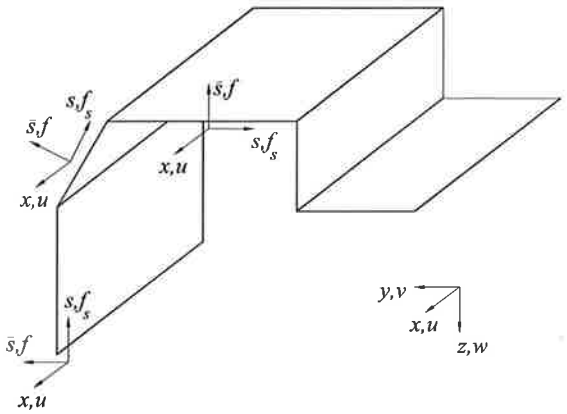


Figure 2.3 Global and local co-ordinate system of a folded plate structure.

In GBT three different degrees of freedom or displacements are required to define displacement in one plate strip. The three displacements are the axial displacement, u, the in-plane displacement, f_s, and the out-of-plane displacement, f. In Fig. 2.4 the displacement functions of one (deformed) plate strip of a profile are shown in addition to the local co-ordinate system of a profile:

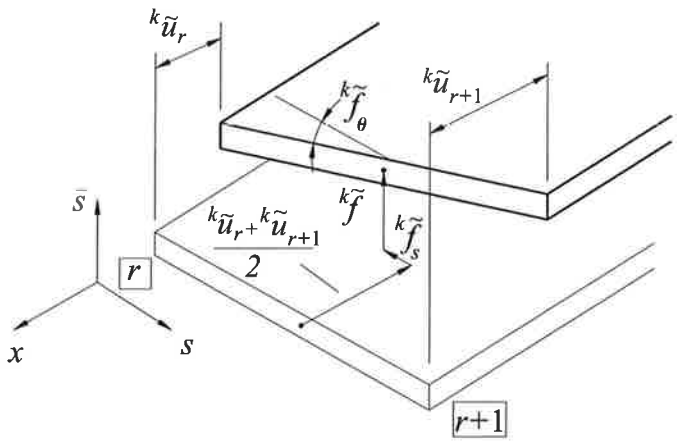


Figure 2.4 The (unit) displacement functions $\tilde{u}, \tilde{f}$ and $\tilde{f}_s$ of a plate strip and the local co-ordinate system.

Thus the total deformation of a prismatic thin-walled structure may be formulated in the following terms as

$$\left.\begin{aligned} u(x,s) &= \sum_k {}^k\tilde{u}(s)\,{}^kV'(x) \\ f_s(x,s) &= \sum_k {}^k\tilde{f}_s(s)\,{}^kV(x) \\ f(x,s) &= \sum_k {}^k\tilde{f}(s)\,{}^kV(x) \end{aligned}\right\} \equiv \sum F(s)V(x). \tag{2.2}$$

Warping in the cross section can be divided into two parts: primary warping or membrane warping caused by in-plane-displacement, f_s, and the secondary warping or plate bending warping caused by out-of-plane displacement, f. In plate structures only out-of-plane displacement is present in all the elementary deformation modes.

The fundamental displacement function is the warping function ${}^ku$, which is the basic displacement quantity in defining the cross-sectional properties as well as the other displacement functions of each mode. The warping function ${}^ku$ is the unifying feature of GBT as it describes the axial strain of each mode k and gives the orthogonality condition

$$\int_A {}^i\tilde{u}\,{}^k\tilde{u}\ \mathrm{d}A = 0, \qquad i \neq k\,. \tag{2.3}$$

Eq. 2.3 indicates that the strain energy of the normal stress of one warping function (mode i) and the strain of another warping function (mode k) is zero. Each deformation mode, k, is defined by corresponding warping function, ${}^ku$, which is directly linked to the lateral cross-section deformations f and f_s in the local co-ordinate system (x,s). These warping functions, more familiar from Vlasov's torsion theory of beams, are thus functions describing the axial displacement along the centre line of a wall of a thin-walled profile as the cross-section moves according to certain deformation mode. These warping functions must be orthogonal and, consequently, the deformation modes appear orthogonal. When these orthogonal warping functions are found, the resulting orthogonal deformation modes and their cross-sectional properties can be calculated.

The derivation of the basic equation of GBT is obtained by using the variational principle. The total virtual work can be obtained by integrating the element virtual work over the whole continuum. The membrane and bending parts are separated which leads to an expression

$$\begin{aligned} \delta W = \int_L \int_A (&\underbrace{\sigma_x^{\mathrm{M}} \cdot \delta\varepsilon_x^{\mathrm{M}}}_{1.} + \underbrace{\sigma_x^{\mathrm{B}} \cdot \delta\varepsilon_x^{\mathrm{B}}}_{2.} + \underbrace{\sigma_s^{\mathrm{M}} \cdot \delta\varepsilon_s^{\mathrm{M}}}_{=0} + \underbrace{\sigma_s^{\mathrm{B}} \cdot \delta\varepsilon_s^{\mathrm{B}}}_{3.} + \underbrace{\tau_{sx}^{\mathrm{M}} \cdot \delta\gamma_{sx}^{\mathrm{M}}}_{=0} + \underbrace{\tau_{sx}^{\mathrm{B}} \cdot \delta\gamma_{sx}^{\mathrm{B}}}_{4.})\,\mathrm{d}A\,\mathrm{d}x \\ &- \int_L \int_s \left(p_v(x,y)\cdot\delta v + p_w(x,y)\cdot\delta w\right)\mathrm{d}s\,\mathrm{d}x = 0. \end{aligned} \tag{2.4}$$

Thus in the first-order theory of GBT the membrane transverse normal stresses (σ_s^{M}) and the membrane shear deformation (γ_{sx}^{M}) are neglected. When variational principle is used in deriving expressions from Eq. 2.4 the variational symbol δV can be directly attached to the displacement functions in Eq. 2.2. The integrations lead to strongly coupled differential equations.

After the orthogonalising process of the deformation functions, which is basically solution of the general eigenvalue problem, the uncoupled differential equations emerge and they can be recognised by the forward subscript k as seen in the basic equation of the first-order GBT, i.e. Eq. 2.1. The modal cross-sectional properties for each deformation mode k are then obtained as (tilde above a symbol denotes normalised value)

$$
\begin{aligned}
{}^{k}C &= \int_{A} {}^{k}\tilde{u}^{2}\,\mathrm{d}A + \frac{K}{E}\int_{s} {}^{k}\tilde{f}^{2}\,\mathrm{d}s \\
{}^{k}D &= {}^{k}D_{1} - \frac{\nu \cdot E}{G}({}^{k}D_{2} + {}^{k}D_{2}) = \frac{t^{3}}{3}\int_{s} {}^{i}\dot{\tilde{f}}^{2}\,\mathrm{d}s - \frac{2\nu K}{G}\int_{s} {}^{k}\ddot{\tilde{f}}\;{}^{k}\tilde{f}\,\mathrm{d}s \qquad (2.5) \\
{}^{k}B &= K\int_{s} {}^{k}\ddot{\tilde{f}}^{2}\,\mathrm{d}s \qquad K = \frac{Et^{3}}{12(1-\nu^{2})}.
\end{aligned}
$$

In addition to the warping function, u, in-plane displacement f_s and out-of-plane displacements f, as well as plate rotation f_θ, transverse bending distribution m_r, shear force S_r and transverse displacements v and w must be calculated for each mode k.

The loading of the structure is actually a virtual work done by the external load according to the first-order GBT:

$$
{}^{k}q_{r,v} = q_{r,y}\,{}^{k}\tilde{v}_{r}, \qquad {}^{k}q_{r,w} = q_{r,z}\,{}^{k}\tilde{w}_{r}. \qquad (2.6)
$$

When the deformation function ${}^{k}V$ is obtained after solving the differential equations for each mode, the solutions are combined to obtain the total deflections and stresses of the structure. The transverse deflections of any point r in the cross-section are calculated according to

$$
w_{r} = \sum_{k=2}^{l+1} {}^{k}V\,{}^{k}\tilde{w}, \qquad v_{r} = \sum_{k=2}^{l+1} {}^{k}V\,{}^{k}\tilde{v}. \qquad (2.7)
$$

The generalised force component is called the stress resultant (or generalised warping moment) ${}^{k}W$ based on the second derivative of the deformation function:

$$
{}^{k}W = -E\,{}^{k}C\,{}^{k}V''. \qquad (2.8)
$$

The longitudinal stress distribution is once again the superposition of stresses of different modes:

$$
\sigma_{x}(s) = -\sum_{k=1}^{l+1} \frac{{}^{k}W\;{}^{k}\tilde{u}(s)}{{}^{k}C}. \qquad (2.9)
$$

The preceding outline of first-order GBT is a general one and can be applied to arbitrary cross-sections and boundary conditions of a prismatic structure.

3. CASE STUDY OF A MONORAIL BEAM

3.1 The cross-section, loading, boundary conditions and the deformation modes of the beam

A stress and displacement analysis of a monorail beam is performed using the first-order GBT. The beam is simply supported at the ends with wheel load in the middle of the beam. The dimensions of the cross-section are shown in Fig. 3.1 as well as the location of the wheel loads. The length of the beam is $L = 6000$ mm and the wheel load is $F = 4$ kN.

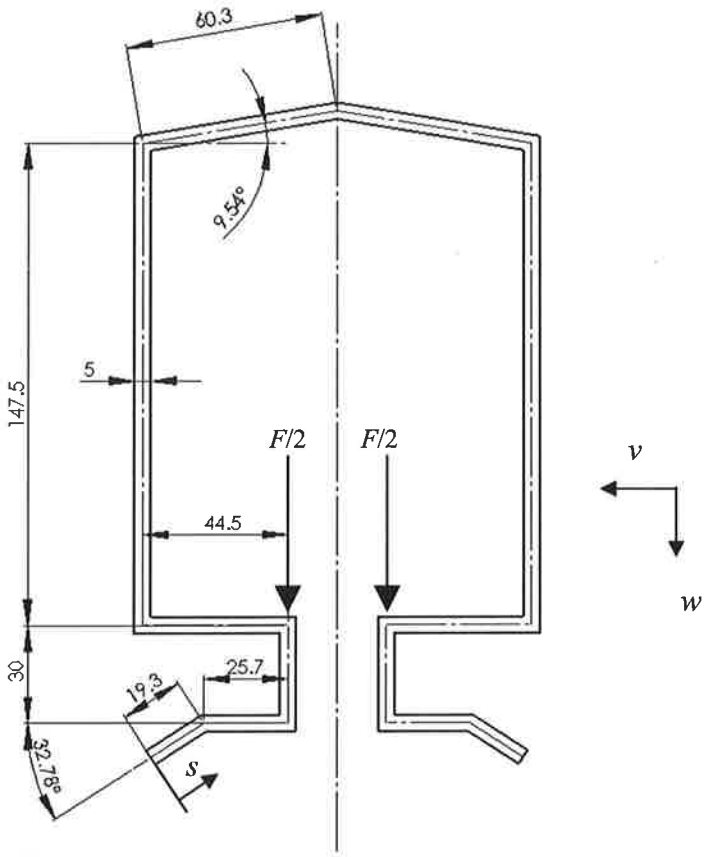


Figure 3.1. *Cross-section of the beam showing the dimensions in* [mm], *loading points and coordinate system*

Material is steel with material properties $E = 210000$ MPa, $G = 80000$ MPa and $v = 0{,}3$. The axial membrane normal stress at the free end of the beam ($s = 0$) was calculated. Additionally, the vertical displacement w was calculated at the line of the wheel load ($s = 75$ mm). The first fifteen deformation modes of the cross-section are shown in Fig. 3.2 based on the cross-sectional analysis according to GBT using program VTB, Schardt (VTB-program, 1996). Units kN and cm were used in the calculation.

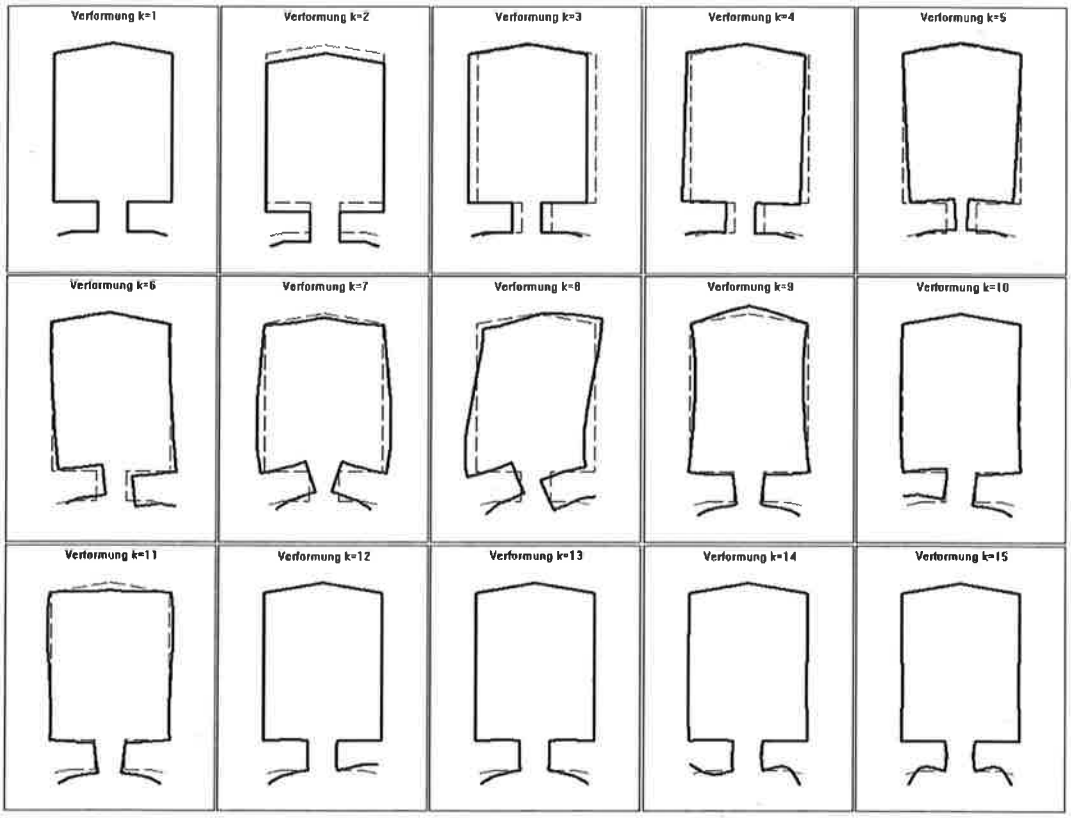


Figure 3.2 *First fifteen deformation modes of the cross-section*

The first deformation mode is the axial displacement. In this mode the warping function is constant, i.e. ${}^{1}\tilde{u} = -1$. The next two modes are the bending deformation about the principal axes and the fourth one is the torsion mode, i.e. rotation about the shear centre. The following are the distortional modes. As the loading is symmetrical, only the symmetrical modes are required in the analysis. From these only three modes, i.e. modes 2, 5 and 7 were chosen for the analysis. The reason is that these modes produce the most significant axial membrane stresses at the chosen point of interest. The other modes produce mainly transverse normal stresses due to plate bending, as can be seen from the displacements in Fig. 3.2. However, these stresses were not calculated in this case study. The main cross-sectional properties of the chosen three modes are presented in Table 3.1.

Table 3.1. Cross-sectional properties of modes 2, 5 and 7

	Warping constant	Torsion constant	Transverse bending stiffness
Deformation mode k	${}^{k}C$ [cm^4]	${}^{k}D$ [cm^2]	${}^{k}B$ [kN/cm^2]
$k = 2$	1520,63	0	0
$k = 5$	3,7187	8,8068E-4	0,013324
$k = 7$	6,4567	6,9248E-4	0,463127

The loading of the beam in GBT is based on the virtual work done by the external forces as was expressed in Eq. 2.6. Table 3.2 shows the normalised eigendisplacements in the direction of the load for the chosen modes. These values are also obtained in the process of calculating the cross-sectional properties of the profile.

Table 3.2. The normalised eigendisplacement of the loading point

${}^{2}\tilde{v}_F$	${}^{5}\tilde{v}_F$	${}^{7}\tilde{v}_F$
1,000	0,125	0,220

Table 3.3 shows the warping function value at the free end of the cross-section where the axial normal stress is calculated in post-processing phase using Eq. 2.9.

Table 3.3 Warping function value at the free end of the cross-section [cm]

${}^{2}\tilde{u}(0)$	${}^{5}\tilde{u}(0)$	${}^{7}\tilde{u}(0)$
-9,5978	-1,0000	-1,0000

The differential equations for the chosen modes are thus

$$\begin{aligned} &E\,{}^{2}C\,{}^{2}V''''(x) = {}^{2}q(x,s) \\ &E\,{}^{5}C\,{}^{5}V''''(x) - G\,{}^{5}V''(x) + {}^{5}B\,{}^{5}V(x) = {}^{5}q(x,s) \\ &E\,{}^{7}C\,{}^{7}V''''(x) - G\,{}^{7}V''(x) + {}^{7}B\,{}^{7}V(x) = {}^{7}q(x,s) \end{aligned} \tag{3.1}$$

The generalised warping moment (Eq. 2.8) and the elastic line along any point of the cross-section (Eq. 2.7) can then be calculated from the resulting mode amplitudes ${}^{k}V(x)$.

Deformation mode 2, i.e. bending about the main axis, can be directly solved from textbook tables as the differential equations for the GBT and for the classical beam theory are analogous. The other two differential equations can be solved using solutions of the beam on elastic foundation (BEF). The general expression of the differential equation of an axially loaded beam on elastic foundation is

$$E\,I\,v''''(x) - N\,v''(x) + k\,v(x) = q(x)\,, \tag{3.2}$$

which is clearly analogous with the differential equation of the first-order GBT. The conversions on the next page must be made when using the available analytical solutions either from the classical beam theory or from the BEF.

Moment of inertia = modal warping constant	$I = {}^{k}C$
Axial load = modal torsional stiffness	$N = G\,{}^{k}D$
Foundation modulus = modal transverse bending stiffness	$k = {}^{k}B$
Elastic line = product of eigendisplacement and amplitude	$v(x,s) = {}^{k}\tilde{v}(s)\,{}^{k}V(x)$
	$w(x,s) = {}^{k}\tilde{w}(s)\,{}^{k}V(x)$
Bending moment = generalised warping moment of a mode	$M = {}^{k}W$
Distributed load = modal virtual work done by distributed load	$q_y = q_v\,{}^{k}\tilde{v}_q,\ q_z = q_w\,{}^{k}\tilde{w}_q$
Point load = modal virtual work done by point load	$F_y = F_v\,{}^{k}\tilde{v}_F,\ F_z = F_w\,{}^{k}\tilde{w}_F$

The distortional deformation modes due to point loads typically dampen quite rapidly. As the beam is long compared to the cross-section dimensions, the analytical results for an infinite beam loaded by a point load F_y, see Fig. 3.3, can be used to solve modes 5 and 7. The bending moment and elastic line are (Hetenyi, 1946):

$$M(x) = \frac{F_y}{4\alpha\beta} e^{-\alpha x}\left(\beta\cos(\beta x) - \alpha\sin(\beta x)\right)$$

$$v = \frac{F_y}{2k}\frac{\lambda^2}{\alpha\beta} e^{-\alpha x}\left(\beta\cos\beta x + \alpha\sin\beta x\right) \tag{3.3}$$

$$\lambda = \sqrt[4]{\frac{k}{4EI}},\quad \alpha = \sqrt{\lambda^2 + \frac{N}{4EI}},\quad \beta = \sqrt{\lambda^2 - \frac{N}{4EI}},\ N > 2\sqrt{kEI}.$$

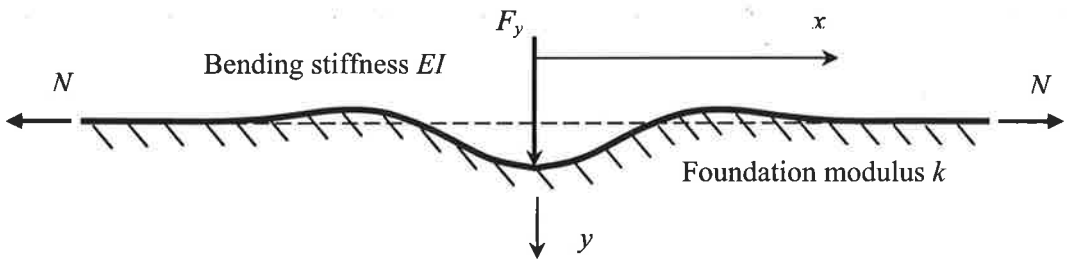


Figure 3.3 *Infinite beam on elastic foundation with transverse point load F_y*

3.2 Finite Element Model

The beam was modeled with finite element method (FEM) for comparison with the results from GBT. Due to double symmetry only a quarter FE model was needed. The FE modeling was done with Femap v9.1 pre- and postprocessor and the solution was obtained with NX/Nastran solver. The model (Fig. 3.4) consisted of ca. 2800 linear four-node shell elements with uniform thickness of 5 mm and ca. 2900 nodes. The material properties were E = 210 GPa and ν = 0,3. In addition to symmetry boundary conditions applied to symmetry planes, the transverse translational degrees of freedom were constrained at the end of the beam to model the simply supported beam ends. A point load of ¼ of the total load was applied to the point of wheel load.

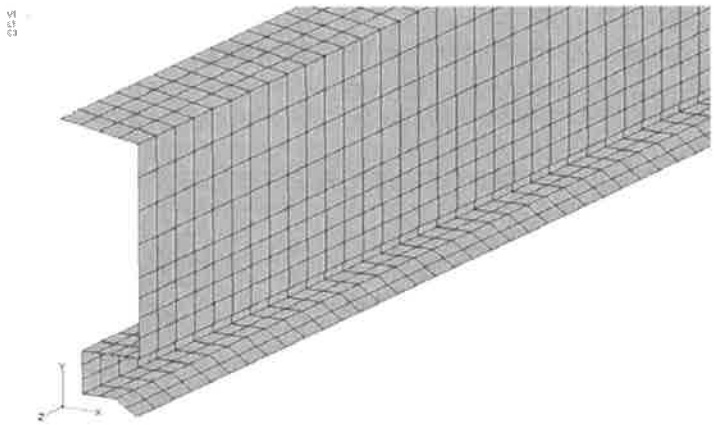


Figure 3.4 *Finite element model*

3.3 Results

The results from GBT for the axial membrane normal stress due to distortion at the free edge of the cross-section is given in Fig. 3.5 between lengths 1000-3000 mm from the end support. The dampening of the distortional normal stresses is clear already in this area.

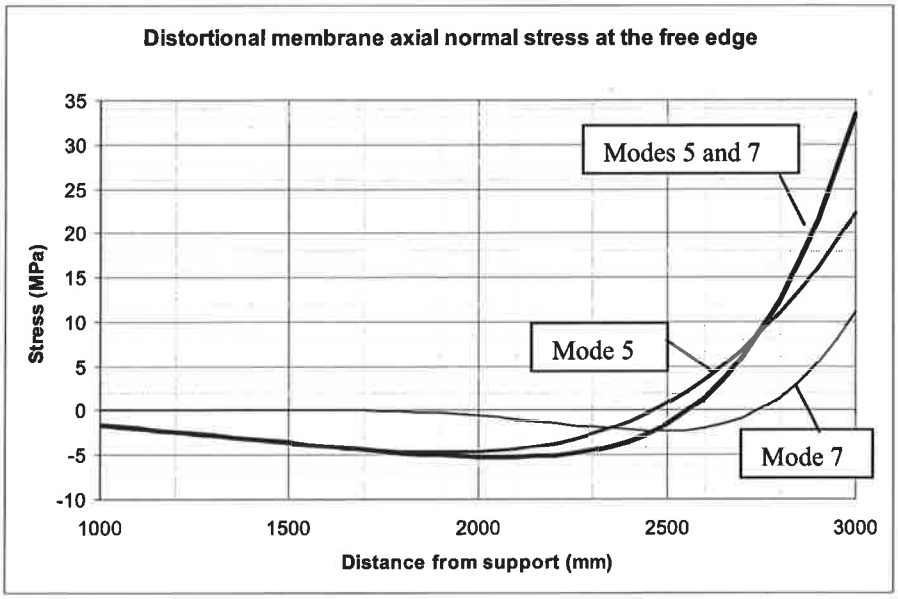


Figure 3.5 *Normal stresses (modes 5 and 7) at the free edge of the profile*

The axial stresses as the sum of all chosen deformation modes is shown in Fig. 3.6, which gives also comparison from the results of the finite element analysis.

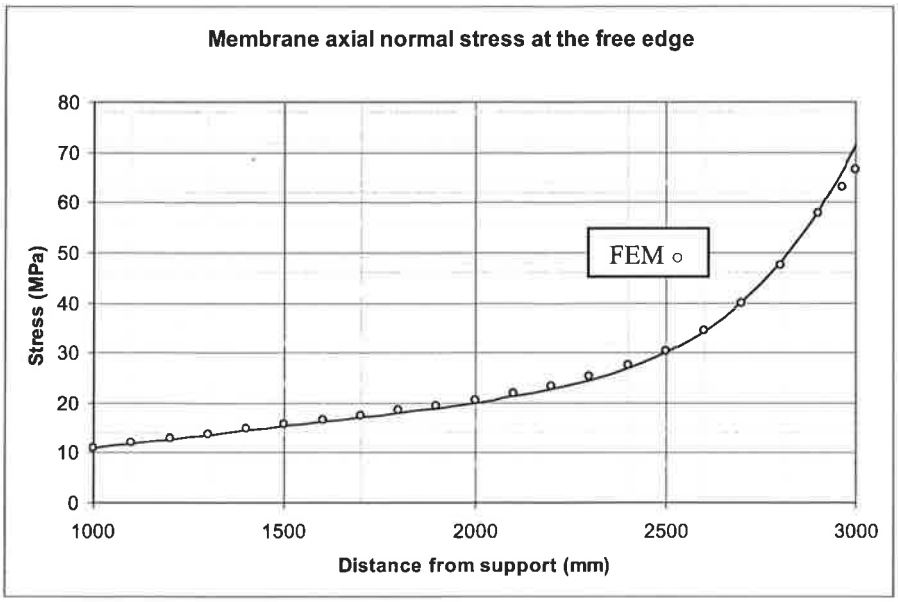


Figure 3.6 *Normal stresses at the free end of the profile compared to results from FE-analysis*

Displacement along the load line of the cross-section is presented in Figs. 3.7 and 3.8. The vertical displacement w due to distortional deformation of modes 5 and 7 is presented in Fig. 3.7. The investigated section is between 2000-3000 mm from the support. The displacement due to mode 7 is already insignificant at a distance of ca 800 mm from the point load.

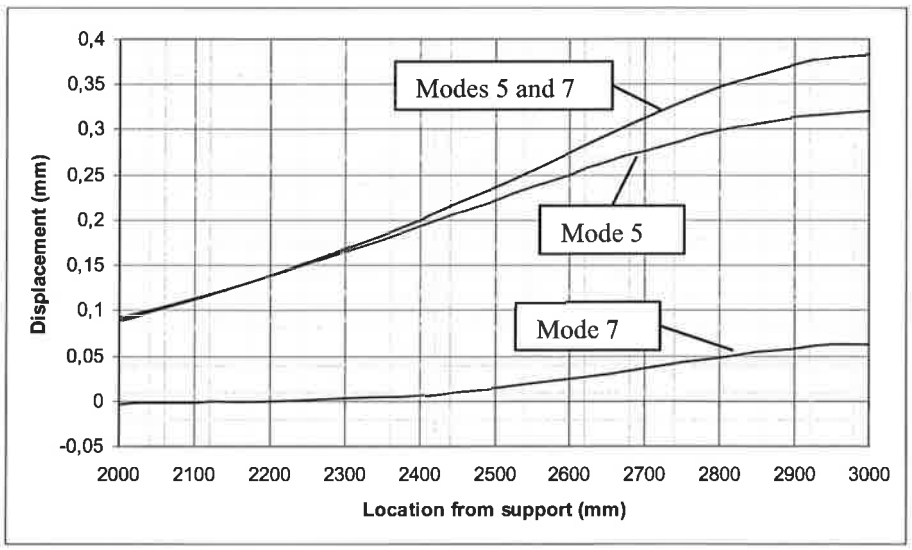


Figure 3.7 *Vertical displacement (modes 5 and 7) at the load line*

The total vertical displacement is presented in Fig. 3.8. The analytical solution is slightly below the finite element solution, which can at least partly be explained by the exclusion of the shear deformation in the first-order GBT.

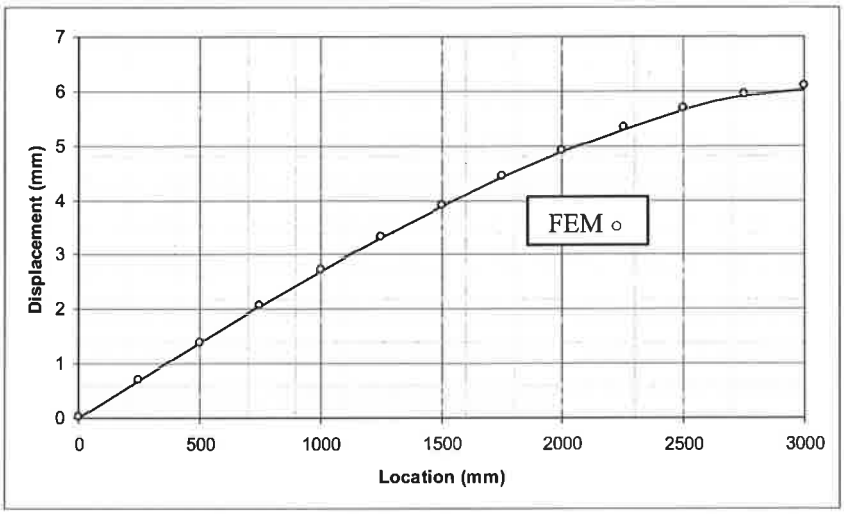


Figure 3.8 *Vertical displacement at the load line compared to results from FE-analysis*

4. CONCLUSIONS

The basic assumption in the classical Euler-Bernoulli-Timoshenko beam theory is the non-deformable cross-section of the beam in structural analysis. However, using Generalised Beam Theory, the deformation or distortion of the cross-section can be included in the analysis. The generalisation is based on the work of Vlasov and the theory unifies the different beam deformations, i.e. classical elongation and compression as well as bending and torsion, to distortional deformation modes using the so-called warping functions. When these warping functions are found, all the required modal cross-sectional properties of a deformation mode can be calculated for a given cross-section. The second phase involves solving the governing differential equation to obtain the required information in the longitudinal direction. This structural analysis usually requires a numerical method, e.g. the finite element or finite difference methods, to solve the governing equation for different analysis options. However, as the form of the governing equation is analogous to the differential equation of beam on elastic foundation, certain simple cases can be analytically solved using the results obtained for BEF. In this paper a monorail beam is analysed using GBT. The results are compared with analysis results from FEM. The comparison clearly shows the efficiency of the first-order GBT to predict the stresses and displacements of a prismatic structure made of thin-walled folded plates.

REFERENCES

Hetenyi, M., Beams on Elastic Foundation, University of Michigan Press, Ann Arbor, MI, 1946.

Schardt, R., Verallgemeinerte Technische Biegetheorie. Springer-Verlag, Berlin, 1989. ISBN 3-540-51339-6.

Schardt, R., Program VTB, 1996.

Vlasov, V.Z., Thin-walled Elastic Beams. Israel program for scientific calculations. Translation of: Tonkostennye uprugie sterzhni, Jerusalem 1961, 493 p.

EXPERIMENTAL SET-UP AND NUMERICAL SIMULATION OF SOFT MISSILE IMPACT SMALL SCALE TESTS

P. VARPASUO Fortum Nuclear
Services Ltd
Keilaniementie 1, Espoo, Finland
P.O.B 10, 00048 Fortum
Tel. +358104532223
Fax. +358104533985
Mobile. +358407250397
e-mail. pentti.varpasuo@fortum.com

ABSTRACT

This paper describes the experimental set-up and numerical simulation of soft missile impact small scale tests. The main aim of the testing program is to provide data for the calibration and verification of the numerical simulation models for soft missile impact time-history simulation.
The experiments are conducted both with dry and fluid filled missiles. The objective of the tests with fluid filled missiles is to investigate the release speed and the droplet size of the fluid spray release in the impact. This information is vital in assessing the combustion hazard of potentially flammable liquid after the release. The spray release velocity and droplet size are also important input data for analytical and numerical simulation methods which can be applied to assess the spreading of the liquid in the impact.
The second object of the testing program is to develop and adopt in use the numerical methods for predicting the response of reinforced concrete structures to impacts of deformable projectiles with or without liquid content. The pre-test and post-test analyses are used for predicting and simulating the impact load time-history, structural behavior of both the missile and the target with the aid of experimental text book formulas and also with the aid of non-linear finite element method.

1. INTRODUCTION

The studies of the effects of hard missiles on the impact target have been described in the past by various researchers (Riera, 1968, Johnson, 1972). Due to its stiffness properties and mass properties the impacting aircraft can be regarded as a soft missile. Riera has derived the algorithm for the calculation of the reaction force time-history in case of soft missile. For further development of this algorithm, an experimental test program has been started in VTT within the framework of the Finnish program for the enhancement of nuclear safety (Hakola, Karna, 2005). The test facility uses the pneumatic cannon to shoot the missile to its target.
The principal aim of the test program is to acquire information for the calibration and verification purposes simulation models. The experiments with fluid filled missiles are also used to acquire information about the release speed, drop size and spreading area of the liquid fuel in impact. This data is important for studies of the fuel deflagration, and for analytical and numerical simulation methods for modeling the phenomena connected with jet fuel in case of aircraft impact.

2. IMPACT TEST SET UP

The requirements for the test facility were specified as follows. In the first place the impact tests should be devised to obtain information on impact force-time function in case of the soft missile impacting against a rigid wall. Secondly, the validity of the published softening factors used to determine the penetration depth of deformable missiles using correlations derived from hard missile data should be investigated. Thirdly, the effect of multiple small-sized hard missiles coupled with the effect of large soft missile impact should be studied. Fourthly, the failure modes of a concrete wall due to impacting missile and release and spreading of liquid contained within the missile are researched. Lastly, the shedding of the missile debris after an impact is studied.
The shape of the missile in most test cases is cylindrical; later on in the test program the missile shape will be varied in order to study the coupled effect of large soft missile impact and smaller hard missile impacts. The materials of the missile are steel and aluminum. The mass of the missile is variable up to 100 kg. The impact velocity of the missile is variable between 100 and 200 meters per second. Part of the missiles is dry and part of the missiles is liquid filled. The test monitoring instruments include high-speed video cameras, force time function measuring sensors, target wall displacement and deformation measuring sensors and gauges and measuring devices for the debris scatter. The test setup is depicted in the Figures 1 and 2:

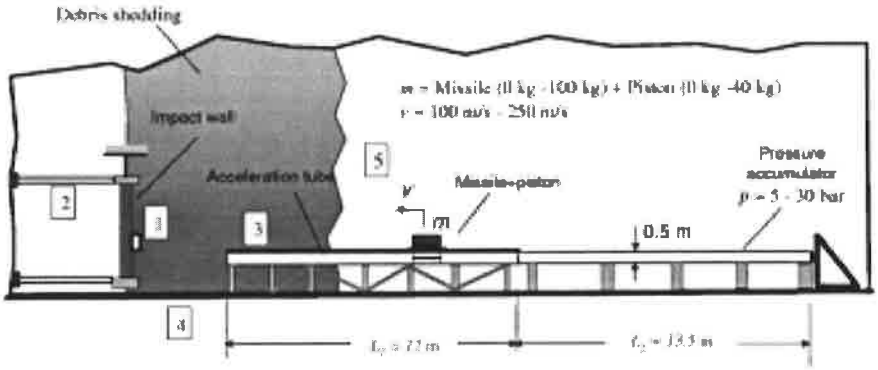


Figure 1 The schematic test set-up for pneumatic soft missile impact tests

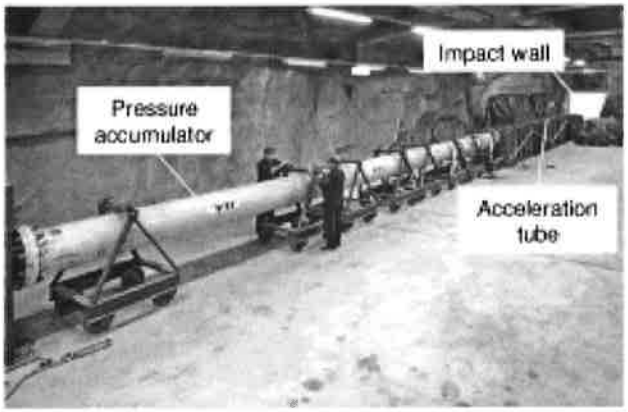


Figure 2 The photograph of test set-up for soft missile impact

3. INSTRUMENTATION SCHEME

The measuring devices used in the tests are shock accelerometers, strain gauges and force transducers. The location of the accelerometers are is shown in Figure 3. The notation for accelerometers in Figure 3 is AM1, AM2, AM3 and AM4. The impact force is also measured using the thick steel plate in front of the target wall and the force transducers behind the steel plate. The notation for the force transducers is FT1, FT2 and FT3. The force plate in front of the wall is shown in Figure 3. The third system for measuring the impact forces are the strain gauges installed on the supporting pipes behind the wall. The notation of the strain gauges in Figure 3 is SG1, SG2, SG3 and SG4.

The speed of the missile is measured with laser equipments in front of the target wall. The notation for lasers in Figure 3 is LA1 and LA2. The second system for the impact speed of the missile are the pressure sensors installed inside the acceleration tube.

All the data values are collected using National Instruments PXI-1000B rack with 2 PXI 4472B data acquisition cards. The sampling frequency is 102.4 kHz with anti-aliasing frequency of 50 kHz. The data is saved onto the hard drive of the PC for subsequent analysis.

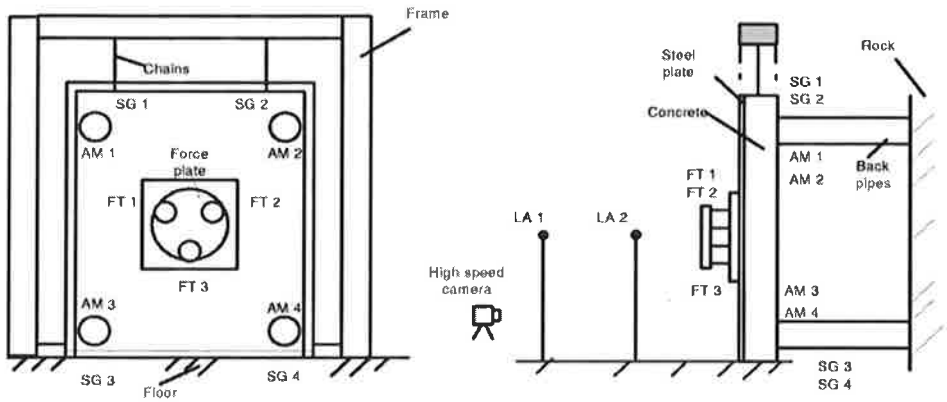


Figure 3 Instrumentation of the test

Figure 4 shows the schematic location of the cameras used for liquid spread monitoring for liquid filled missile tests.

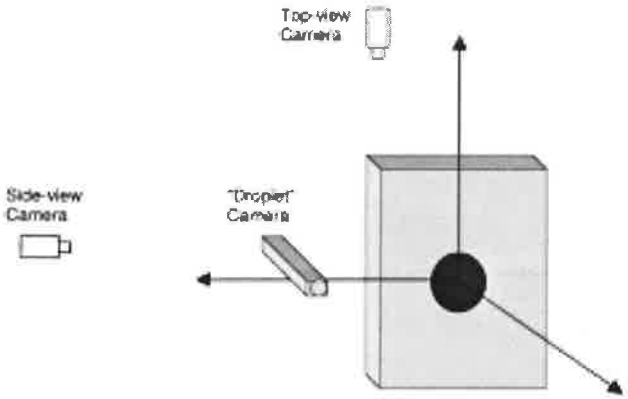


Figure 4 Camera locations for liquid filled missile tests

3. SPECIFICATION TABLE FOR TESTS USED IN NUMERICAL SIMULATION

The test set-up has been used during the year 2005 to obtain preliminary information on the missile impact. The purpose of the preliminary tests was to gain experience and proficiency in the monitoring of the and in numerical simulation methods in order to calculate the reaction force time-history during impact. In the preliminary test, a wide spectrum of different impact velocities and masses were used. The weights of the missiles varied between 22 and 95 kilograms and the impact speeds of the missiles varied between 70 and 170 meters per second. The first sixteen missiles in the year 2005 were spiral-welded steel pipes and the last four were aluminum pipes. Seven of the missiles shot during the year 2005 were filled with water.

Test No	Velocity (m/s)	Missile	Weight of the missile (kg)
616	143.0	Original steel pipe	22.7 kg
629	136.0	Al pipe	39.3 kg

Table 1 Specification table of tests used in numerical simulation

4. PRELIMINARY RESULTS

In order to approximately simulate the experimental missile impact tests, the straight pipe model of the length of 1.5 m was developed. Two materials were considered, namely, steel and aluminum. For the steel pipe wall thickness the value of 2 millimeters was chosen in order to simulate the test #616 in the Table 1. For the aluminum pipe wall thickness the value of 5 millimeters was chosen in order to simulate the test #629. The aluminum material used in the test #629 was EN AW 6060 T66 and the minimum yield strength of the material was 160 megapascals and the tensile strength was 215 megapascals. No material tests were performed for the material before the impact test #629.

5. NUMERICAL SIMULATION OF TEST #616 WITH THE AID OF MSC/DYTRAN AND ABAQUS/ EXPLICIT PROGRAMS

According to the Table 1, the basic data for test #616 is as follows: The mass of the missile was 22.7 kilograms and the impact velocity was 143 meters per second. The missile type was plain steel pipe and test was conducted on 17th of June, 2005.
To simulate the test numerically the finite element pipe model was constructed. The model had 3977 nodal points and 3920 four node shell elements. The number of degrees of freedom was about 15000. The von Mises material properties for the steel were adopted with the following parameter values: Young's modulus was selected to 206000 megapascals and the Poisson ratio was selected to be 0.3. For the yield strength of the steel material the value of 200 megapascals was chosen. The mass density of the material was selected to be 0.00785 gigagrams per cubic meter. The megapascals, meters, seconds and gigagrams form a consistent unit system for dynamic analysis of the impact phenomenon. The plot of the finite element model of the impacting steel pipe is given in Figure 5.

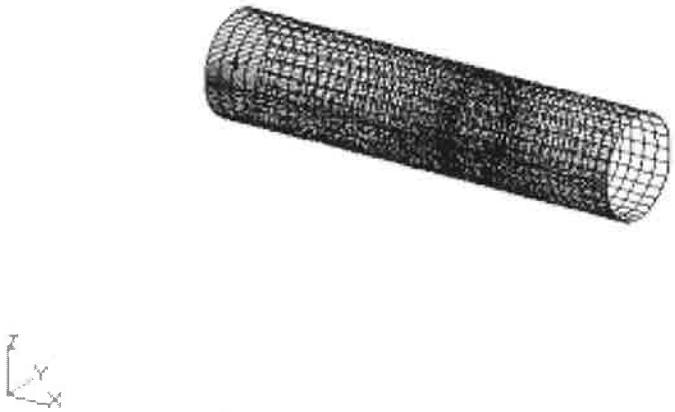


Figure 5 The finite element model of the impacting steel pipe

The following results of the simulation analysis were plotted for qualitative comparison with the measured results from the pipe missile impact test. The reaction force time history in the impact direction from one point of the impacting front edge of the pipe was plotted. Also the displacement and velocity time histories in the impact direction of one point in the rear edge of the pipe were plotted. Finally the deformed shape of the pipe at end of the impact on time 33 milliseconds from the beginning. The time-histories and the deformed shape of the steel pipe at the end of the impact simulated by MSC/DYTRAN are presented in the Figures 6, 7, 8 and 9.

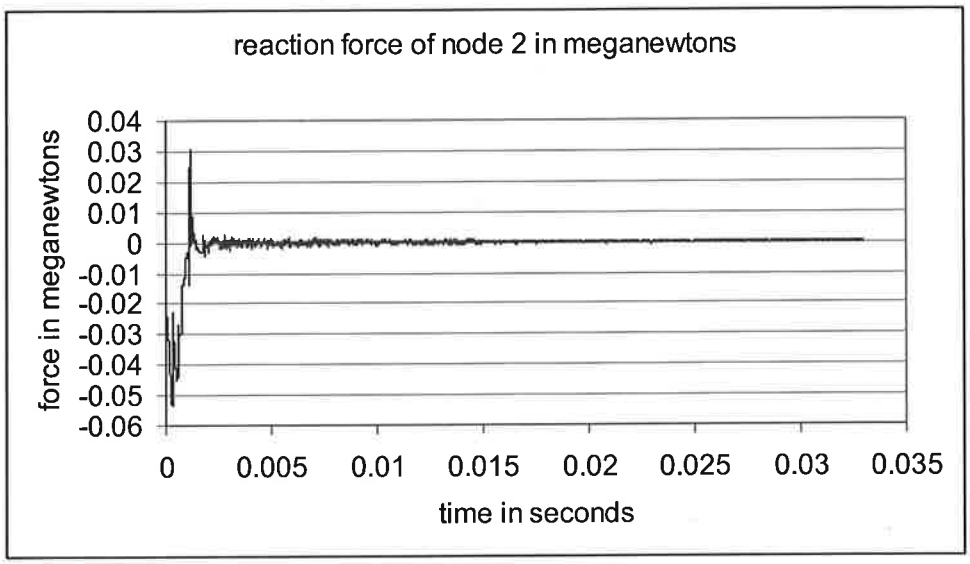


Figure 6 The time history of the reaction force in megapascal of the node 2 in the impacting edge of the steel pipe missile

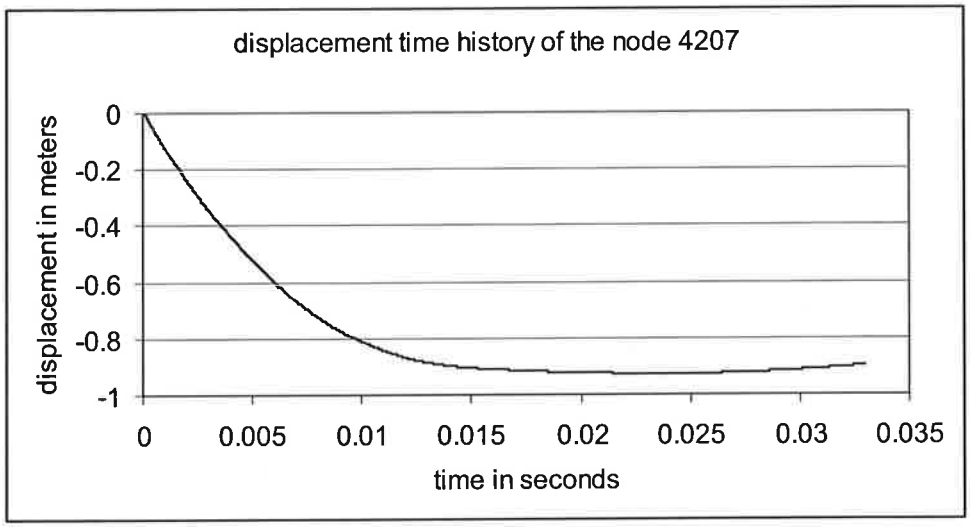


Figure 7 The time history of the displacement in the impacting direction in meters of the node 4207 in the rear edge of the steel pipe missile

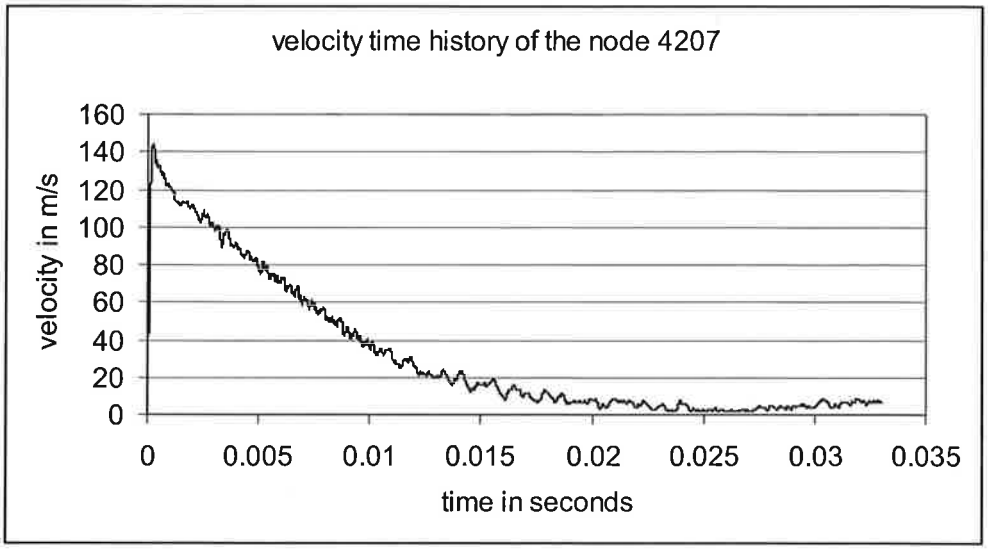


Figure 8 The time history of the velocity in the impacting direction in meters per second of the node 4207 in the rear edge of the steel pipe missile

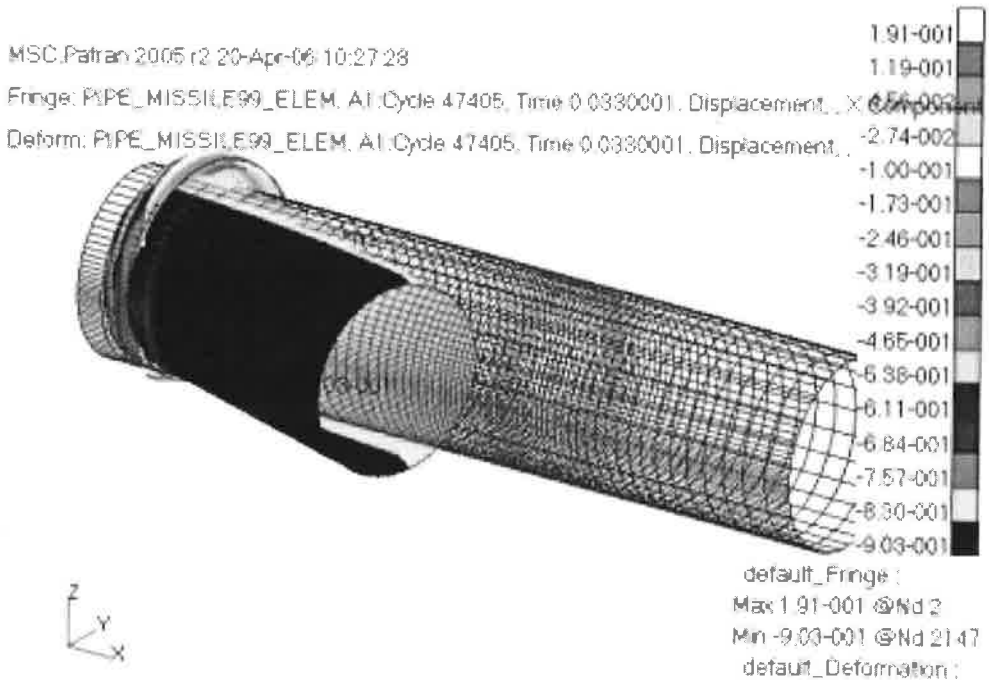


Figure 9 The deformed shape of the steel pipe missile at the end of the impact

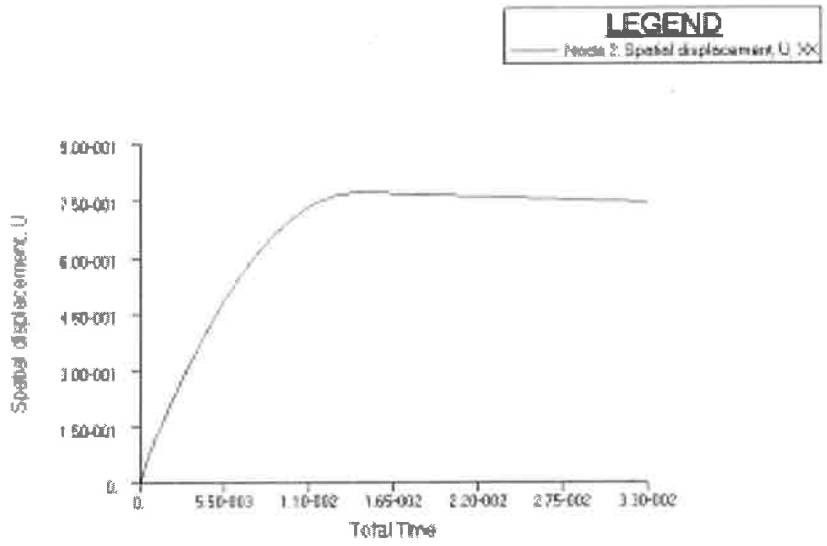


Figure 10 The time history of the displacement in the impacting direction in meters of the node 2 in the rear edge of the steel pipe missile calculated by ABAQUS/EXPLICIT

Figure 11 The time history of the velocity in the impacting direction in meters per second of the node 2 in the rear edge of the steel pipe missile calculated by ABAQUS/EXPLICIT

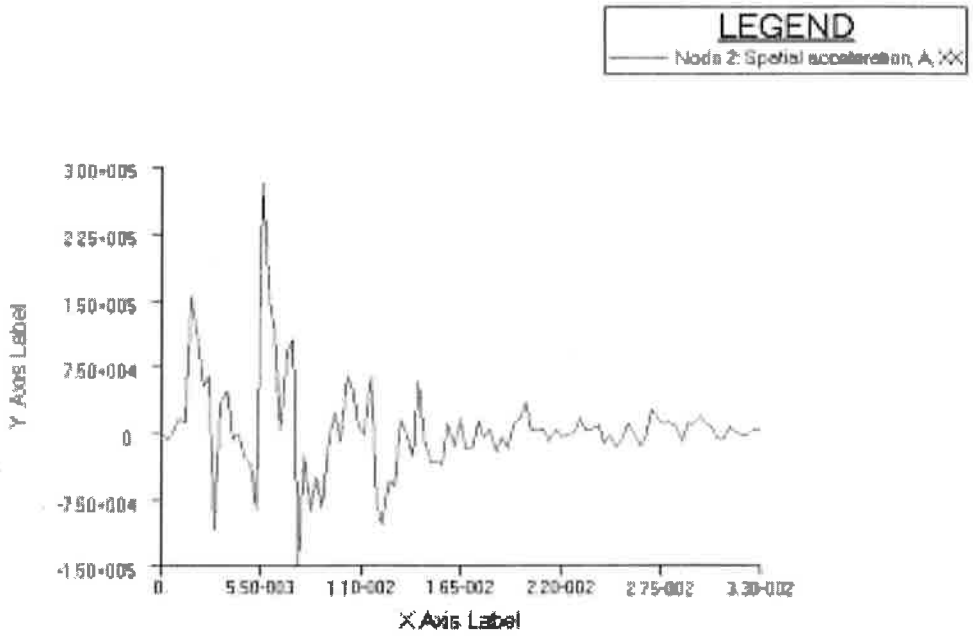


Figure 12 The time history of the acceleration in the impacting direction in meters per second square of the node 2 in the rear edge of the steel pipe missile calculated by ABAQUS/EXPLICIT

The reaction force measured as an average value of the sensors FT1, FT2 and FT3 shown in Figure 3 are given in the Figure 13.

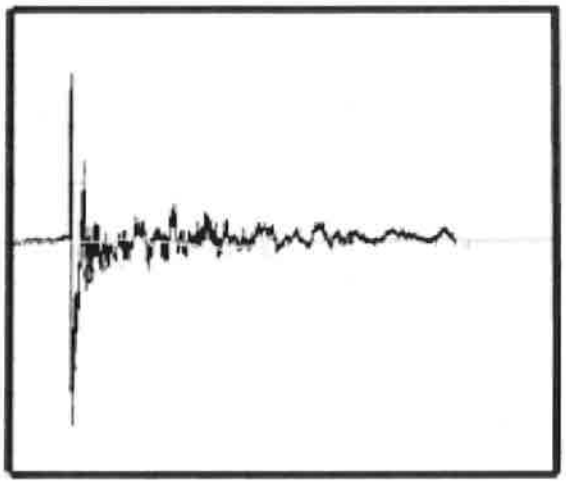


Figure 13 The measured reaction force of the test #616

6. NUMERICAL SIMULATION OF THE TEST #629 WITH THE AID OF MSC/DYTRAN

According to the Table 1 the basic data for test #629 is as follows: The mass of the missile was 39.3 kilograms and the impact velocity was 136 meters per second. The missile type was plain aluminum pipe and test was conducted on 16th of December, 2005.
To simulate the test numerically the finite element pipe model already used for the steel missile was utilized. The model had the same 3977 nodal points and 3920 four node shell elements. The number of degrees of fredom was about 15000. The von Mises material properties for the aluminum were adopted with the following parameter values: Young's modulus was selected to 72000 megapascals and the Poisson ratio was selected to be 0.01. For the yield strength of the steel material the value of 160 megapascals was chosen. The mass density of the material was selected to be 0.00270 gigagrams per cubic meter. The megapascals, meters, seconds and gigagrams form consiststent unit system for dynamic analysis of the impact phenomenon. The rear end displacement ad velocity histories of the impacting aluminum pipe missile are given in Figures 14 and 15. The reaction force time history of one node of the impacting front edge of the missile is given in Figure 16. The reaction force measured as an average og the recordings of sensors FT1, FT2 and FT3 is given in Figure 17.

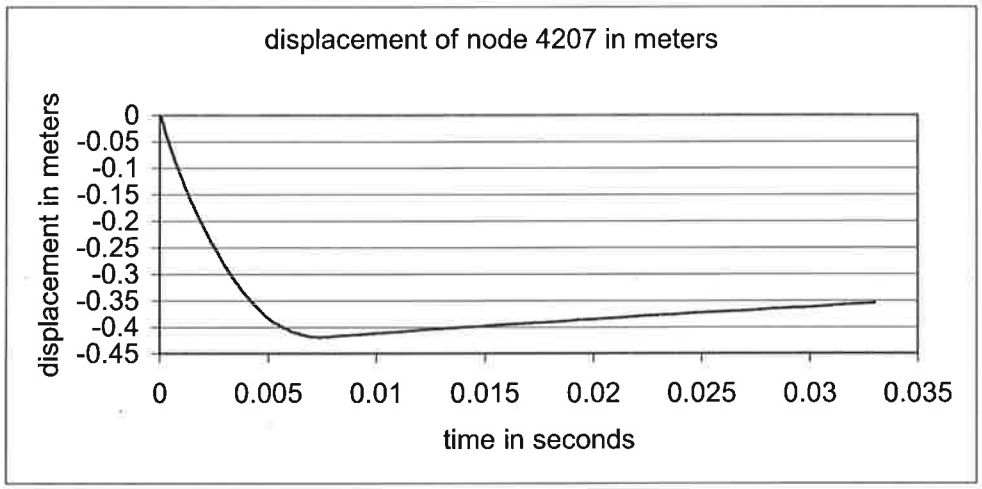


Figure 14 The time history of the displacement in the impacting direction in meters of the node 4207 in the rear edge of the aluminum pipe missile

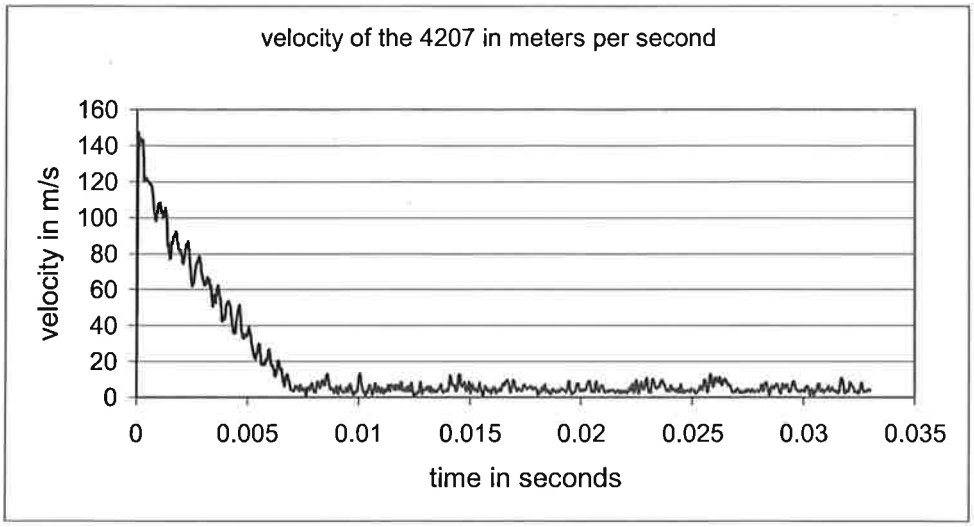


Figure 15 The time history of the velocity in the impacting direction in meters of the node 4207 in the rear edge of the aluminum pipe missile

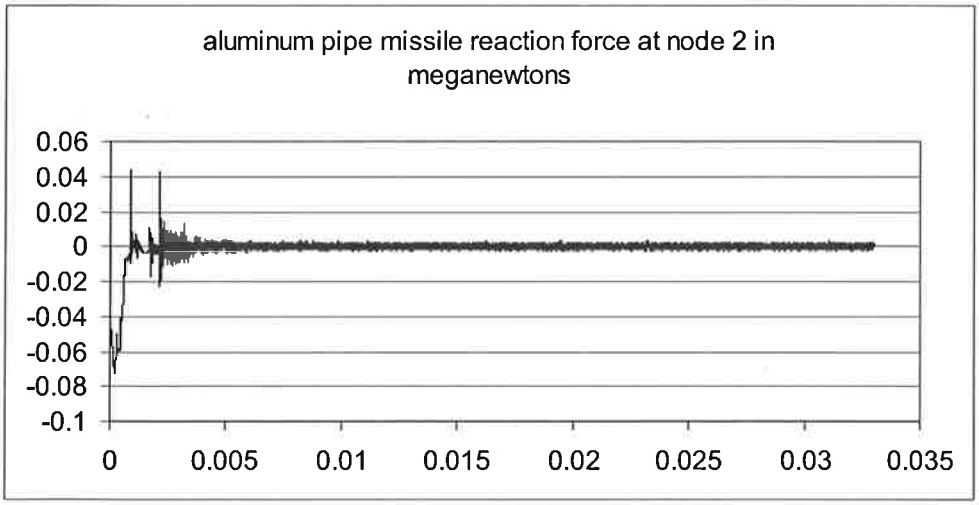


Figure 16 The time history of the reaction force in megapascals of the node 2 in the impacting edge of the aluminum pipe missile

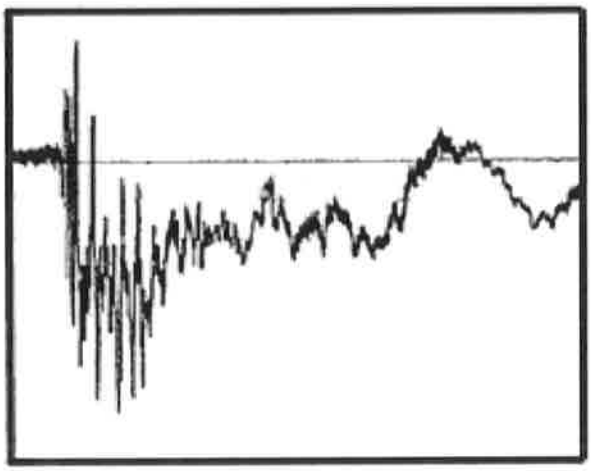


Figure 17 The measured reaction force of the test #629

7. CONCLUSION

In this paper the outline of the SAFIR experimental program for investigating the effects of soft missile impact on a rigid target was reported. During the year 2005 about 20 tests were carried out. Two different materials were utilized in the for soft missiles, namely, steel and aluminum. Part of the were conducted with missiles filled with liquid.

Preliminary numerical simulation results for one steel missile test as well as for one aluminum missile test were reported in this paper. The simulations were carried out with two different analysis programs: MSC/DYTRAN and ABAQUS/EXPLICIT. The obtained numerical simulation results agree fairly with each other.

8. REFERENCES

Riera, J. On the stress analysis of structures subjected to Aicraft impact forces. Nuclear Engineering and Design, 8, 1968, 41.5-426.

Johnson, W., Impact Strength of Materials. Edward Arnold Ltd, London 1972, 361 p.

Hakola, Ilkka, Kärnä, Tuomo: IMPACT TESTs Testing Facility. SAFIR Report December 2005

STRESSES OF INSULATING GLASS UNITS

J. AALTO, P. HASSINEN, I. KOTSIONIS

Helsinki University of Technology
Structural Mechanics
P.O. Box 2100, 02015 TKK
e-mail: jukka.aalto@tkk.fi

ABSTRACT

The paper demonstrates physical behaviour of multiply insulating glass units. It also introduces a practical technique to solve stresses and deflections of glass units without expensive multipurpose finite element codes. Only programs, which are able to solve small and large deflection plates, is needed. The paper also demonstrates the presented technique with numerical examples, which are carried out in detail. IGUs of three panes, which are in common use in Finland, are considered in these examples.

1. INTRODUCTION

Insulating glass units (IGU) consist typically of two or three glass plates, which have been fixed air-tight together through spacer profiles. The cavities between these glass panes are filled with a mixture of gases and closed in the manufacturing process. The mixture may consist of dry air and of argon or krypton gases to improve the thermal insulation power of the unit. The pressure in the closed cavity changes due to the change of the temperature. Further, the changes of the outside air pressure due to the barometric pressure changes and the change of the height position of the unit in geographical map and due to the wind speed pressure, result in pressure differences between the closed cavity and outside air. The sum of the differences of the pressure between the closed cavity and outside air make them to deflect in the direction of a lower pressure. The deflections of the panes cause a change in the volume of the cavity, which have effects in the pressure inside of the closed cavity.

The length and width of the glass plates are usually large compared to the thickness of the plates. For this reason, deflections of the plates increase easily over the thickness of the plates, which is a practical limit for the validity of the small deflection theory. Large deflections and membrane stresses have to be taken into account in the analysis and design calculations of glass plates.

In reference [1] physical behaviour of IGUs of two panes is demonstrated and practical techniques to analyse such structures are outlined. This paper is a generalization of those ideas to IGUs of more than two panes. It also presents an efficient technique of large deflection analysis of IGUs.

2. INSULATING GLASS UNITS

2.1 Pressure-volume chance relation of a gas-filled cavity

In a gas filled cavity $\boxed{i}$ of an IGU the state equation of an ideal gas is assumed to hold. Thus

$$\frac{p_i V_i}{T_i} = \frac{p_P V_{Pi}}{T_P}, \tag{1}$$

where p_i is the pressure, V_i is the volume and T_i is the absolute temperature of the cavity and subscripts P refers to the state of production. Here a rectangle around i in text and subscript i in a symbol refer to cavity $\boxed{i}$. The volume V_i of the cavity $\boxed{i}$ is thus

$$V_i = \frac{T_i}{T_P}\frac{p_P}{p_i}V_{Pi} = \frac{T_i}{T_P}\frac{p_P}{p_B + \Delta p_i}V_{Pi}, \tag{2}$$

where Δp_i the pressure change with respect to the barometric pressure p_B of the surrounding air.

2.2 Volume change caused by deflection of a glass pane

When a glass pane deflects, the volumes of the cavities around it change. This volume change caused by the deflection of pane i is

$$\Delta V^i = \int_A w^i(x,y)dA, \tag{3}$$

where $w^i(x,y)$ is the deflection of the pane and A is the area of the plate. Here superscript i in a symbol refers to glass pane i. It is practical to divide the loading of the pane in two parts, uniform load q^i and non-uniform load. The uniform load consists of the pressure differences between the pane surfaces and the non-uniform load of possible non-uniformly distributed surface loads, line loads and concentrated loads of the pane. The volume change is here expressed as a function of the uniform load q^i as

$$\Delta V^i = \Delta V^i(q^i). \tag{4}$$

If the pane can be assumed to follow a small deflection theory, the volume change is written as

$$\Delta V^i = v^i_{uni} q^i + \Delta V^i_{non}. \tag{5}$$

The quantity v^i_{uni} is the volume change caused by unit (uniform) load. The volume change v^i_{uni} can be obtained by first analysing the plate under uniform load with unit intensity resulting to the deflection $w^i_{uni}(x,y)$ and then using equation (3). The volume change ΔV^i_{non} can be obtained by analysing the plate under the (known) non-uniform load resulting to the corresponding deflection $w^i_{non}(x,y)$ and then using equation (3). This two step analysis can be performed in advance

analytically using Fourier series solutions or numerically using the finite element method. The series solutions apply only to simple geometries and restricted boundary conditions.

If large deflection plate behaviour is assumed, expression (5) is no more valid. It is, however, possible to construct in advance an approximation to the relation (4) with the help of a standard large deflection finite element plate analysis program.

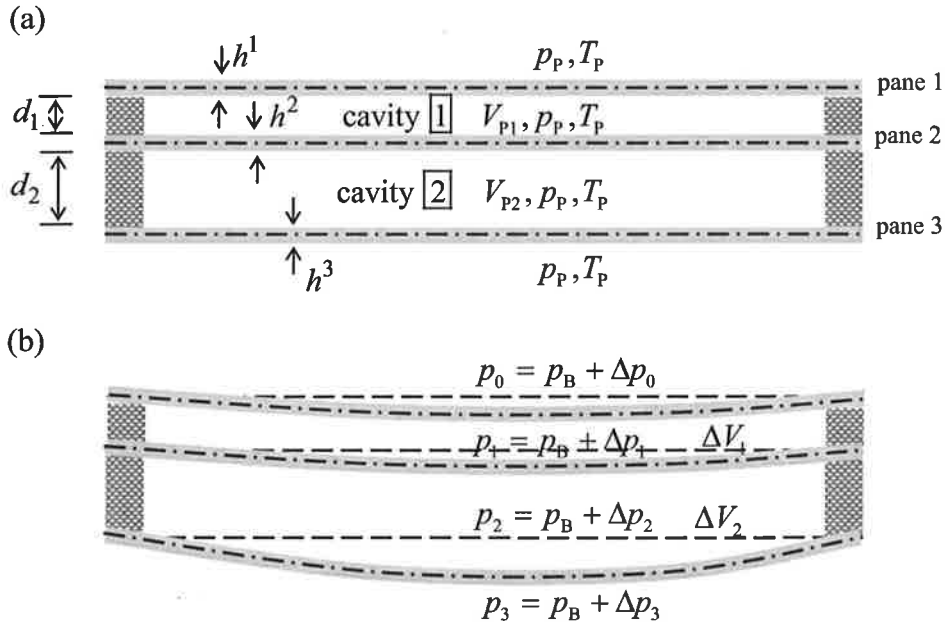


Fig. 1: Cross-section of an IGU of three panes: (a) initial and (b) deformed.

2.3 IGU of *n* panes

Consider an IGU of n panes and $n-1$ cavities (see Fig. 1, where $n=3$). The volume V_i of cavity $\boxed{i}$ between panes i and $i+1$ is thus

$$\Delta V_i = V_{Pi} - \Delta V^i + \Delta V^{i+1}, \tag{6}$$

where ΔV^i and ΔV^{i+1} are the volume changes caused by panes i and $i+1$, respectively.

In large deflection case it can be written as

$$V_i = V_{Pi} - \Delta V^i(q^i) + \Delta V^{i+1}(q^{i+1}), \tag{7}$$

where

$$q^i = \Delta p_{i-1} - \Delta p_i, \; q^{i+1} = \Delta p_i - \Delta p_{i+1} \tag{8}$$

Δp_{i-1}, Δp_i and Δp_{i+1} are the pressure changes in cavities $\boxed{i-1}$, $\boxed{i}$ and $\boxed{i+1}$, respectively. Thus the volume of the cavity $\boxed{i}$ is

$$V_i = V_{Pi} - \Delta V^i(\Delta p_{i-1} - \Delta p_i) + \Delta V^{i+1}(\Delta p_i - \Delta p_{i+1}) \tag{9}$$

being a nonlinear function of the pressure changes Δp_{i-1}, Δp_i and Δp_{i+1}. Equating the right-hand-sides of equations (2) and (9) gives the following non-linear equation

$$f_i(\Delta p_{i-1}, \Delta p_i, \Delta p_{i+1}) = 0, \tag{10}$$

where

$$f_i = (1 + \frac{\Delta p_i}{p_B})[-\Delta V^i(\Delta p_{i-1} - \Delta p_i) + \Delta V^{i+1}(\Delta p_i - \Delta p_{i+1})] + \frac{V_{Pi}}{p_B}(\Delta p_i - \Delta p_{0i}), \tag{11}$$

between the pressure changes of cavities $\boxed{i-1}$, $\boxed{i}$ and $\boxed{i+1}$. In equation (11) the isochoric pressure of cavity i

$$\Delta p_{0i} = \frac{T_i}{T_P} p_P - p_B \tag{12}$$

has been adopted.

In small deflection case the volume change ΔV_i of cavity $\boxed{i}$ can be written as

$$\begin{aligned} V_i &= V_{Pi} - v^i_{uni} q^i - \Delta V^i_{non} + v^{i+1}_{uni} q^{i+1} + \Delta V^{i+1}_{non} \\ &= V_{Pi} - v^i_{uni}\Delta p_{i-1} + (v^i_{uni} + v^{i+1}_{uni})\Delta p_i - v^{i+1}_{uni}\Delta p_{i+1} - \Delta V^i_{non} + \Delta V^{i+1}_{non} \end{aligned} \tag{13}$$

and it is a linear function of the pressure changes Δp_{i-1}, Δp_i and Δp_{i+1}. Equating the right-hand-sides of equations (2) and (13) gives the following quadratic equation

$$(1 + \frac{\Delta p_i}{p_B})[-v^i_{uni}\Delta p_{i-1} + (v^i_{uni} + v^{i+1}_{uni})\Delta p_i - v^{i+1}_{uni}\Delta p_{i+1} - \Delta V^i_{non} + \Delta V^{i+1}_{non}] + \frac{V_{Pi}}{p_B}(\Delta p_i - \Delta p_{0i}) = 0, \tag{14}$$

between the pressure changes of cavities $\boxed{i-1}$, $\boxed{i}$ and $\boxed{i+1}$. In practical applications the pressure change Δp_i is small compared to the barometric pressure p_B. If we can assume that the ratio $\Delta p_i / p_B$ is so small compared to unity, that it can be cancelled, equation (14) can be written as

$$-v^i_{uni}\Delta p_{i-1} + (v^i_{uni} + v^{i+1}_{uni} + \frac{V_{Pi} - \Delta V^i_{non} + \Delta V^{i+1}_{non}}{p_B})\Delta p_i - v^{i+1}_{uni}\Delta p_{i+1} = \frac{V_{Pi}}{p_B}\Delta p_{0i} + \Delta V^i_{non} - \Delta V^{i+1}_{non}. \tag{15}$$

This equation is a linear equation of the pressure changes Δp_{i-1}, Δp_i and Δp_{i+1}.

Equations (10), (14) or (15) can be written for each cavity and they form a set of $n-1$ equations. Unknowns in these equations are the pressure changes $\Delta p_1,\cdots,\Delta p_{n-1}$ of the cavities, because the pressure changes Δp_0 and Δp_n on the outer faces of the IGU are known. If $\overline{q}_1$ and $\overline{q}_n$ denote uniform loads on the first and last pane, respectively, we have $\Delta p_0 = \overline{q}_1$ and $\Delta p_n = -\overline{q}_n$. In practice IGUs consist of two or thee panes. Corresponding set of equations has only one or two equations and thus, can be solved very effectively.

2.4 Solving the cavity pressure changes

An IGU of three panes is considered here as an example. In large deflection case the corresponding set of two equations of type (10) can be written as

$$\mathbf{f}(\Delta\mathbf{p}) = \mathbf{0}, \tag{16}$$

where

$$\mathbf{f} = \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix},\ \Delta\mathbf{p} = \begin{Bmatrix} \Delta p_1 \\ \Delta p_2 \end{Bmatrix} \tag{17}$$

and

$$\begin{aligned} f_1(\Delta p_1,\Delta p_2) &= (1+\frac{\Delta p_1}{p_B})[-\Delta V^1(\overline{q}_1-\Delta p_1)+\Delta V^2(\Delta p_1-\Delta p_2)]-\frac{V_{P1}}{p_B}(\Delta p_1-\Delta p_{01}),\\ f_2(\Delta p_1,\Delta p_2) &= (1+\frac{\Delta p_2}{p_B})[-\Delta V^2(\Delta p_1-\Delta p_2)+\Delta V^3(\Delta p_2+\overline{q}_3)]-\frac{V_{P2}}{p_B}(\Delta p_2-\Delta p_{02}). \end{aligned} \tag{18}$$

The solution of this equation is obtained by using the Newton-Raphson iteration

$$\Delta\mathbf{p}^{k+1} = \Delta\mathbf{p}^k - (\mathbf{S}_T^k)^{-1}\mathbf{f}^k \tag{19}$$

starting with $\Delta\mathbf{p}^0 = \mathbf{0}$. Here the superscript k refers to an iteration cycle and

$$\mathbf{S}_T^k = \frac{\partial\mathbf{f}}{\partial\Delta\mathbf{p}}(\Delta\mathbf{p}^k) \tag{20}$$

is the tangential matrix. The elements of this matrix can be written as

$$\begin{aligned} S_{11}^T &\equiv \frac{\partial f_1}{\partial\Delta p_1} = (1+\frac{\Delta p_1}{p_B})[\frac{d\Delta V^1}{dq^1}(q^1)+\frac{d\Delta V^2}{dq^2}(q^2)]+\frac{V_{P1}-\Delta V^1(q^1)+\Delta V^2(q^2)}{p_B},\\ S_{12}^T &\equiv \frac{\partial f_1}{\partial\Delta p_2} = -(1+\frac{\Delta p_1}{p_B})\frac{d\Delta V^2}{dq^2}(q^2),\\ S_{21}^T &\equiv \frac{\partial f_2}{\partial\Delta p_1} = -(1+\frac{\Delta p_2}{p_B})\frac{d\Delta V^2}{dq^2}(q^2),\\ S_{22}^T &\equiv \frac{\partial f_2}{\partial\Delta p_2} = (1+\frac{\Delta p_2}{p_B})[\frac{d\Delta V^2}{dq^2}(q^2)+\frac{d\Delta V^3}{dq^3}(q^3)]+\frac{V_{P2}-\Delta V^2(q^2)+\Delta V^3(q^3)}{p_B}, \end{aligned} \tag{21}$$

where $q^1 = \overline{q}_1 - \Delta p_1$, $q^2 = \Delta p_1 - \Delta p_2$ $q^3 = \Delta p_2 + \overline{q}_3$.

In small deflection case the corresponding set of two equations of the type (14) can be written as

$$\mathbf{S}(\Delta\mathbf{p})\Delta\mathbf{p} = \mathbf{t}, \tag{22}$$

where the elements of matrix $\mathbf{S}(\Delta\mathbf{p})$ and column vector $\mathbf{t}$ are

$$\begin{aligned}
&S_{11} = (1+\frac{\Delta p_1}{p_B})(v^1_{uni} + v^2_{uni}) + \frac{V_{P1} - \Delta V^1_{non} + \Delta V^2_{non}}{p_B}, \quad S_{12} = -(1+\frac{\Delta p_1}{p_B})v^2_{uni}, \\
&S_{21} = -(1+\frac{\Delta p_2}{p_B})v^2_{uni}, \quad S_{22} = (1+\frac{\Delta p_2}{p_B})(v^2_{uni} + v^3_{uni}) + \frac{V_{P2} - \Delta V^2_{non} + \Delta V^3_{non}}{p_B}, \\
&t_1 = v^1_{uni}\overline{q}_1 + \Delta V^1_{non} - \Delta V^2_{non} + \frac{\Delta p_{01}}{p_B}V_{P1}, \quad t_2 = -v^2_{uni}\overline{q}_3 + \Delta V^2_{non} - \Delta V^3_{non} + \frac{\Delta p_{02}}{p_B}V_{P2}.
\end{aligned} \tag{23}$$

The solution of this equation is obtained by using the Picard iteration

$$\Delta\mathbf{p}^{k+1} = \mathbf{S}(\Delta\mathbf{p}^k)^{-1}\mathbf{t}, \tag{24}$$

starting with $\Delta\mathbf{p}^0 = \mathbf{0}$. The corresponding solution of the linearized equation (22) is simply

$$\Delta\mathbf{p} = \mathbf{S}^{-1}\mathbf{t}, \tag{25}$$

where the elements of matrix $\mathbf{S}$ are

$$\begin{aligned}
&S_{11} = v^1_{uni} + v^2_{uni} + \frac{V_{P1} - \Delta V^1_{non} + \Delta V^2_{non}}{p_B}, \; S_{12} = -v^2_{uni}, \\
&S_{21} = -v^2_{uni}, \; S_{22} = v^2_{uni} + v^3_{uni} + \frac{V_{P2} - \Delta V^2_{non} + \Delta V^3_{non}}{p_B}.
\end{aligned} \tag{26}$$

It is easy to note, that first iteration cycle of the Picard iteration gives the solution of the linearized problem. The small deflection solution of the three pane IGU problem can thus be obtained with a minimum effort, even by hand.

2.5 Calculating the volume change from the results of a finite element plate analysis

The finite element approximation of the volume change (3) is simply

$$\Delta\hat{V} = \sum_e \int_{A^e} \hat{w}^e(x,y)dA, \tag{27}$$

where A^e is the element area and the summation is performed over all the elements of the plate. The element approximation of the deflection is

$$\hat{w}^e(x,y) = \sum_{i=1}^{m} N^e_i(x,y)a^e_i, \tag{28}$$

where $N_i^e(x,y)$ are the corresponding element shape functions, a_i^e are the element nodal parameters and m is the number of element degrees of freedom. Integration in each element in equation (27) is usually performed numerically. In the examples of this paper a nine degree of freedom discrete Kirchhoff triangular (DKT) plate element [2] has been used, which is one of the most widely used plate elements. The degrees of freedom of this element are the deflection and two rotations at each node. The equations of this element have traditionally been derived by defining the deflection only on the element sides. Therefore the approximation of type (28) is not available. This problem can, however, be easily avoided by using the cubic shape functions of the non-conforming triangular plate element of reference [3], which has identical nodal parameters. Numerical integration with three integration points [4] suffices to integrate the corresponding deflection approximation $\hat{w}^e(x,y)$ exactly.

2.6 Expressing the volume change as a function of the uniform load

The relation (4) for pane i is obtained from large deflection analysis of the pane. The superscript i referring to a pane number is here omitted. Finite element analysis of the pane is performed with equally spaced values q_j, $j=1,\cdots,m$ of the uniform load q. Values of the corresponding volume changes ΔV_j can thus be calculated. In the vicinity of each point q_j the function $\Delta V(q)$ is here approximated using a second degree Lagrange polynomial. This approximation can be expressed as

$$\Delta V(q)=\frac{1}{\Delta q^2}[\frac{(q-q_j)(q-q_{j+1})}{2}\Delta V_{j-1}-(q-q_{j-1})(q-q_{j+1})\Delta V_j+\frac{(q-q_{j-1})(q-q_j)}{2}\Delta V_{j+1}], \quad (29)$$

where Δq is the uniform load interval. In the Newton-Raphson iteration the corresponding derivative is also needed

$$\frac{d\Delta V}{dq}(q)=\frac{1}{\Delta q^2}[\frac{-q_j+2q-q_{j+1}}{2}\Delta V_{j-1}-(-q_{j-1}+2q-q_{j+1})\Delta V_j+\frac{-q_{j-1}+2q-q_j}{2}\Delta V_{j+1}]. \quad (30)$$

The piecewise approximation within the range $q_1 \le q \le q_n$ is defined as follows: In the first interval $q_1 \le q \le (q_2+q_3)/2$ equations (29) and (30) with $j=2$ are used, in an intermediate interva $(q_{j-1}+q_j)/2 \le q \le (q_j+q_{j+1})/2$ ($j=3,\cdots,m-2$) these equations with j are used and in the last interval $(q_{m-2}+q_{m-1})/2 \le q \le q_m$ these equations with $j=m-1$ are used. The derivative of this approximation is not continuous at the midpoints $q=(q_{j-1}+q_j)/2$ of the intermediate intervals ($j=3,\ldots,m-2$). Therefore, if the value of the derivative exactly at such point is needed, it can be calculated by averaging the two existing values.

2.7 Calculating pane deflections and stress resultants

After the cavity pressure changes Δp_i have been solved, the uniform loads on each pane can be calculated simply by using

$$q^i=\Delta p_{i-1}-\Delta p_i,\ i=1,\cdots,n. \quad (31)$$

Consequently the loading of each pane is known and they can be analysed separately with the help of handbook formulas, series solutions or the finite element method.

3. EXAMPLES

The presented technique is demonstrated using three examples. In the first two examples the panes are assumed to follow small deflection plate theory and in the last example large deflection theory. In the first example the IGU is loaded by a temperature change in the cavities and in the last two examples by a uniform line load on the first pane.

Example 1:

Consider a simply supported rectangular IGU of three panes, with the side the lengths of $350\text{mm} \times 500\text{mm}$. The thickness of all the glass panes and cavities are 4mm and 12mm respectively. The temperature and the pressure at the time of production were $20°\text{C}$ and 100kPa, respectively. The temperature of the cavities rises up to $58°\text{C}$, but the pressure does not change. Obtain the pressure change in the cavities, maximum deflection and maximum normal stress of the panes. Assume small deflection plate model.

The initial volume of the cavities:

$$V_{\text{P}i} = (0{,}350 \cdot 0{,}500 \cdot 0{,}012)\text{m}^3 = 2{,}1 \cdot 10^{-3}\,\text{m}^3.$$

Because the panes are not subjected to no-uniform load

$$\Delta V^i_{\text{non}} = 0, \quad i = 1,2,3.$$

Using double Fourier series solution [5] of a simply supported rectangular plate we obtain

$$v^1_{\text{uni}} = v^2_{\text{uni}} = v^3_{\text{uni}} = 2{,}0583 \cdot 10^{-8}\,\frac{\text{m}^3}{\text{kPa}}.$$

The isochoric pressure of the two cavities:

$$\Delta p_{0i} = \frac{T_i}{T_{\text{P}}}\,p_{\text{P}} - p_{\text{B}} = \left(\frac{273+58}{273+20}100 - 100\right)\text{kPa} = 12{,}97\text{kPa}, \quad i = 1,2.$$

The elements of matrix $\mathbf{S}$ and column vector $\mathbf{t}$:

$$S_{11} = \left(1+\frac{\Delta p_1}{p_{\text{B}}}\right)(v^1_{\text{uni}} + v^2_{\text{uni}}) + \frac{V_{\text{P1}}}{p_{\text{B}}} = \left[\left(1+\frac{\Delta p_1}{p_{\text{B}}}\right)4{,}1166 + 2{,}1\right]\cdot 10^{-5}\,\frac{\text{m}^3}{\text{kPa}},$$

$$S_{12} = -\left(1+\frac{\Delta p_1}{p_{\text{B}}}\right)v^2_{\text{uni}} = -\left(1+\frac{\Delta p_1}{p_{\text{B}}}\right)2{,}0583 \cdot 10^{-5}\,\frac{\text{m}^3}{\text{kPa}}, \quad S_{21} = -\left(1+\frac{\Delta p_2}{p_{\text{B}}}\right)v^2_{\text{uni}} - \left(1+\frac{\Delta p_2}{p_{\text{B}}}\right)2{,}0583 \cdot 10^{-5}\,\frac{\text{m}^3}{\text{kPa}},$$

$$S_{22} = \left(1+\frac{\Delta p_2}{p_{\text{B}}}\right)(v^2_{\text{uni}} + v^3_{\text{uni}}) + \frac{V_{\text{P2}}}{p_{\text{B}}} = \left[\left(1+\frac{\Delta p_1}{p_{\text{B}}}\right)4{,}1166 + 2{,}1\right]\cdot 10^{-5}\,\frac{\text{m}^3}{\text{kPa}},$$

$$t_1 = \frac{\Delta p_{01}}{p_B} V_{P1} = 0,2724 \cdot 10^{-3}\,\mathrm{m}^3, \quad t_2 = \frac{\Delta p_{02}}{p_B} V_{P2} = 0,2724 \cdot 10^{-3}\,\mathrm{m}^3.$$

Iteration:

$$\Delta\mathbf{p}^0 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \quad \mathbf{t} = 0,2724 \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \cdot 10^{-3}\,\mathrm{m}^3.$$

Iteration cycle 1:

$$\mathbf{S}(\Delta\mathbf{p}^0) = \begin{bmatrix} 6,2166 & -2,0583 \\ -2,0583 & 6,2166 \end{bmatrix} \cdot 10^{-5} \frac{\mathrm{m}^3}{\mathrm{kPa}}, \quad \Delta\mathbf{p}^1 = \mathbf{S}(\Delta\mathbf{p}^0)^{-1}\mathbf{t} = \begin{Bmatrix} 6,551 \\ 6,551 \end{Bmatrix} \mathrm{kPa}$$

Iteration cycle 2:

$$\mathbf{S}(\Delta\mathbf{p}^1) = \begin{bmatrix} 6,4863 & -2,1931 \\ -2,1931 & 6,4863 \end{bmatrix} \cdot 10^{-5} \frac{\mathrm{m}^3}{\mathrm{kPa}}, \quad \Delta\mathbf{p}^2 = \mathbf{S}(\Delta\mathbf{p}^1)^{-1}\mathbf{t} = \begin{Bmatrix} 6,345 \\ 6,345 \end{Bmatrix} \mathrm{kPa}$$

Iteration cycle 3:

$$\mathbf{S}(\Delta\mathbf{p}^2) = \begin{bmatrix} 6,4778 & -2,1889 \\ -2,1889 & 6,4778 \end{bmatrix} \cdot 10^{-5} \frac{\mathrm{m}^3}{\mathrm{kPa}}, \quad \Delta\mathbf{p}^3 = \mathbf{S}(\Delta\mathbf{p}^2)^{-1}\mathbf{t} = \begin{Bmatrix} 6,351 \\ 6,351 \end{Bmatrix} \mathrm{kPa}$$

Iteration cycle 4:

$$\mathbf{S}(\Delta\mathbf{p}^3) = \begin{bmatrix} 6,4780 & -2,1890 \\ -2,1890 & 6,4780 \end{bmatrix} \cdot 10^{-5} \frac{\mathrm{m}^3}{\mathrm{kPa}}, \quad \Delta\mathbf{p}^4 = \mathbf{S}(\Delta\mathbf{p}^3)^{-1}\mathbf{t} = \begin{Bmatrix} 6,351 \\ 6,351 \end{Bmatrix} \mathrm{kPa}$$

Pressure changes in cavities 1 and 2:

$$\Delta p_1 = \Delta p_2 = 6,351\mathrm{kPa}\,.$$

Loading of the glass panes:

$$q^1 = \overbrace{\Delta p_0}^{0} - \Delta p_1 = -6,351\mathrm{kPa}, \quad q^2 = \Delta p_1 - \Delta p_2 = 0, \quad q^3 = \Delta p_2 - \overbrace{\Delta p_3}^{0} = 6,351\mathrm{kPa}.$$

Using double Fourier series solution of pane 1, 2 and 3 loaded by uniform loads q^1, q^2 and q^3, respectively, the corresponding deflection and normal stress distributions can be calculated. In this special case pane 2 is unloaded and panes 1 and 3 are loaded with equal, but opposite loads. Thus only the glass pane 3 is studied. Maximum deflection of the pane 3 is $w^3_{\max} = 1,8\mathrm{mm}$ and maximum normal stress is $\sigma^3_{\max} = 21,8\mathrm{MPa}$. It can be concluded, that the linearized analysis (assuming $\Delta p_i / p_B$ to vanish compared to unity) of the quadratic equations gives $\Delta p_1 = \Delta p_2 = 6,551\mathrm{kPa}$. This result is 3% higher than the accurate result and it is on the safe side. Thus the result of this example justifies the applicability of this linearizing assumption.

Example 2:

Consider a simply supported rectangular IGU of three panes, with the side lengths of $1000\text{mm}\times 3000\text{mm}$. The thickness of all the glass panes and cavities are 6mm and 12mm, respectively. The glass pane 1 is loaded by uniformly distributed straight line load $q_l = 0,5\text{kN/m}$ in the direction of the short side and 1000mm from the nearest side. Obtain the pressure change in the cavities, maximum deflection and maximum normal stress of the panes. Assume small deflection model and use simpler linearized solution for the cavity pressures.

The initial volume of the cavities:

$$V_{\text{P}i} = (3\cdot 1\cdot 0,012)\text{m}^3 = 3,6\cdot 10^{-2}\text{m}^3.$$

Using double Fourier series solution of a simply supported rectangular plate we obtain

$$v_{\text{uni}}^1 = v_{\text{uni}}^2 = v_{\text{uni}}^3 = 1,2881\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}},$$

$$\Delta V_{\text{non}}^1 = 0,27735\cdot 10^{-2}\text{m}^3, \quad \Delta V_{\text{non}}^2 = \Delta V_{\text{non}}^3 = 0.$$

The isochoric pressure of the two cavities is:

$$\Delta p_{0i} = 0, \quad i = 1,2.$$

The elements of matrix $\mathbf{S}$ and column vector $\mathbf{t}$:

$$S_{11} = v_{\text{uni}}^1 + v_{\text{uni}}^2 + \frac{V_{\text{P1}} - \Delta V_{\text{non}}^1}{p_{\text{B}}} = 2,6094\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}}, \quad S_{12} = -v_{\text{uni}}^2 = -1,2881\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}},$$

$$S_{21} = -v_{\text{uni}}^2 - 1,2881\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}}, \quad S_{22} = v_{\text{uni}}^2 + v_{\text{uni}}^3 + \frac{V_{\text{P2}}}{p_{\text{B}}} = 2,6122\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}},$$

$$t_1 = \Delta V_{\text{non}}^1 = 1,2881\cdot 10^{-2}\text{m}^3, \quad t_2 = 0.$$

Equation and solution:

$$\mathbf{S} = \begin{bmatrix} 2,6094 & -1,2881 \\ -1,2881 & 2,6122 \end{bmatrix}\cdot 10^{-2}\frac{\text{m}^3}{\text{kPa}}, \quad \mathbf{t} = \begin{Bmatrix} 0,27735 \\ 0 \end{Bmatrix} 10^{-2}\text{m}^3$$

$$\Delta\mathbf{p} = \mathbf{S}^{-1}\mathbf{t} = \frac{1}{5,1571}\begin{bmatrix} 2,6122 & 1,2881 \\ 1,2881 & 2,6094 \end{bmatrix}\begin{Bmatrix} 0,27735 \\ 0 \end{Bmatrix}\text{kPa} = \begin{Bmatrix} 0,1405 \\ 0,0693 \end{Bmatrix}\text{kPa}$$

Pressure changes in cavities 1 and 2:

$$\Delta p_1 = 0,1405\text{kPa}, \quad \Delta p_2 = 0,0693\text{kPa}.$$

Loading of the glass panes:

$$q^1 = -\Delta p_1 = -0,1405\text{kPa}, \quad q^2 = \Delta p_1 - \Delta p_2 = 0,0712\text{kPa}, \quad q^3 = \Delta p_2 = 0,0693\text{kPa}.$$

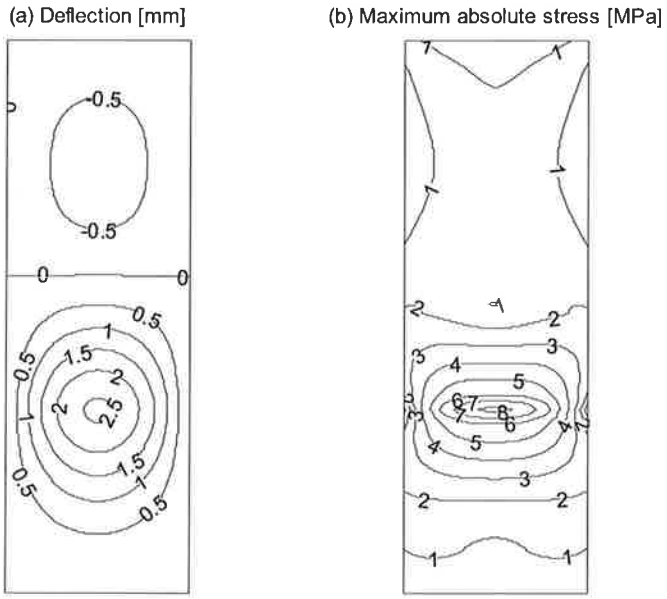


Fig. 2: Deflection and maximum absolute stress of the glass pane 1 in the example 2.

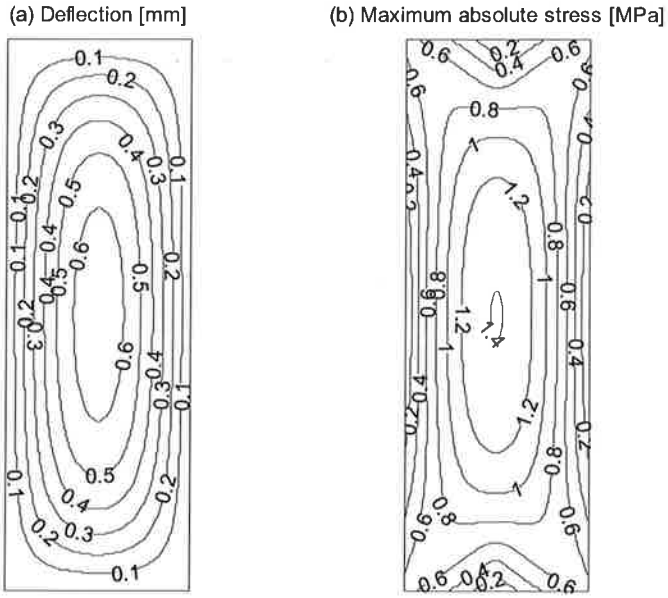


Fig. 3: Deflection and maximum absolute stress of the glass pane 2 in the example 2.

Now the loading of each glass pane is known and double Fourier series solution of each separate pane can be performed. Maximum deflection and normal stress of the pane 1 are $w^1_{max} = 2,56\text{mm}$ and $\sigma^1_{max} = 8,15\text{MPa}$, those of the pane 2 are $w^2_{max} = 0,66\text{mm}$ and $\sigma^2_{max} = 1,41\text{MPa}$, and those of

the pane 3 are $w^3_{\max} = 0,64\text{mm}$ and $\sigma^3_{\max} = 1,37\text{MPa}$, respectively. Fig. 2 and 3 show deflection and maximum absolute normal stress distributions of panes 1 and 2, respectively.

Example 3:

Consider the same rectangular IGU as in example 2. The line load is now 5 times larger $q_l = 2,5\text{kN/m}$ and large deflection plate model is used.

Large deflection finite element analysis of the pane 1 under the given line load q_l and three values of the uniform load q^1_j gives

j	q^1_j [kPa]	ΔV^1_j [m^3]
1	-0,8	0,00158392
2	-0,4	0,00593359
3	0	0,01054298

Large deflection finite element analysis of the pane 2 under three values of the uniform load q^2_i gives

j	q^2_j [kPa]	ΔV^2_j [m^3]
1	0	0
2	0,4	0,00495158
3	0,8	0,00975130

Because the panes 2 and 3 are identical, this result holds also to the pane 3.

Iteration:

$$\Delta\mathbf{p}^0 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}.$$

Iteration cycle 1:

$$\Delta V^1(q^1) = 0,0105430\text{m}^3, \quad \Delta V^2(q^2) = 0, \quad \Delta V^3(q^3) = 0,$$

$$\frac{d\Delta V^1}{dq^1}(q^1) = 0,011848\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^2}{dq^2}(q^2) = 0,012569\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^3}{dq^3}(q^3) = 0,012569\frac{\text{m}^3}{\text{kPa}},$$

$$\mathbf{f}^1 = \begin{Bmatrix} -0,0105430 \\ 0 \end{Bmatrix}\text{m}^3, \quad \mathbf{S}^1_T = \begin{bmatrix} 0,0245442 & -0,0125688 \\ -0,0125688 & 0,0253176 \end{bmatrix}\frac{\text{m}^3}{\text{kPa}},$$

$$\Delta\mathbf{p}^1 = \Delta\mathbf{p}^0 - \mathbf{S}^1_T\mathbf{f}^1 = \begin{Bmatrix} 0,57598 \\ 0,28594 \end{Bmatrix}\text{kPa}.$$

Iteration cycle 2:

$$\Delta V^1(q^1) = 0,0039880\text{m}^3, \quad \Delta V^2(q^2) = 0,0036055\text{m}^3, \quad \Delta V^3(q^3) = 0,0035551\text{m}^3,$$

$$\frac{d\Delta V^1}{dq^1}(q^1) = 0,0109132\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^2}{dq^2}(q^2) = 0,0122935\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^3}{dq^3}(q^3) = 0,0122974\frac{\text{m}^3}{\text{KPa}},$$

$$\mathbf{f}^2 = \begin{Bmatrix} -2,79917 \\ 0,001047 \end{Bmatrix} 10^{-3}\text{m}^3, \quad \mathbf{S}_T^2 = \begin{bmatrix} 0,0234516 & -0,0123289 \\ -0,0123289 & 0,0248058 \end{bmatrix} \frac{\text{m}^3}{\text{kPa}}$$

$$\Delta\mathbf{p}^2 = \Delta\mathbf{p}^1 - \mathbf{S}_T^2\mathbf{f}^2 = \begin{Bmatrix} 0,59210 \\ 0,29390 \end{Bmatrix} \text{kPa}.$$

Iteration cycle 3:

$$\Delta V^1(q^1) = 0,0038123\text{m}^3, \quad \Delta V^2(q^2) = 0,0037058\text{m}^3, \quad \Delta V^3(q^3) = 0,0036530\text{m}^3,$$

$$\frac{d\Delta V^1}{dq^1}(q^1) = 0,010887\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^2}{dq^2}(q^2) = 0,0122857\frac{\text{m}^3}{\text{kPa}}, \quad \frac{d\Delta V^3}{dq^3}(q^3) = 0,0122898\frac{\text{m}^3}{\text{KPa}},$$

$$\mathbf{f}^3 = \begin{Bmatrix} -0,220957 \\ 0,001463 \end{Bmatrix} 10^{-6}\text{m}^3, \quad \mathbf{S}_T^3 = \begin{bmatrix} 0,0234521 & -0,0123221 \\ -0,0123038 & 0,0247914 \end{bmatrix} \frac{\text{m}^3}{\text{kPa}},$$

$$\Delta\mathbf{p}^3 = \Delta\mathbf{p}^2 - \mathbf{S}_T^3\mathbf{f}^3 = \begin{Bmatrix} 0,59211 \\ 0,29391 \end{Bmatrix} \text{kPa}.$$

Loading of the glass panes:

$$q^1 = -\Delta p_1 = -0,5921\text{kPa}, \quad q^2 = \Delta p_1 - \Delta p_2 = 0,2983\text{kPa}, \quad q^3 = \Delta p_2 = 0,2939\text{kPa}.$$

Now the loading of each pane is known and large deflection finite element analysis of the panes 1, 2 and 3 can be performed. Maximum deflection and normal stress of the pane 1 are $w_{\max}^1 = 9,56\text{mm}$ and $\sigma_{\max}^1 = 35,21\text{MPa}$, those of the pane 2 are $w_{\max}^2 = 2,77\text{mm}$ and $\sigma_{\max}^2 = 5,93\text{MPa}$, and those of the pane 3 are $w_{\max}^3 = 2,69\text{mm}$ and $\sigma_{\max}^3 = 5,78\text{MPa}$, respectively. Fig. 4 and 5 show the deflection and the maximum absolute normal stress distributions of the panes 1 and 2, respectively.

4. CONCLUSIONS

The effective use of building materials requires accurate calculation models and analysis tools in order to utilise the strength in an optimal way. The paper studies behaviour of multi-pane insulating glass units (IGU), which have became popular today. The loads of the units may be external pressure, line or point load, or internal pressure load caused by the change of temperature or barometric pressure. The response of each pane has first to be studied. Secondly, the pressure loads based on the interaction of the changes of the volumes of the cavities, are calculated for each pane. In connection with small deflection plate model this can be done by solving the simple linearized equations or by solving the corresponding accurate quadratic equations iteratively starting form the linearized solution. In connection with large deflection plate model the equations are non-linear and iterative solution is always necessary. Finally, the deflections and stresses of each pane are calculated numerically using the finite element method. In connection with rectangular IGUs and small deflection plate model Fourier series solutions [5] can also be used.

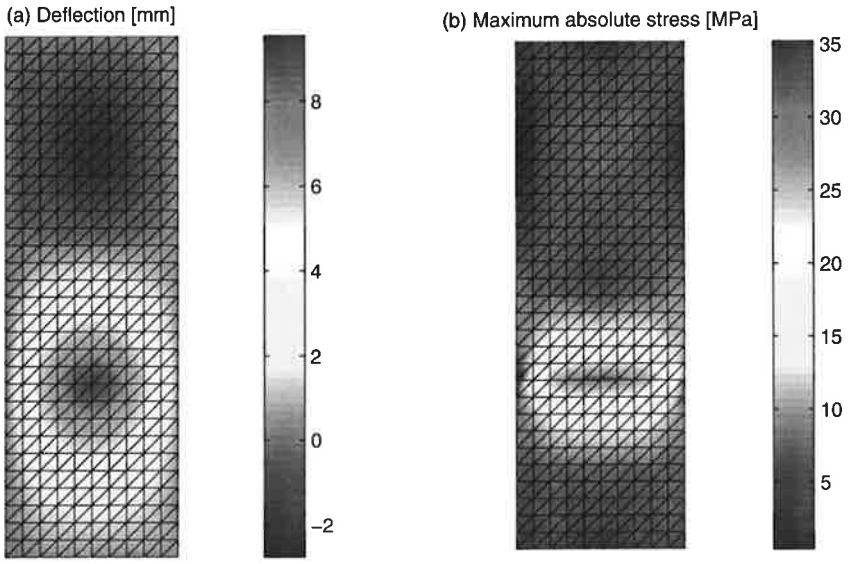


Fig. 4: Deflection and maximum absolute stress of the glass pane 1 in example 3.

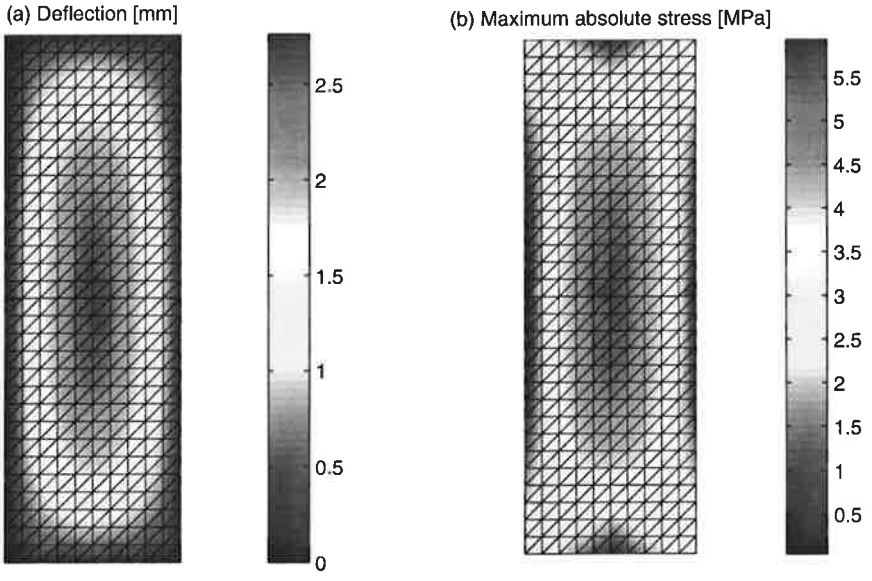


Fig. 5: Deflection and maximum absolute stress of the glass pane 2 in example 3.

REFERENCES

1. F. Feldmeier, *Insulating units exposed to wind and weather – Load shearing and internal loads.* Proceedings of Glass Processing Days, Tampere 2003, Finland, pp. 633-636.
2. J.L. Batoz, K-J. Bathe and L.W. Ho, *A study of three node plate bending elements*, Int. J. Num. Meth. Engn. **15** (1980) pp. 1771-1812.
3. G.P. Bazeley, Y.K. Cheung, B.M. Irons and O.C. Zienkiewicz, *Triangular elements in plate bending – Conforming and non-conforming solutions,* Journal of Strain Analysis, **6** (1971), pp. 547-576.
4. O.C. Zienkiewicz and R.L. Taylor, *The finite element method –Volume 1: The Basis*, Butterworth-Heinemann, 2000, p. 222.
5. S.P. Timoshenko and S. Wloinowsky-Krieger, *Theory of plates and shells*, McGraw-Hill, 1970.

ANALYSIS OF LINE LOAD DETERMINATION BY INFLUENCE COEFFICIENT METHOD

A-J. ROMPPANEN, E. KESKINEN

TTY, Konedynamiikan laboratorio
P. O. Box 589, FIN-33101 Tampere
e-mail: antti-jussi.romppanen@tut.fi

ABSTRACT

In paper manufacturing processes the paper web is typically led through several roll pair contacts. In many cases the web is pressed between two rolls to have better properties on the final product. The line load distribution in the contact zone is very important and yet very hard to measure accurately. One choice of measuring the line load is to use the influence coefficient method which utilizes strain measurements of the roll in contact.

The goal is to study this measuring method computationally on a roll structure in order to have a better understanding of the strain measurement set up and behaviour of the system.

1. INTRODUCTION

In manufacturing higher grade paper qualities the paper web is led through several loaded roll pair contacts. The typical unit processes of that kind are calendaring and coating of which the purpose is to make the paper more solid or improve its surface qualities. The conditions in the loading zone and especially the line load distribution in the contact have a great effect on the quality of the final product. However, accurate control and measuring of the line load distribution is still very hard.

Under heavy line loads the rolls deform elastically in both axial and tangential direction which causes deviation in the line load also. The strains of the rolls are measurable and as they reflect the line load conditions in the contact zone one could use them to analyse the actual line load distribution during run. In this situation the influence coefficient method could be applied. In the method an influence coefficient matrix is constructed to give a direct relation between the known strain measurements and the loading of the roll structure.

The goal of this study is to examine the use influence coefficient method on a hollow roll structure under line loading. The particular roll structure discussed here is a part of a nip unit installation in the Laboratory of Machine Dynamics which is shown in figure 1. This analysis will help to get a better understanding of the strain measurement set up as well as the behaviour of a hollow roll structure. The analysis is purely computational and the calculations of the roll structure are done with two computational models. The first one is the analytical Bernoulli-Euler beam theory and the other one a numerical finite element (FE) model.

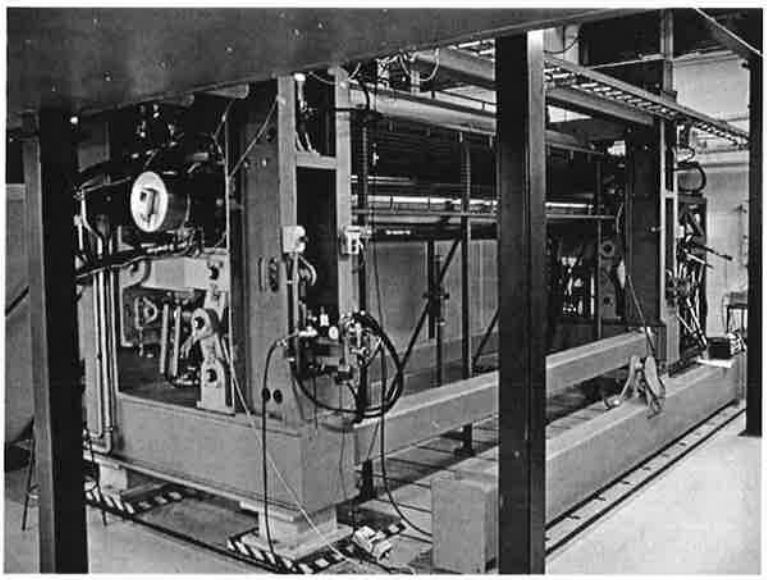


Figure 1. The nip installation of the Laboratory of Machine Dynamics.

The influence coefficient matrix is formed by a special unit load procedure. In this case it means that line loading has to be divided into discrete pieces which are each represented by a certain unit load (point force). Here three different loading systems of unit loads acting on as line load are handled.

As well as the positions of the unit loads also the positions of the strain points are important. The strain set up is chosen after analysing the strains of the hollow roll structure under line load.

The influence coefficient matrices are formed for the chosen set ups with the two computational models. The models are then subjected to an ideal even line load of which the strain response is calculated. Based on the strain information in certain points, the current load distribution is solved by using corresponding influence coefficient matrix. Finally, the computed line load solutions are compared with the known line load distributions. The suitability of different strain measurement set ups and the corresponding influence coefficient matrices are evaluated.

2. INFLUENCE COEFFICIENT MATRIX

2.1 Basic Concept

The basic concept of the influence coefficient method is general and versatile, it lies in the theory of matrix structural analysis. The actual continuous system is formulated to a discrete element system of finite number of degrees of freedom upon which matrix algebra operations can be formed. The degrees of freedom are typically based on the measurements. This kind of force-displacement relationship is actually the basis for finite element method (FEM), force or displacement method [4].

The measured quantity could be almost anything and in our case we have chosen strain. Then the basic equation of influence coefficient method is

$$\boldsymbol{\varepsilon} = \mathbf{S}\mathbf{f} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the strain vector, $\mathbf{S}$ the coefficient matrix and $\mathbf{f}$ the loading vector.

From eq. (1) the loading vector **f** can be inversely solved by

$$\mathbf{f} = \mathbf{S}^{-1}\boldsymbol{\varepsilon} \tag{2}$$

Thus ICM presents us an inverse way to measure loading.

Suppose the force vector is combined from a complete set of linearly independent unit load case vectors

$$\mathbf{f} = \mathbf{L}\boldsymbol{\varphi} \tag{3}$$

where the column vectors of load case matrix **L** are the unit load case vectors.

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & \mathbf{L}_2 & \dots & \mathbf{L}_n \end{bmatrix} \tag{4}$$

The relationship between load case factors and strains is now

$$\boldsymbol{\varepsilon} = \mathbf{SL}\boldsymbol{\varphi} \tag{5}$$

The problem stated now is to solve the unknown loading components **f** and loading case factors $\boldsymbol{\varphi}$ using the measure information coming from the strain gages. In order to use the model one has to determine first the coefficients in matrices **S** and **L**. By definition **L** is known because it is given by the user of the model. Influence matrix **S** needs more analysis.

The method is valid for linear cases and it utilizes the theory of superposition. In order to apply the method to a certain structure, one has to determine first the influence coefficient matrix. This is done by applying unit loads at desired loading points of which the influence, the unit strain response at the measuring points, is stored as the columns of the influence coefficient matrix. The determination process can be done based on experimental measurements under unit loadings or a computational method, an analytical or a FE model, with which one can simulate the unit loading procedure. Here, the analysis is done purely computationally.

Doyle [1] discusses about the same type of method by the name of sensitivity response method. This method uses computational calculations available to construct a sensitivity matrix which is used in the same fashion as the influence coefficient matrix in the inverse problems. The influence coefficients can also be thought as functions as Hurty and Rubinstein [2] show for a simple Bernoulli-Euler beam.

2.2 The Unknown Loading Distribution Cases

It is possible to define the unknown loading distribution based on the information of the strain measurements $\boldsymbol{\varepsilon}$ and the influence coefficient matrix **S**. Whether we define the influence coefficient matrix experimentally or computationally, the determination of the unknown load distribution leads to the following cases:

a) There are more strain points than loading points, $n < m$.
b) There are equal amount of strain points and loading points, $n = m$.
c) There are fewer strain points than loading points, $n > m$.

The number of equations in case (a) is larger than the number of unknown loading factors f_i which means that the strain information available for solving the factors is larger than needed. In this situation the solution can, however, be evaluated numerically by means of pseudo-inverse solution in form

$$\mathbf{f} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\boldsymbol{\varepsilon} \tag{6}$$

The situation (b) is straight forward solved by equation (2).

In case (c) the number of equations is less than the number of unknowns. Thus there is not enough information available for the solution.

For cases (a) and (b) the decomposition of load components to different load cases is obtained by

$$\boldsymbol{\varphi} = \mathbf{L}^{-1}\mathbf{f} \tag{7}$$

3. COMPUTATIONAL MODELS

The basis of the computational models is a real roll structure. The roll in question is the lower roll of nip unit installation in the laboratory of Machine Dynamics. The size of the nip unit is 1:2 as compared with the corresponding mill units. The span of the supports is *5350 mm* and the length of the roll shell is *4400 mm*. Inner and outer diameters of the roll are *425* and *525 mm*, thus the shell thickness is *50 mm*. The nip unit installation is illustrated in figure 1. The strains are calculated on the inner wall surface of the roll for future measuring references.

3.1 Beam theory

The bending strain on the inner surface of the roll illustrated in figure 2 is calculated by using the Bernoulli-Euler beam theory [3].

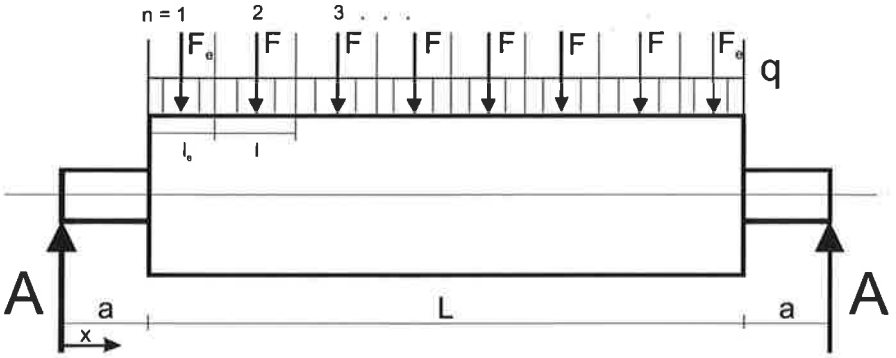


Figure 2. Roll for the beam theory.

The bending strain is calculated using a straight-forward application of Hooke's law

$$\varepsilon(x) = \frac{M_t(x)}{EI} r \tag{8}$$

where E is the elastic modulus, I the moment of the cross-section, x the distance from the left end and r the inner radius of the roll. $M_t(x)$ is the bending moment of the beam which is calculated according to the Bernoulli- Euler beam theory for line load by

$$M_t(x) = Ax - \frac{q(x-a)^2}{2} \tag{9}$$

In equation (9) A is the force at the supports, q is the line load, L the length of the roll and a the length of the unloaded ends of the roll.

For equally distributed point forces F ($F_e = F$, $l = l_e$, in figure 2), starting from the left, the bending moment is calculated by

$$M_t(x) = \frac{qL}{2}x - Fl\sum_{0}^{n-1}k \tag{10}$$

where l is the distance between point forces and n the location number of the measuring point starting from the left end.

For cases in which the point forces on the edges F_e act on a different-size length l_e then the other forces F, the bending moment can be calculated starting from the left end

$$M_t(x) = \frac{qL}{2}x - F_{edge}l_{edge}(n-1), \quad n \le 2 \tag{11}$$

$$M_t(x) = \frac{qL}{2}x - F_{edge}\left[l_{edge} + (n-2)l\right] - Fl\sum_{0}^{n-2}k, \quad n > 2 \tag{12}$$

3.2 FE model

The FE model used is constructed of three solids, the ends and the shell. It has 128462 10-node hexagonal elements and 206008 nodes. There are three element layers on the shell and both ends are simply supported from the rotation centres of the bearings. The model is shown in figure 3.

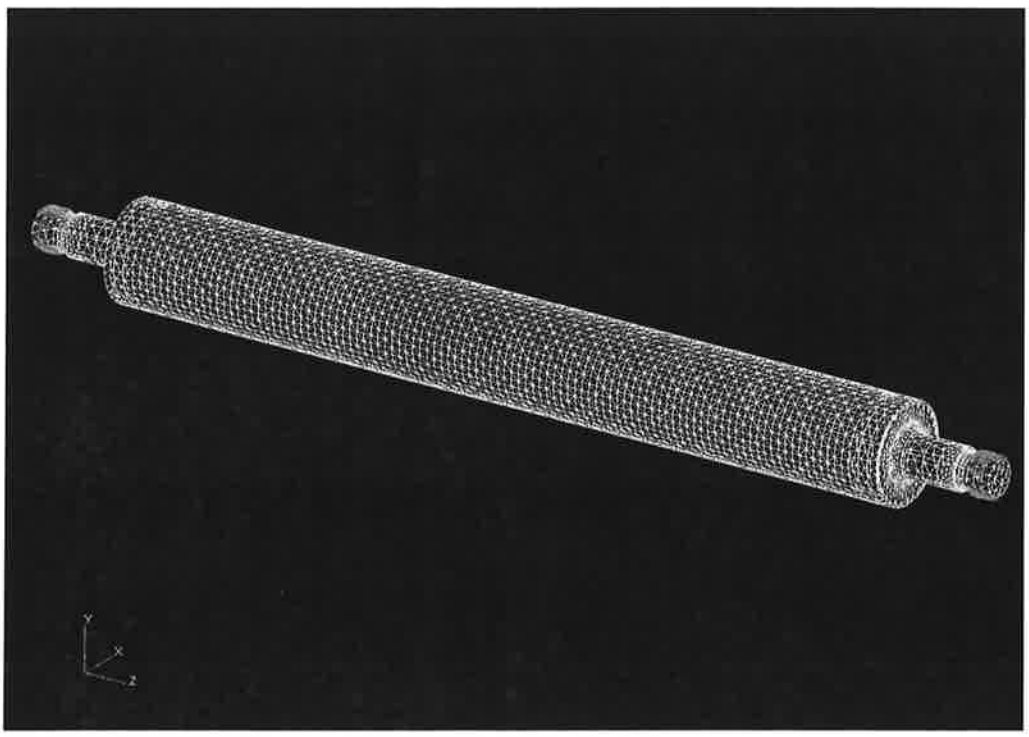


Figure 3. The FE model.

4. LOAD DISTRIBUTION VS. LINE LOAD

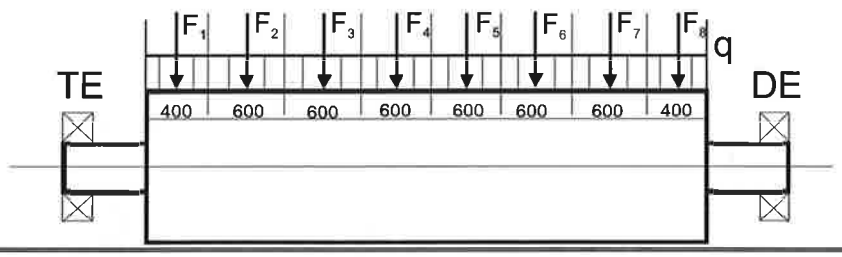
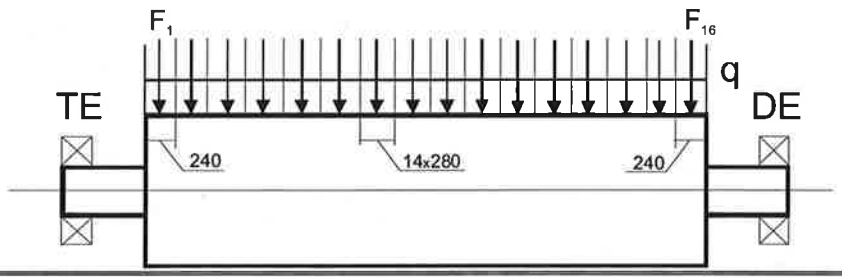
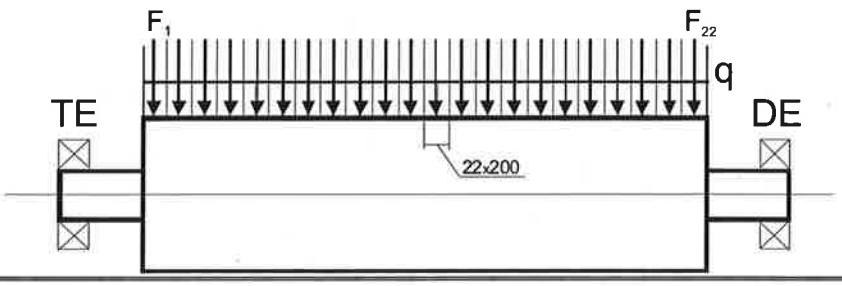
Pressing of two rolls together creates a continuous line load in the contact zone between the two rolls. If we were to measure the continuous line load with a discrete amount of sensors using the influence coefficient method, we would have to divide the line load into smaller pieces. Here, these smaller pieces of load distribution will each be represented by a point force. The point forces are each an element of the loading vector **f**. This approach gives us the opportunity to form the required equations for the influence coefficient method. When solving individual point forces acting on a structure, the application of the method is a more straight-forward procedure [5].

Using point forces is the simplest way of replacing the line load with smaller pieces. Another more complex option would be to divide the line load into certain amount of continuous functions, which would be characterized by parameters.

4.1 Point force distributions

For the analysis three different point force distributions were chosen, which respectively consist of 8, 16 and 22 point forces. The first two cases are equally spaced expect for the two smaller "line load pieces" at each end, while the third has all pieces equally spaced. The load cases are illustraded in table 1.

Table 1. The three different point force cases. TE tending end, DE drive end.

#	Loading case
I	
II	
III	

4.2 Plotting the strains

The three point force cases are compared to continuous line load solutions of both models. The bending strains are calculated in both top and bottom (opposite) surfaces of the inner wall with both computational models. Also the tangential strains are calculated with the FE model. The strain results of these calculations are presented in figures 4 - 9.

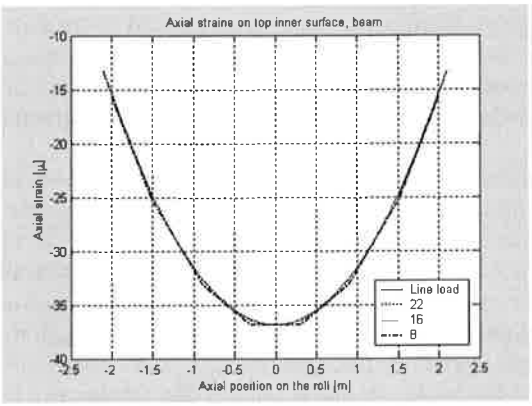


Figure 4. Top strain, beam theory.

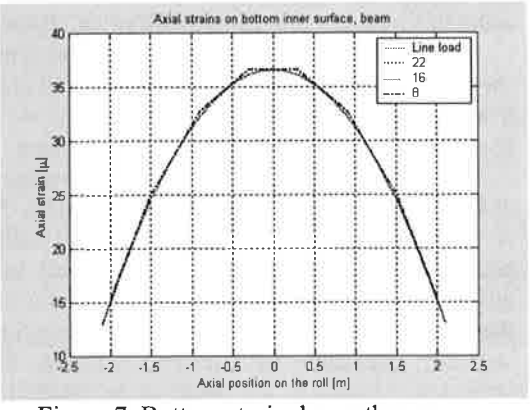


Figure 7. Bottom strain, beam theory.

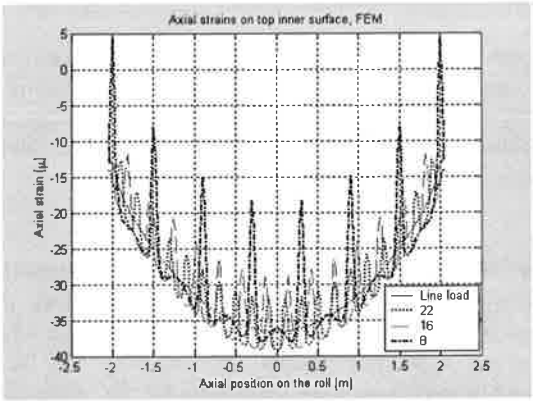


Figure 5. Top axial strain, FE model.

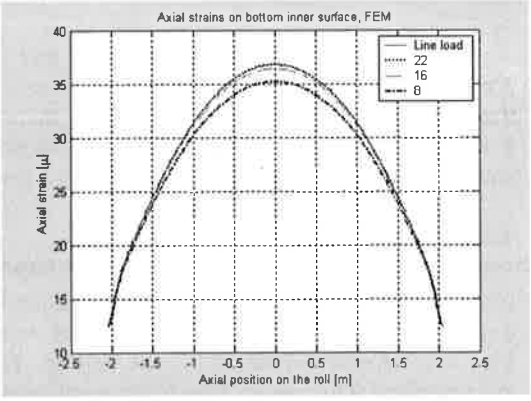


Figure 8. Bottom axial strain, FE model.

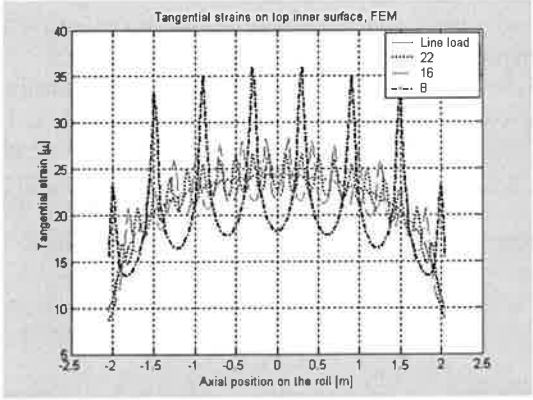


Figure 6. Top tangential strain, FE model.

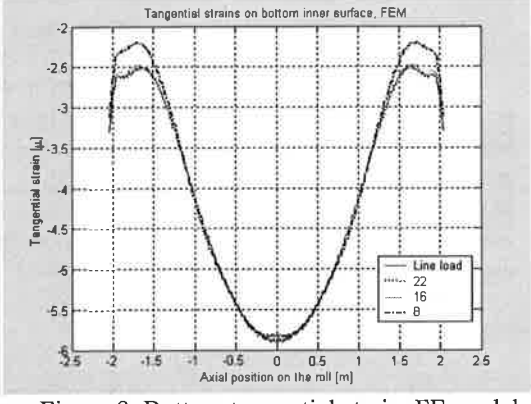


Figure 9. Bottom tangential strain, FE model.

Replacing a continuous line load with a discrete load distribution has noticeable effect on the behaviour of the roll structure. On a local scale deformations increase near the point forces. If we take a closer look of the bending strain of the roll structure of the beam theory, in figures 4 and 7 one can in fact notice the differences between the continuous line load and the coarsest point force distribution near the point forces. But in one point between each point force location, the bending strain is the same for both loading cases. These points are exactly in the place where the boundaries of the smaller pieces of line load are. Thus, if one would like to measure line load in using this kind of point force distribution, these points between the point forces would be the best.

When looking at the strains computed with the FE model, especially the axial direction, it shows that the two coarser load cases (8 and 16) tend to give a bit lower strains than the actual line load case. The 22 case is almost identical to the actual line load, thus 22 point forces on a structure like this seems to suitable for line load sensing.

The beam theory does not take into account the shell deformations and therefore it does not distinguish the top and bottom loading cases. From the strain plots of the FE model in figures 5 and 6 it can clearly be seen that the shell deformation has an influence to the strains of the top position. The local strains near the point forces are significantly large and, the coarser the distribution, the bigger the effect of local strains is. On the top position the strain distribution from the point force cases does not represent the strain distribution of the line loading well although it is statically equivalent with it. This refers to Saint Venant's principle, which states that in the immediate vicinity of the points of application of the loads, the stress (strain) distribution can not be assumed to be independent of the actual mode of application of the loading. Also, the computation in the immediate vicinity of the points of loading is unreliable, even when FEM is used [3].

Judging from Saint Venant's principle and figures 5 and 6 we can make a conclusion that if we are trying to measure line loading with discrete measuring system on this kind of a roll structure, the measuring points have to be located a bit away from the loading points. As it shows in figures 8 and 9, measuring points located on bottom position are suitable for line load sensing as the local strains of the discrete loading are almost non-existent.

4.3 Choosing the measuring points

According to the results of the beam theory, best correlation between the line load and the point force distributions is attained at the boundary points of line load pieces. The results of the FE model were more delicate to the amount of point forces representing the line loading. On the other hand the strains measured in the bottom position were not strongly affected by the local deformations. Thus, these boundary points on the bottom position are chosen to be the measuring points for the analysis of the influence coefficient method. The only problem is, that there are not enough of these boundary points (real measuring in the edge of the shell would be impossible) because the method requires at least as many measuring points as there is loading points. This means that one has to choose one other than a boundary point also.

From beam models three strain set ups is chosen each having the same amount of measuring points as loading points. The measuring points contain respectively the boundary points and a point at *200 mm* from the shells DE. The same bending set ups as well as same tangential set ups are chosen from the FE models. In addition, mixed set ups using the preceding measurement points, which utilize both the axial and tangential strains are chosen for load cases II and III. This means there is twice the number of strain as there is loading points. Thus equation (6) has to be used.

5. THE RESULTS

The three different line load cases were handled with the ten different influence coefficient matrices for both computational models. The roll was objected with a typical even *15 kN/m* line load. Based on the strain computations in measurement points, the inverse solution was calculated.

The condition of the matrices was examined by determining a condition number respectively for each matrix using a singular value decomposition analysis.

The results are assembled in tables 2 and 3. The "known forces" section indicates the actual magnitude of the elements of the load vector, which should be in the inverse result also.

Table 2. The results from beam calculations.

Loading case		Known forces	The force vector from the inverse solution	Cond. number
BEAM	8	$F_e = 6$ $F = 9$	$\mathbf{f}=[7.3 \;\; 7.0 \;\; 11.0 \;\; 7.0 \;\; 11.0 \;\; 7.0 \;\; 11.0 \;\; 4.6]^T$	330
	16	$F_e = 3.6$ $F = 4.2$	$\mathbf{f}=[5.4 \;\; 1.2 \;\; 7.2 \;\; 1.2 \;\; 7.2 \;\; \ldots \;\; 1.2 \;\; 7.2 \;\; 1.8]^T$	2900
	22	$F = 3$	$\mathbf{f}=[5.0 \;\; -0.5 \;\; 6.5 \;\; -0.5 \;\; 6.5 \;\; \ldots \;\; -0.5 \;\; 6.5 \;\; 1.0]^T$	7700

Table 3. The results from FEM calculations.

Loading case	ax8	ax16	ax22	tang16	tang22	mix16	mix22
The force vector from the inverse solution of ICM	$\mathbf{f}=\begin{bmatrix}6.9\\7.6\\9.7\\9.0\\8.1\\10.5\\7.2\\7.0\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}4.1\\5.2\\-0.6\\8.0\\5.4\\2.5\\3.3\\6.7\\0.0\\6.9\\5.6\\6.8\\-10.4\\22.6\\-4.9\\4.8\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}1.9\\5.7\\-9.8\\20.2\\0.3\\10.0\\-17.4\\11.3\\1.7\\11.8\\-1.9\\-4.5\\17.3\\0.0\\-2.0\\-0.3\\-4.0\\17.5\\8.8\\-4.6\\-3.3\\7.2\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}17.7\\0.8\\-3.3\\18.6\\-14.5\\15.1\\-0.4\\-1.7\\7.3\\4.4\\-8.5\\21.5\\-18.1\\25.2\\-14.7\\21.4\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}-0.4\\2.1\\7.0\\-1.1\\6.1\\0.3\\7.2\\0.6\\1.7\\9.8\\-6.4\\13.9\\-6.3\\10.9\\-0.5\\2.4\\6.5\\0.1\\6.5\\-1.9\\8.3\\-2.1\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}17.7\\0.8\\-3.3\\18.6\\-14.5\\15.1\\-0.4\\-1.7\\7.3\\4.4\\-8.5\\21.5\\-18.1\\25.2\\-14.7\\21.4\end{bmatrix}$	$\mathbf{f}=\begin{bmatrix}0.6\\10.1\\-4.6\\6.2\\0.4\\10.0\\-3.5\\1.3\\9.1\\2.2\\0.7\\-2.2\\13.6\\0.0\\-1.5\\0.7\\13.0\\0.8\\-7.1\\13.8\\-1.1\\3.3\end{bmatrix}$
Known forces	$F_e = 6$ $F = 9$	$F_e = 3.6$ $F = 4.2$	$F = 3$	$F_e = 3.6$ $F = 4.2$	$F = 3$	$F_e = 3.6$ $F = 4.2$	$F = 3$
Cond. number	29000	2.2×10^{12}	3.9×10^{12}	12300	5300	4.1×10^{11}	4.8×10^{11}

The results were somewhat disappointing. The matrix equation appeared to be extremely sensitive. This means that a small change in some element of the strain vector causes big deviations in the solving of the force distribution vector from the matrix equation (1). This can be seen clearly from the high condition number of the matrices also. The phenomena happened even though the actual inverse of the influence coefficient matrix was not computed. The solving of the unknown vector was done by using Gaussian elimination which should be more efficient. Sensitivity could also be an elementary problem of the whole method regardless of structure, because deviations occur using beam theory also.

If we take a closer look on the beam theory, indications of efficiency can be noticed. Good and in the matter of a fact precise inverse results can be reached when one uses the appropriate measuring points, which were previously discussed in chapter 4.2 and 4.3. These ideal measuring points are located in the boundaries of the discrete line load pieces. In order to get enough measuring points, one of the edge points of the roll shell has to be chosen. These results are shown in table 4. In reality, that would be an impossible measuring point.

Table 4. The results from beam calculations, ideal measurement points

Loading case		Known forces	The force vector from the inverse solution	Cond. number
BEAM	8	$F_e = 6$ $F = 9$	$\mathbf{f}=[6\ 9\ 9\ 9\ 9\ 9\ 9\ 6]^T$	276
	16	$F_e = 3.6$ $F = 4.2$	$\mathbf{f}=[3.6\ 4.2\ 4.2\ 4.2\ 4.2\ \ldots\ 4.2\ 4.2\ 3.6]^T$	2700
	22	$F = 3$	$\mathbf{f}=[3\ 3\ 3\ 3\ 3\ \ldots\ 3\ 3\ 3]^T$	7700

Although the ideal measurement points according to the beam theory were chosen, the condition number of the matrices remains bad. This means that if the measured structure would behave exactly according to the beam theory, the inverse measurements would give exactly the right answer. No beam structure, much less a hollow roll structure, behaves ideally and exactly according to the beam theory. The inverse results would be affected from the slightest error or non- ideality in the strain measurements because of the sensitivity of the matrix. This would also be the case if these same ideal measurement points were to be used on the inverse calculations of the FE model.

The cause of the sensitivity might be the fewness of the measuring points in comparison with the loading points. More information of the strain behaviour and less elements to be solved could improve the behaviour of the equation.

The equations could also be improved by conditioning the matrix equations some way. There are quite a lot of different conditioning algorithms available.

Finally, the modelling of the continuous line load with discrete point forces might not be a good way of constructing matrices. Modelling continuous function with a set of continuous functions for determining the influence coefficients could be a better solution to the problem.

6. CONCLUSIONS

The influence coefficient method was analysed using two different computational models, Bernoulli-Euler beam theory and a FE model, on a typical roll structure of a paper machine unit. The method was used to calculate line loading on the roll surface utilizing the strain measurements from the inner wall of the roll.

The continuous line loading was first divided into smaller discrete pieces of line loads which were each represented by a point force for the determination of the influence coefficients.

The strain points used in the inverse solution were chosen according to the compatibility of the three different discretization of the line load. Strain distribution computed from the bottom inner wall opposite to loading proved to be less affected by the local deformations and that way best suited to be used in the inverse solution.

The inverse solution were calculated with 3 different set up of beam theory and 7 different set ups of the FE model. The roll was subjected to a typical even line loading and the strain results were used to compute the inverse solutions. Unfortunately none of the set ups were found to be good to solve the line loading. It was shown that the influence matrices were very sensitive to errors in the input (strain vector), which led to bad results. If the strain points were chosen ideally, the equations from the beam theory gave exact results. Still the condition of the matrices was bad and thus sensitivity high.

To overcome the difficulties one would have to condition the influence coefficient matrix, have more information from the behaviour of the structure or use different way of forming the influence coefficient matrices for example using continuous functions.

REFERENCES

[1] Doyle, James F., Modern Experimental Stress Analysis, John Wiley & Sons Ltd., Chichester, England, 2004, 424 p.

[2] Hurty, Walter C. and Rubinstein, Moshe F., Dynamics of Structures, Prentice-Hall, New Jersey, USA, 1964, 455 p.

[3] Outinen, H. and Salmi, T., Lujuusopin perusteet, Pressus Oy, Tampere, Finland, 2004, 464p.

[4] Przemieniecki, J.S., Theory of Matrix Structural Analysis, Dover Publications, New York, USA, 1968, 468 p.

[5] Romppanen A-J, Keskinen E, Miettinen J, Järvinen . Line Load Determination of Press Rolls by Inverse Coefficient Method. 2005 Society of Experimental Mechanics Annual Conference and Exposition on Experimental and Applied Mechanics, June 7-9, 2005, Portland, Oregon, USA.

COMBINED FINITE-DISCRETE ELEMENT METHOD FOR APPLICATIONS IN ICE MECHANICS

A. POLOJÄRVI, J. TUHKURI

Helsinki University of Technology
Laboratory for Mechanics of Materials
P.0. Box 4300
02015 HUT, Finland
e-mail: arttu.polojarvi@tkk.fi

ABSTRACT

Some problems in ice mechanics can be approached by modelling the ice as a discontinuous media. Examples of such problems are ridge keel loads and ice pile-up against structures. Earlier modelling of discontinuous problems in ice mechanics has applied the discrete element method (DEM). However, a more recent approach on modelling a discontinuum is the combined finite-discrete element method (FEM-DEM).

In the research reported in this paper, a 3D combined FEM-DEM application was implemented. The usability of the application in ice mechanics and the differences between DEM and FEM-DEM are discussed. Small example cases demonstrating the use of FEM-DEM in ice mechanics are given.

1 INTRODUCTION

Marine structures in northern seas are affected by different ice features which can cause high loads on the structures. Some of the ice features have discontinuous nature. Examples of this kind of features include ice ridges and also rubble piles formed against marine structures. Earlier studies in ice mechanics using discontinuum mechanics include studies of ice ridging, simulations on ridge keel punch-through tests and studies on ice pile-up against inclined structures [1]-[5].

In the research reported in this paper, a 3D finite-discrete element method application was implemented. The application includes features needed in research on ice mechanics. These features include buoyancy, friction and fluid drag of arbitrary shaped polyhedral rigid bodies. The basis of these features are presented first and then small example cases are given.

2 COMBINED FINITE-DISCRETE ELEMENT METHOD

Discrete element method (DEM) and combined finite-discrete element method (FEM-DEM) are techniques for modelling granular media. In both methods, the interaction between particles in the discontinuous system occurs through pairwise collisions from which the forces acting on the particles are derived according to the chosen interaction laws.

DEM models are divided in two groups according to the way which contact forces are derived [6]. In soft contact approach the particles are allowed to interpenetrate and contact

forces are derived by using chosen contact stiffness and damping etc. to model deformation on the surface of the contacting particles (e.g. [3],[7],[8]). The physical assumption in hard contact approach is that that no point is space is occupied by bodies of two separate particles simultaneously and the contact forces are determined to enforce the condition of no interpenetration of the particles. In the current FEM-DEM application the hard contact approach is approximated as presented in the following section.

The main phases during one time step of a FEM-DEM simulation are (a) neighbor search, (b) contact resolution analysis and (c) updating particle positions for the next time step. On phase (a), the particles in the vicinity of each other and possibly interacting through contacts on a given time step, are found without referring to their exact geometry. This is done to avoid computationally intensive phase (b) on every pair of particles within the system. On phase (b) the contact geometries are solved and the contact forces are determined. The particles are then moved into their new positions on phase (c) according to the forces acting on them using Newton's laws and finite difference methods.

In FEM-DEM the discrete element geometries are discretized into finite elements. The contact resolution analysis and calculation of external forces acting on discrete elements is done finite element by finite element. The total force acting on a discrete element is achieved by superposing the forces acting on its finite elements. In the current application discrete elements were discretized into tetrahedron shaped finite elements.

2.1 Contact forces

In the current FEM-DEM application rigid particles are used and the contact forces are derived by using an approximation of the hard contact approach through using penalty function and a potential contact force method [9].

In the potential contact force method the contact forces are derived from a scalar potential φ. $\varphi = \varphi(P)$ is a continuous positive scalar potential with continuous first partial derivatives with respect to spatial coordinates in every point P within element volume Ω and should vanish on element boundaries Γ, i.e

$$\varphi(P) > 0, \qquad P \in \Omega \qquad \wedge \qquad \varphi(P) = 0, \qquad P \in \Gamma. \tag{1}$$

Suitable functions for φ can be achieved from volume coordinates when bodies of discrete elements are meshed into tetrahedron shaped finite elements.

The contact force applied to an infinitesimal volume element $\mathrm{d}\Omega_o$ penetrating into φ is determined from the gradient of φ (Fig. 1):

$$\frac{\mathrm{d}\mathbf{f_c}(P)}{\mathrm{d}\Omega_o} = -\nabla\varphi(P). \tag{2}$$

As the force acting on $\mathrm{d}\Omega_o$ is conservative and φ vanishes on Γ, the total work done on $\mathrm{d}\Omega_o$ penetrating through Ω is independent of path and vanishes.

The total contact force $\mathbf{f}_c$ is determined by computing volume integral over the overlap volume Ω_o of two elements which is reduced into computationally more efficient form by generalized Gauss's theorem (Fig. 2)

$$\mathbf{f}_c = -\int_{\Omega_o} \nabla\varphi(P)\ \mathrm{d}\Omega = -\int_{\Gamma_o} \varphi(P)\mathbf{n}\ \mathrm{d}\Gamma, \tag{3}$$

where $\mathbf{n}$ is the unit outer normal of Γ_o.

The distributed load due to φ given by Equation 3 on the surface of the finite element is then reduced into equivalent force system acting on the nodes belonging to the surfaces

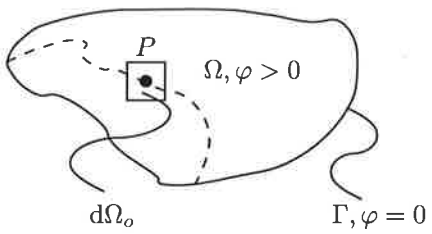


Figure 1: An infinitesimal volume element $d\Omega_o$ at point P penetrating into a finite element with volume Ω.

of the finite element. The total contact force acting on the discrete element is achieved by superposition of nodal loads acting on its nodes.

In the penalty function method the potential φ in Equations 2 and 3 is replaced by $\varphi' = s\varphi$, where s is a positive penalty term. It should be noted that s is not a material parameter, but enforces the boundary condition, hence it affects the maximum allowed penetration within the system of particles.

2.2 Buoyancy and fluid drag

If the discrete elements representing ice blocks are submerged in water they become acted on by buoyant force $\mathbf{f}_b$. The buoyant force is determined using hydrostatic pressure $p = p(h)$ in depth h from the surface. The total buoyant force for a partly submerged element is determined from the surface integral (Fig. 3)

$$\mathbf{f}_b = -\int_{\Gamma_s} p(h)\mathbf{n}\, d\Gamma = -\rho_w g \int_{\Gamma_s} h\mathbf{n}\, d\Gamma, \tag{4}$$

where Γ_s is the submerged part of the surface Γ of the discrete element, ρ_w is the mass density of water and g is the gravitational acceleration.

As discrete element geometries in the current application were discretized into tetrahedrons, Γ is a union of triangular areas and the previous integral has an analytical solution. For a partly submerged triangular surface element, the solution for $\mathbf{f}_b$ is achieved by first finding the intersection of water surface and the element. This may lead to quadrilateral surface element-water intersection areas, which are then divided into triangular areas for a solution to Equation 4.

To add global dissipation into the system, fluid drag was approximated by the following

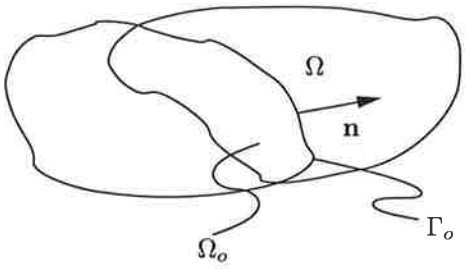


Figure 2: The overlap volume Ω_o of two overlapping elements.

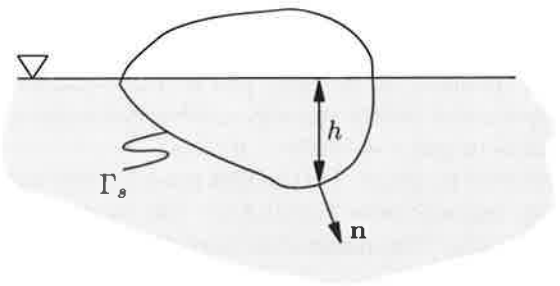


Figure 3: The submerged surface Γ_s of a discrete element and surface normal $\mathbf{n}$ at depth h. Gray area represents water.

surface integral over Γ_s:

$$\mathbf{f}_d = -\frac{1}{2}\rho_w C_d \int_{\Gamma_s} (\mathbf{u}_n \cdot \mathbf{u}_n)\,\mathbf{n}\, \mathrm{d}\Gamma, \qquad \text{if } \mathbf{u}_n \cdot \mathbf{n} > 0 \tag{5}$$

and $\mathbf{f}_d = 0$ else. In the previous equation ρ_w is the mass density of the water, C_d is the drag coefficient and $\mathbf{u}_n$ is the velocity into the normal direction of Γ_s.

2.3 Friction

A simple model of the frictional forces during sliding was used. This was done using Coulomb's law and relative velocities of contacting particles [10]. According to Coulomb's law the frictional force $\mathbf{f}_\mu$ has a magnitude proportional to the normal force $\mathbf{f}_n$ with proportionality of friction coefficient μ

$$|\mathbf{f}_\mu| = \mu |\mathbf{f}_n|. \tag{6}$$

In the previous equation, the normal force is achieved from the contact resolution analysis, i.e. $\mathbf{f}_n = \mathbf{f}_c$ (Eq. 3).

When bodies are in contact and sliding, the force exerted on them has a direction parallel with the sliding velocity $\mathbf{u_s}$

$$\mathbf{f}_\mu \times \mathbf{u_s} = 0 \tag{7}$$

but with opposite sense of direction

$$\mathbf{f}_\mu \cdot \mathbf{u_s} < 0. \tag{8}$$

Sliding velocity $\mathbf{u_s}$ is determined at the point of contact $\mathbf{x}_c$ from the known velocities $\mathbf{u}_i$ and $\mathbf{u}_j$ and angular velocities $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$ of contacting discrete elements i and j

$$\mathbf{u}_s = \mathbf{u}_i + [\boldsymbol{\omega}_i \times (\mathbf{x}_c - \mathbf{x}_i)] - \mathbf{u}_j - [\boldsymbol{\omega}_j \times (\mathbf{x}_c - \mathbf{x}_j)]\,. \tag{9}$$

The point of contact, $\mathbf{x}_c$, is determined in calculation of equivalent nodal loads due to collision.

3 EXAMPLE

A small example case representing features of the current application was set up as follows. Three dimensional ice blocks of the shape of rectangular parallelepiped were realeased

one by one from the water depth of $y = -0.3$ m. The origin of the coordinate system was fixed onto the plane of the water surface with positive y-axis pointing upwards making the x- and z-axes lie on the plane of the surface. The x- and z-coordinates of the blocks at the time of release were generated with a random number generator and were allowed to vary within $x = -0.15 \ldots 0.15$ m and $z = -0.15 \ldots 0.15$ m.

The simulation was run two times. On the first run a friction coefficient $\mu = 0.3$ was used and on the second run contacts were frictionless. The location of the release of particles was the same for both runs. The parameters used in the simulations are summarized in Table 1.

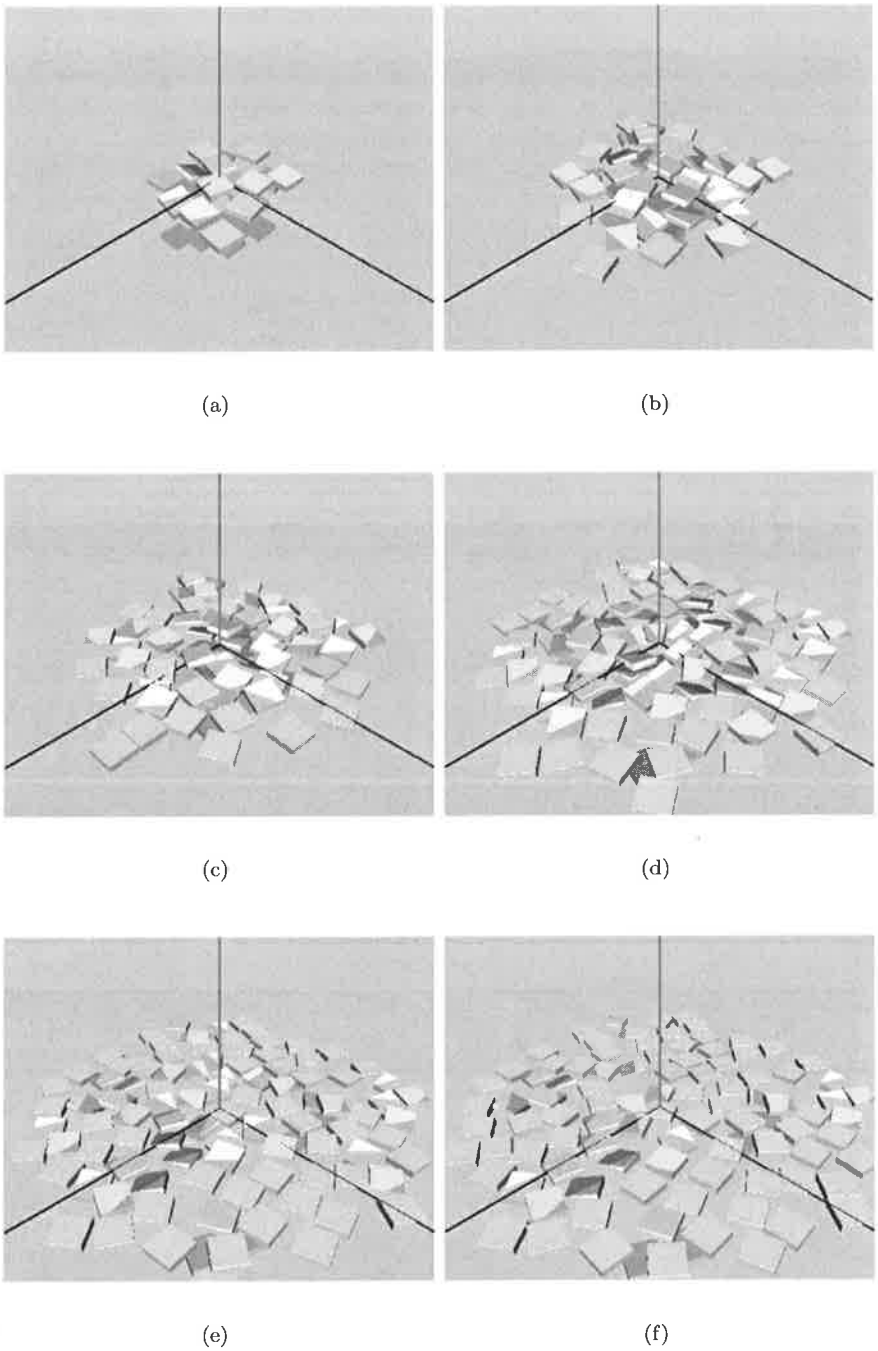
The snapshots of the simulation run with $\mu = 0.3$ are represented in Figures 4a-f. The snapshots were taken at intervals of 10 s of the simulation time t starting from $t = 10$ s (Fig. 4a). The change in the configuration of the system is faster during the earlier stages of the simulation due to new particles surfacing and forcing the particles on top of them to displace (Figs. 4b-d). The last particle was released at $t = 36.26$ s and as most of the particles have surfaced, the configuration of the system becomes more stable (Figs. 4e and 4f). At $t = 60$ s, most of the particles are floating free and only a few of them are forced to remain below surface due to particles above them (Fig. 4f).

In Figures 5a-c, series of underwater snapshots of the simulations are presented at times $t = 20$ s, $t = 25$ s and $t = 30$ s. In the simulation with $\mu = 0.3$ (left column of Figures 5a-c), friction causes the particles to form a small pile of ice underwater. In the case with no friction, this type of behaviour of the system was not detected.

The energetics of the two simulations are presented in Figures 6 and 7. As in both the simulations the particles were released from a known depth, the total potential energy E_p of the system could be calculated beforehand. From the figures it becomes obvious that the work W_d of water drag accounts for the major part of the energy dissipation in both simulated systems. The total kinetic energy E_k was a small fraction (~ 0.3 %) of the total energy of the system in both simulations due to high dissipation through W_d, and thus E_k is not presented in the figures. From Figure 6 it is noticed that the dissipation W_μ due to friction has the highest rate before $t = 40$ s of simulation time. This is due to the high amount of particles in sliding contact. The rate of W_μ drops close to zero as the particles have surfaced and only a few sliding contacts occur. The stability of the simulation was monitored through the energy balance. The variation of the total energy was less than 1.48 % in the case with $\mu = 0.3$ and less than 1.03 % in the frictionless case.

Table 1: Parameters used in the simulations.

	Parameter	Value
General	number of particles	100
	number of volume elements	600
	number of surface elements	1200
	penalty term, s	$1.0 \cdot 10^6$
	time step	$2.5 \cdot 10^{-4}$
Ice blocks	length [m]	0.1
	width [m]	0.1
	height [m]	0.025
	mass density [kgm^{-3}], ρ_i	920
	friction coeffient, μ	0, 0.3
	drag coefficient, C_d	1.025
Water	mass density [kgm^{-3}], ρ_w	1010

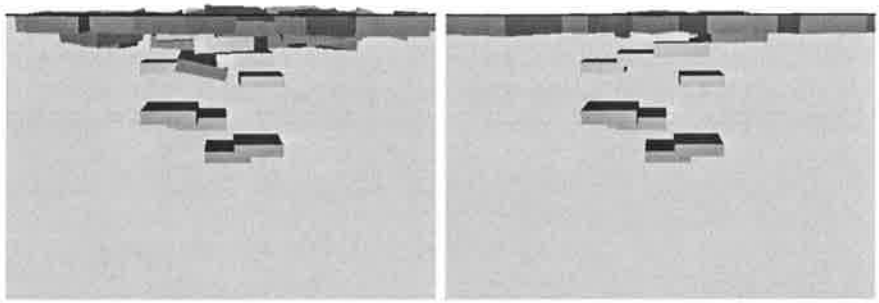


(a) (b)

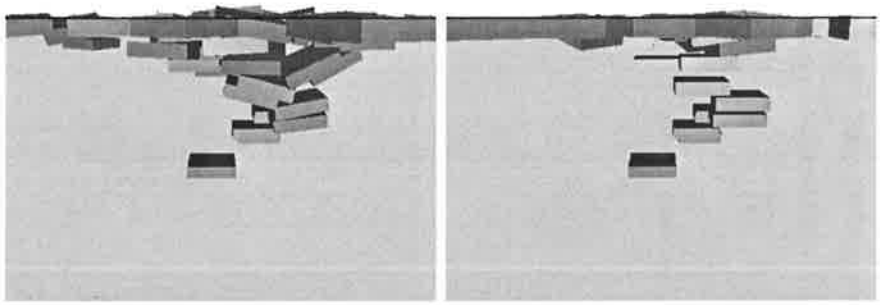
(c) (d)

(e) (f)

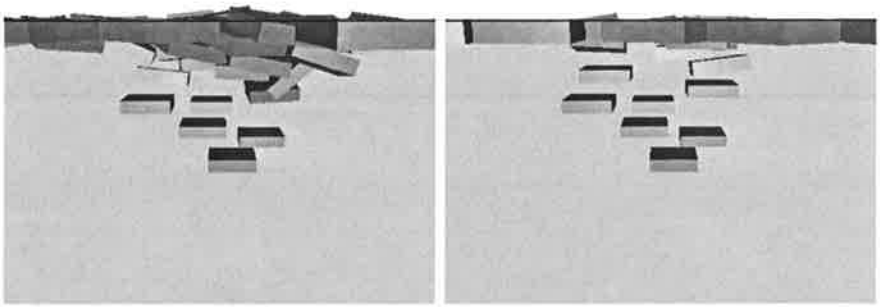
Figure 4: Snapshots of a simulation with $\mu = 0.3$ taken at intervals of 10 s of the simulation time t. The first snapshot (a) is taken at $t = 10$ s and the final (f) at $t = 60$ s. The origin of the coordinates is on the plane of the water surface with y-axis pointing upwards and x- and z-axes lying on the plane of the surface.



(a)



(b)



(c)

Figure 5: Underwater snapshots of the simulations with friction coefficients $\mu = 0.3$ (left column) and $\mu = 0$ (right column) at (a) $t = 20$ s, (b) $t = 25$ s and (c) $t = 30$ s. In the simulation with $\mu = 0.3$, ice blocks form a small pile below the blocks on the surface.

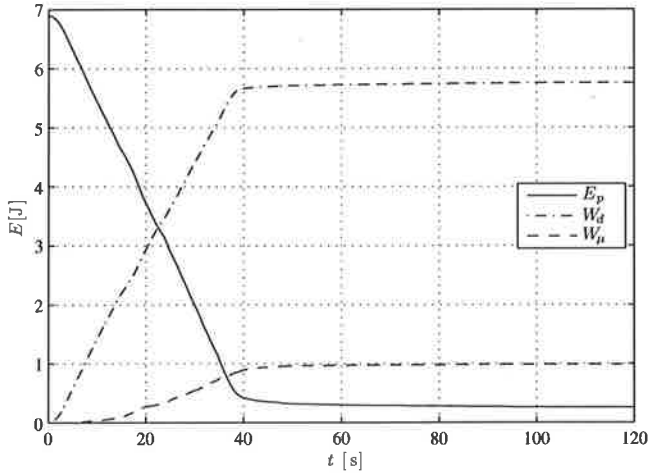


Figure 6: The energetics of the system of the ice blocks with friction coefficient $\mu = 0.3$. Kinetic energy E_k of the system was less than 0.02 J and thus is not presented in the graph.

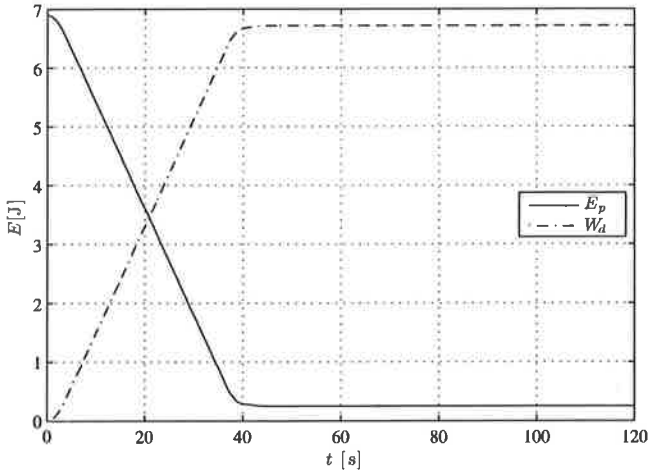


Figure 7: The energetics of the system of frictionless ice blocks. Kinetic energy E_k of the system was less than 0.02 J and thus is not presented in the graph.

In Figure 8, E_k-records of simulations with $\mu = 0.3$ and $\mu = 0$ are presented. At approximately $4 < t < 40$ s the E_k-records for both simulations show high frequency oscillations due to multiple contacts in the simulation.

The E_k-record for the simulation with $\mu = 0.3$ shows multiple separate transients as $t > 40$ whereas in the case with $\mu = 0.0$ E_k vanishes due to W_d after $t \approx 50$ s. The reason for these transients becomes obvious from Figures 9a and 9b where snapshots at $t = 50$ s from simulations with different values of μ are presented. From Figure 9b it is seen that all the particles in the frictionless case have surfaced at $t = 50$ s whereas from Figure 9a some are noticed to be forced to remain beneath surface by other particles. The transients are due to surfacing of these particles.

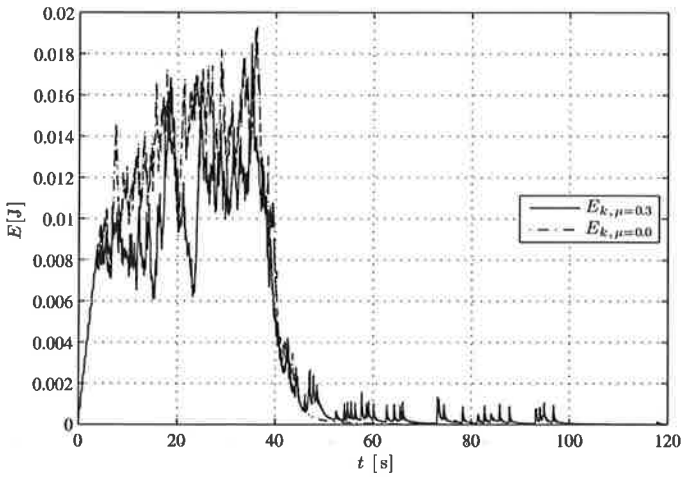
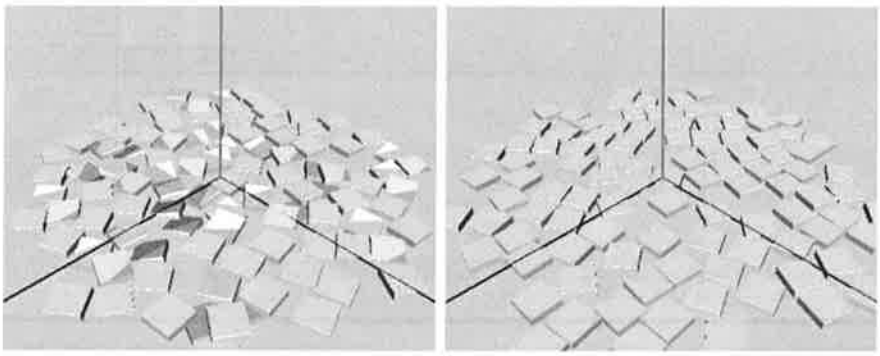


Figure 8: The kinetic energy E_k of the two simulations with different μ.



(a) (b)

Figure 9: Snapshots taken from simulations with $\mu = 0.3$ (a) and $\mu = 0$ (b) at simulation time $t = 50$.

4 CONCLUSIONS

3D combined FEM-DEM simulation was implemented with features needed in the research on ice mechanics. Small example simulations were run to demonstrate the features and the usability of the method.

In the two simulations presented the value of friction coefficient was varied. The first run had frictional contacts whereas in the other run the contacts were frictionless. The behaviour of the simulated system of ice blocks was altered as expected due to friction. The total energy of the system was measured during the simulation and the error in it was noticed to be small.

The simulations represented were run with discrete elements of the shape of rectangular parallelepipeds. It is noted here, that the implementation of the application allows arbitratry shaped polyhedral discrete elements.

REFERENCES

[1] Hopkins, M. A., 1998. *Four Stages of Pressure Ridging.* Journal of Geophysical Research, 103, 21,883-21,891

[2] Hopkins, M. A., Tuhkuri, J., Lensu, M., 1999. *Rafting and Ridging of Thin Ice sheets.* Journal of Geophysical Research, 104, 13,605-13,613

[3] Tuhkuri, J., Polojärvi, A., 2005. *Effect of Particle Shape in 2D Ridge Keel Deformation Simulations.* Proc. of the 18th Int. Conf. on Port and Ocean Engineering, POAC'05, Potsdam, USA. Vol. 2, pp. 939-948

[4] Hopkins, M. A., 1997. *Onshore Ice Pile-up: a Comparison Between Experiments and Simulations.* Cold Regions Science And Technology, 26, 205-214

[5] Paavilainen, J., Tuhkuri, J., Polojärvi, A., 2006. *Discrete Element Simulation of Ice Pile-up Against an Inclined Structure.* To appear in the proceedings of IAHR 2006.

[6] Cundall, P.A., Hart, R.D, 1992. *Numerical modelling of Discontinua.* Engr. Comp., 9(2), 101-113.

[7] Cundall, P.A., Strack, D.L., 1979. *A Discrete Numerical Model for Granular Assemblies.* Géotechnique, 29, 1, 47-65

[8] Hopkins, M.A., 1992. *The Numerical Simulation of Systems of Multitudinous Polygonal Blocks.* United States Army Corps of Engineers, Cold Regions Research & Engineering Laboratory, Report CR 92-22. 69 p.

[9] Munjiza, A., Andrews, K. F. R., 2000. *Penalty Function Method for Combined Finite-Discrete element systems Comprising Large Number of Separate Bodies.* Int. J. Numer. Engng. 2000, 49, 1377-1396

[10] Hahn, J. K., 1988. *Realistic Animation of Rigid Bodies.* Computer Graphics, Vol. 22, 4, 299-308

A CONSTITUTIVE MODEL FOR STRAIN-RATE DEPENDENT DUCTILE-TO-BRITTLE TRANSITION

S. FORTINO[1], J. HARTIKAINEN[1], K. KOLARI[2], R. KOUHIA[1], T. MANNINEN[1]

[1] Helsinki University of Technology, Structural Mechanics, P.O. Box 2100, FIN-02015 TKK
[2] VTT Technical Research Center of Finland, P.O. Box 1000, FIN-02044 VTT

ABSTRACT

Most materials exhibit rate-dependent inelastic behaviour. Increasing strain-rate usually increases the yield stress thus enlarging the elastic range. However, the ductility is gradually lost, and for some materials there exist a rather sharp transition strain-rate after which the material behaviour is completely brittle.

In this paper, a simple phenomenological approach to model ductile-to-brittle transition of rate-dependent solids is presented. The model is based on a consistent thermodynamic formulation using proper expressions for the Helmholtz free energy and the dissipation potential, which is additively split into damage and visco-plastic parts, and the transition behaviour is obtained through a stress dependent damage potential. In addition, the basic features of the model are discussed and a numerical example is presented.

1 INTRODUCTION

A large number of engineering materials, such as metals, polymers, concrete, soils and rock, can show reduction in the load carrying capacity accompanied by increasing localised deformations after the ultimate load is reached. If this phenomenon is considered as material property, it will lead to a negative slope of the stress-strain diagram, which is known as strain softening. In this study, a phenomenological model, which is capable to describe the strain-rate dependent ductile-to-brittle transition, is presented. The ductile behavior is considered as a viscoplastic feature, whereas the strain softening phenomenon, after reaching the transition strain-rate, is dealt with a continuum damage model.

A demerit pertaining to strain-softening models for classical continua is that they result in problems which are not well-posed in general. The field equations of motion lose hyperbolicity and become elliptic as soon as strain softening occurs. The domain is split into an elliptic part, in which the waves are not able to propagate, and into a hyperbolic part with propagating waves. In static and quasi-static problems, localisation of deformation is usually understood as a synonym to the loss of ellipticity of the underlying rate-boundary value problem. When such problems are solved numerically, the solution of the localisation zone of zero thickness can result in mesh sensitivity. A simple remedy is to include viscous effects in the plastic model as proposed in [1]. Other improvements, e.g. [2, 3, 4], are also possible. However, they usually involve additional field unknowns which make numerical computations much more time-consuming. Also, the physical interpretation of the additional boundary conditions can be ambiguous.

2 THERMODYNAMIC FORMULATION

The constitutive model is derived using a thermodynamic formulation, in which the material behaviour is described completely through the Helmholz free energy and the dissipation potential in terms of the variables of state and dissipation and considering that the Clausius-Duhem inequality is satisfied [5].

The Helmholtz free energy

$$\psi = \psi(\epsilon_e, \omega) \tag{1}$$

is assumed to be a function of the elastic strains, ϵ_e, and the so-called continuity or integrity, ω. The elastic strains together with the inelastic strains, ϵ_i, constitute the total strains as

$$\epsilon = \epsilon_e + \epsilon_i. \tag{2}$$

The continuity in turn is a function of the scalar damage parameter, D, as

$$\omega = 1 - D. \tag{3}$$

It is used instead of the damage parameter to simplify the notation.

As usual in the solid mechanics, the dissipation potential

$$\varphi = \varphi(\sigma, Y) \tag{4}$$

is expressed in terms of the thermodynamic forces σ and Y dual to the fluxes $\dot{\epsilon}_i$ and $\dot{\omega}$, respectively. The dissipation potential is associated with the power of dissipation, γ, such that

$$\gamma = \frac{\partial \varphi}{\partial \sigma} : \sigma + \frac{\partial \varphi}{\partial Y} Y. \tag{5}$$

Convexity is not a prerequisite for the dissipation potential but the condition that the product $(\partial\varphi/\partial\sigma) : \sigma + (\partial\varphi/\partial Y)Y$ is non-negative.

The Clausius-Duhem inequality, in the absence of thermal effects, is formulated as

$$\gamma \geq 0, \quad \gamma = -\rho\dot{\psi} + \sigma : \dot{\epsilon}, \tag{6}$$

where ρ is the material density. Using definitions (2) and (5) and defining that $\partial\psi/\partial\omega = Y$, inequality (6) can be reformulated as

$$\left(\sigma - \rho\frac{\partial \psi}{\partial \epsilon_e}\right) : \dot{\epsilon}_e + \left(\dot{\epsilon}_i - \frac{\partial \varphi}{\partial \sigma}\right) : \sigma + \left(-\dot{\omega} - \frac{\partial \varphi}{\partial Y}\right) Y = 0. \tag{7}$$

Then, if eq. (7) holds for any evolution of $\dot{\epsilon}_e$, σ and Y, inequality (6) is satisfied and the following relevant constitutive relations are obtained:

$$\sigma = \rho\frac{\partial \psi}{\partial \epsilon_e}, \qquad \dot{\epsilon}_i = \frac{\partial \varphi}{\partial \sigma}, \qquad \dot{\omega} = -\frac{\partial \varphi}{\partial Y}. \tag{8}$$

3 PARTICULAR MODEL

A particular expression for the free energy, describing the elastic material behaviour with the reduction effect due to damage, is given by

$$\rho\psi = \frac{1}{2}\omega\epsilon_e : C_e : \epsilon_e \tag{9}$$

where C_e is the elasticity tensor.

To model the ductile-to-brittle transition due to increasing strain-rate, the dissipation potential is decomposed into the brittle damaging part, φ_{d}, and the ductile viscoplastic part, φ_{vp}, as

$$\varphi(\sigma, Y) = \varphi_{\mathrm{d}}(Y)\varphi_{\mathrm{tr}}(\sigma) + \varphi_{\mathrm{vp}}(\sigma), \tag{10}$$

where the transition function, φ_{tr}, deals with the change in the mode of deformation when the strain-rate $\dot{\epsilon}_{\mathrm{i}}$ increases. Applying an overstress type of viscoplasticity [6, 7, 8] and the principle of strain equivalence [9, 10], the following choices are made to characterize the inelastic material behaviour:

$$\varphi_{\mathrm{d}} = \frac{1}{r+1}\frac{Y_{\mathrm{r}}}{\tau_{\mathrm{d}}\omega}\left(\frac{Y}{Y_{\mathrm{r}}}\right)^{r+1}, \tag{11}$$

$$\varphi_{\mathrm{tr}} = \frac{1}{pn}\left[\frac{1}{\tau_{\mathrm{vp}}\eta}\left(\frac{\bar{\sigma}}{\omega\sigma_{\mathrm{r}}}\right)^{p}\right]^{n}, \tag{12}$$

$$\varphi_{\mathrm{vp}} = \frac{1}{p+1}\frac{\sigma_{\mathrm{r}}}{\tau_{\mathrm{vp}}}\left(\frac{\bar{\sigma}}{\omega\sigma_{\mathrm{r}}}\right)^{p+1}, \tag{13}$$

where parameters τ_{d}, r and n are associated with the damage evolution, and parameters τ_{vp} and p with the visco-plastic flow. In addition, η denotes the inelastic transition strain-rate. The relaxation times τ_{d} and τ_{vp} have the dimension of time and the exponents $r, p \geq 0$ and $n \geq 1$ are dimensionless. $\bar{\sigma}$ is a scalar function of stress, e.g. the effective stress $\sigma_{\mathrm{eff}} = \sqrt{3J_2}$, where J_2 is the second invariant of the deviatoric stress. The reference values Y_{r} and σ_{r} can be chosen arbitrarily, and they are used to make the expressions dimensionally reasonable. Since only isotropic elasticity is considered, the reference value Y_{r} has been chosen as $Y_{\mathrm{r}} = \sigma_{\mathrm{r}}^2/E$, where E is the Young's modulus.

Making use of eqs. (2) and (8), choices (9)-(13) yield the following constitutive equations:

$$\sigma = \omega C_{\mathrm{e}} : (\epsilon - \epsilon_{\mathrm{i}}), \tag{14}$$

$$\dot{\epsilon}_{\mathrm{i}} = \left[\frac{\varphi_{\mathrm{d}}}{(\tau_{\mathrm{vp}}\eta)^n \omega\sigma_{\mathrm{r}}}\left(\frac{\bar{\sigma}}{\omega\sigma_{\mathrm{r}}}\right)^{np-1} + \frac{1}{\tau_{\mathrm{vp}}\omega}\left(\frac{\bar{\sigma}}{\omega\sigma_{\mathrm{r}}}\right)^{p}\right]\frac{\partial\bar{\sigma}}{\partial\sigma}, \tag{15}$$

$$\dot{\omega} = -\frac{\varphi_{\mathrm{tr}}}{\tau_{\mathrm{d}}\omega}\left(\frac{Y}{Y_{\mathrm{r}}}\right)^{r}. \tag{16}$$

Moreover,

$$Y = \rho\frac{\partial\psi}{\partial\omega} = \frac{1}{2}(\epsilon - \epsilon_{\mathrm{i}}) : C_{\mathrm{e}} : (\epsilon - \epsilon_{\mathrm{i}}) = \frac{1}{2\omega^2}\sigma : C_{\mathrm{e}}^{-1} : \sigma. \tag{17}$$

This particular model has the following general properties:

- Elastic stiffness is reduced monotonously due to damage (eq. (14));
- The model does not include any specific yield stress;
- In the absence of damage evolution, the inelastic model (15) behaves under a constant uniaxial strain-rate loading as

  $$\sigma \to (\tau_{\mathrm{vp}}\dot{\epsilon}_0)^{1/p}\sigma_{\mathrm{r}} \quad \text{when} \quad t \to \infty,$$

  where $\dot{\epsilon}_0$ is a prescribed strain-rate;
- In the evolution of damage based on eq. (16), the constraint for the continuity ω that $\omega \in [0, 1]$ is satisfied automatically, since $\omega(x, 0) = 1$, $\dot{\omega} \leq 0$ and $\dot{\omega} \to 0$ as $\omega \to 0$;

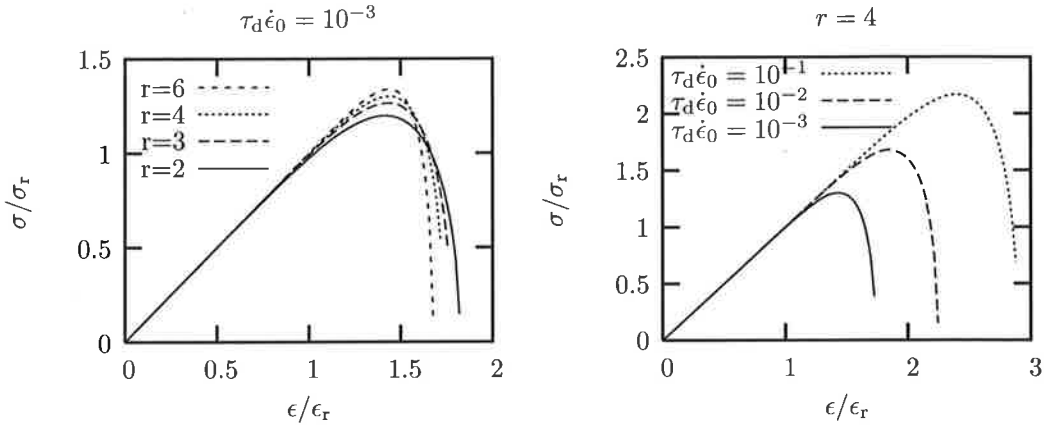


Figure 1: Uniaxial constant strain-rate loading for elastic material with damage.

- The transition function φ_{tr} in eq. (16) deals with the change in the mode of deformation through the damage evolution such that

$$\varphi_{tr} \geq 0 \quad \text{and} \quad \varphi_{tr} \approx 0 \text{ when } \|\dot{\epsilon}_i\| < \eta \quad \text{and} \quad \varphi_{tr} > 1 \text{ when } \|\dot{\epsilon}_i\| > \eta;$$

- Inequality (6) is satisfied *a priori* for any admissible isothermal process. Moreover, the dissipation potential (10) is a non-convex function with respect to the thermodynamic forces σ and Y.

The sensitivity of the material parameters on the damage evolution is studied next for constant strain-rate loading ($\epsilon = \dot{\epsilon}_0 t$). First, the behaviour of the pure damage model, i.e. when $\epsilon_i \equiv 0$ and $\varphi_{tr} \equiv 1$, is shown in fig. 1 for varying relaxation time τ_d and exponent r. In this case the solution can be obtained in the closed form

$$\frac{\sigma}{\sigma_r} = \sqrt{1 - \left(\frac{1}{2}\right)^r \frac{2}{2r+1} \frac{\epsilon_r}{\tau_d \dot{\epsilon}_0} \left(\frac{\epsilon}{\epsilon_r}\right)^{2r+1}} \left(\frac{\epsilon}{\epsilon_r}\right), \tag{18}$$

where the reference strain is defined as $\epsilon_r = \sigma_r/E$.

For the full model the strain rate $\dot{\epsilon}_0$ is varied on both sides of the transition strain-rate η. Both exponents p and r have the value of 4 and the visco-plastic relaxation time τ_{vp} is set to η^{-1} in all cases shown in fig. 2.

Figure 2a shows the stress-strain behaviour for varying damage relaxation times τ_d, as $\tau_d\eta = 10^{-5}$, 10^{-3} and 10^{-2}, and strain rates $\dot{\epsilon}_0$, as $\dot{\epsilon}_0/\eta = 0.1$, 1 and 10, considering that $n = 4$. For the lowest strain rate $\dot{\epsilon} = \eta/10$ all the curves coincide. At the loading rate equal to the transition strain rate, the results for $\tau_d = 10^{-2}\eta^{-1}$ and $10^{-3}\eta^{-1}$ are indistinguishable. As it can be expected, the difference is largest for the fastest loading rates and the deepest strain softening is obtained for the smallest relaxation time $\tau_d = 10^{-5}\eta^{-1}$.

The effect of the transition function exponent n is studied in fig. 2b for the fastest loading rate $\dot{\epsilon} = 10\eta$ and considering that $\tau_d = 10^{-3}\eta^{-1}$. The model seems to be relatively insensitive to the value of the exponent in the power law type of transition function.

The evolution of damage and its thermodynamic conjugate force for the fastest loading rate and varying damage relaxation times are shown in figs. 2c and 2d, respectively. As it can be seen from fig. 2c, largest damage occurs for the smallest values of τ_d.

In figs. 2e and 2f the inelastic strain and strain-rate are shown for $\tau_d\eta = 10^{-3}$ and different loading rates. Observe that $\dot{\epsilon}_i > \dot{\epsilon}_0$ for the fastest loading in the softening region.

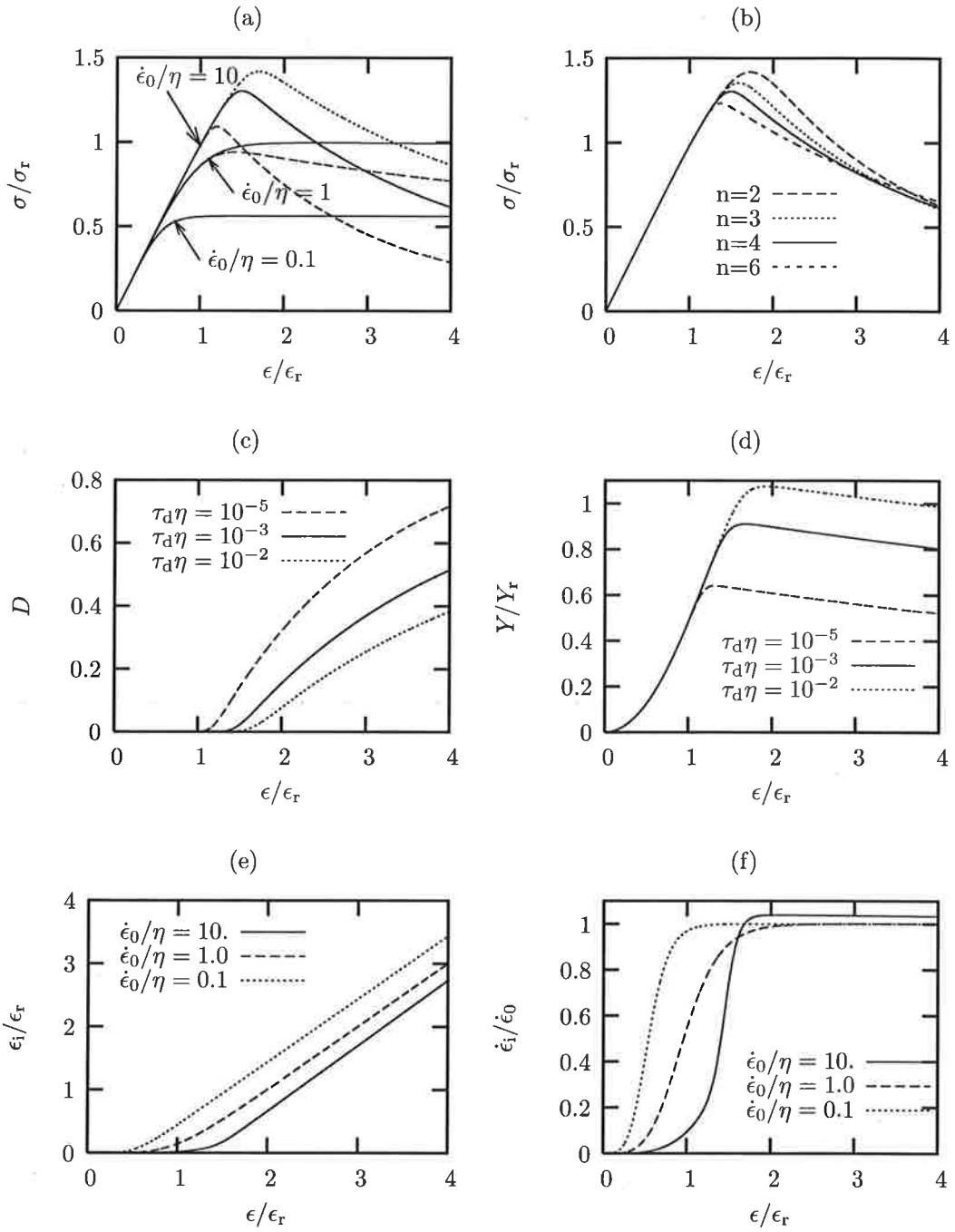


Figure 2: Uniaxial constant strain-rate loading. The line types in (a) equal to those in (c) and (d).

4 ALGORITHMIC TREATMENT

For rate-dependent solids implicit time integrators are preferable. In this study, the backward Euler scheme is used to integrate the constitutive model at the integration point level. Although the backward Euler scheme is asymptotically only first order accurate, it has good accuracy properties for large, practically relevant time-steps [11].

The constitutive model (14)-(16) is rewritten using matrix notation in the form

$$\dot{\boldsymbol{\sigma}} = \mathbf{f}_\sigma(\boldsymbol{\sigma}, \omega), \tag{19}$$

$$\dot{\omega} = f_\omega(\boldsymbol{\sigma}, \omega) \tag{20}$$

such that

$$\mathbf{f}_\sigma(\boldsymbol{\sigma}, \omega) = \omega \mathbf{C}_\mathrm{e}(\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}_\mathrm{i}) + \frac{f_\omega}{\omega}\boldsymbol{\sigma}, \tag{21}$$

$$f_\omega(\boldsymbol{\sigma}, \omega) = -\frac{\varphi_\mathrm{tr}}{\tau_\mathrm{d}\omega}\left(\frac{Y}{Y_\mathrm{r}}\right)^\tau, \tag{22}$$

where

$$\dot{\boldsymbol{\epsilon}}_\mathrm{i} = \left[\frac{\varphi_\mathrm{d}}{(\tau_\mathrm{vp}\eta)^n \omega \sigma_\mathrm{r}}\left(\frac{\bar{\sigma}}{\omega\sigma_\mathrm{r}}\right)^{np-1} + \frac{1}{\tau_\mathrm{vp}\omega}\left(\frac{\bar{\sigma}}{\omega\sigma_\mathrm{r}}\right)^p\right]\frac{\partial\bar{\sigma}}{\partial\boldsymbol{\sigma}}, \tag{23}$$

$$Y = \frac{1}{2\omega^2}\boldsymbol{\sigma} : \mathbf{C}_\mathrm{e}^{-1} : \boldsymbol{\sigma}. \tag{24}$$

Applying the backward Euler scheme and the Newton's linearisation method to the evolution equations (19) and (20) results in the linear system of equations[1]

$$\begin{bmatrix} \mathbf{H}_{11} & \mathbf{h}_{12} \\ \mathbf{h}_{21}^T & H_{22} \end{bmatrix} \begin{Bmatrix} \delta\boldsymbol{\sigma} \\ \delta\omega \end{Bmatrix} = \Delta t \begin{Bmatrix} \mathbf{f}_\sigma \\ f_\omega \end{Bmatrix} - \begin{Bmatrix} \Delta\boldsymbol{\sigma} \\ \Delta\omega \end{Bmatrix}, \tag{25}$$

where

$$\mathbf{H}_{11} = \mathbf{I} - \Delta t \frac{\partial \mathbf{f}_\sigma}{\partial \boldsymbol{\sigma}}, \tag{26}$$

$$\mathbf{h}_{12} = -\Delta t \frac{\partial \mathbf{f}_\sigma}{\partial \omega}, \tag{27}$$

$$\mathbf{h}_{21}^T = -\Delta t \frac{\partial f_\omega}{\partial \boldsymbol{\sigma}}, \tag{28}$$

$$H_{22} = 1 - \Delta t \frac{\partial f_\omega}{\partial \omega}. \tag{29}$$

The algorithmic tangent matrix, i.e. the Jacobian of the algorithmic stress-strain relation has the simple form

$$\mathbf{C} = \omega \widetilde{\mathbf{H}}_{11}^{-1} \mathbf{C}_\mathrm{e}, \tag{30}$$

where

$$\widetilde{\mathbf{H}}_{11} = \mathbf{H}_{11} - \mathbf{h}_{12} H_{22}^{-1} \mathbf{h}_{21}^T. \tag{31}$$

As it can be seen, the Jacobian matrix is in general nonsymmetric due to the damage. The algorithmic tangent matrix is a necessity for the Newton's method to obtain asymptotically quadratic convergence of the global equilibrium iterations.

[1]The symbols Δ and δ refer to incremental and iterative values, $\boldsymbol{\sigma}_n^{i+1} = \boldsymbol{\sigma}_n^i + \delta\boldsymbol{\sigma}_n^i$, $\Delta\boldsymbol{\sigma}_n^i = \boldsymbol{\sigma}_n^i - \boldsymbol{\sigma}_{n-1}$, where the sub- and superscripts refer to step- and iteration numbers, respectively.

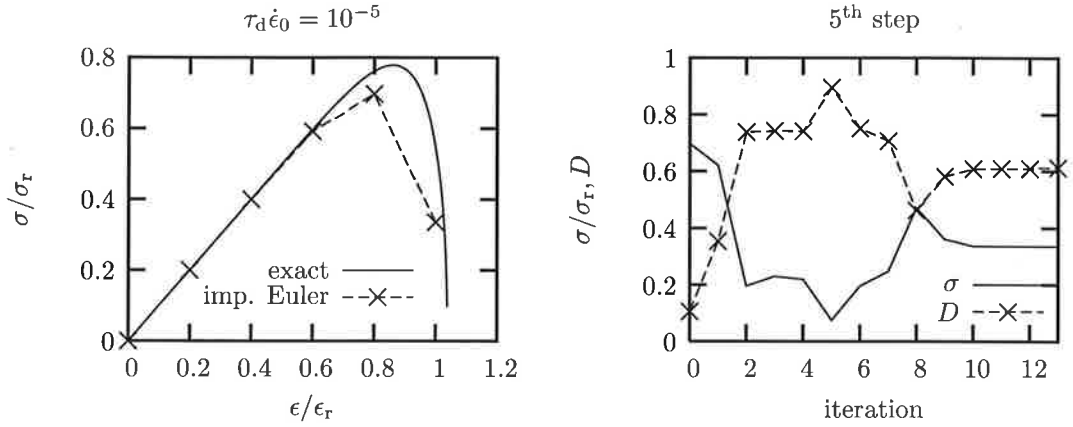


Figure 3: Instability in the implicit backward Euler scheme.

Although the backward Euler scheme is simple and gives accurate results in visco-plastic computations for large time-steps, in the case of damage the scheme can generate oscillations which destroy the convergence of the local Newton's process. This can be seen clearly, when only elastic-damaging material is considered. Applying the backward Euler scheme to a uniaxial case with pure damage evolution

$$\dot{\omega} = f_\omega(\omega) = -\frac{1}{\tau_d \omega}\left(\frac{Y}{Y_r}\right)^r \tag{32}$$

yields the iteration process

$$\omega_n^{i+1} = \omega_n^i + \frac{\Delta t f_\omega(\omega_n^i) - \Delta\omega_n^i}{1 - \Delta t f_{\omega,\omega}(\omega_n^i)}, \tag{33}$$

where the denominator is similar to the element H_{22} in eq. (25). When the denominator is negative, oscillations occur, so that the size of time-step Δt needs to be restricted. The derivative $f_{\omega,\omega}$ is

$$f_{\omega,\omega} = \frac{\partial f_\omega}{\partial \omega} = \frac{1}{\tau_d \omega^2}\left(\frac{Y}{Y_r}\right)^r, \tag{34}$$

and the requirement, $1 - \Delta t f_{\omega,\omega} > 0$, results in the following restriction on the time-step:

$$\frac{\Delta t}{\tau_d} < 2^r \omega^2 \left(\frac{\epsilon_r}{\epsilon}\right)^{2r}. \tag{35}$$

The oscillations are illustrated in fig. 3, in which the analytical stress-strain curve together with the numerical solution using the implicit backward Euler scheme are shown on the left, and the iteration histories of stress and damage for the fifth step ($\epsilon/\epsilon_r \in (0.8, 1.0)$) on the right. The following parameters have been used: $r = 4$, $\tau_d = 10^{-2}$ s, $\dot{\epsilon}_0 = 10^{-3}$ s^{-1} and $Y_r = 0.01$ ($E = 40$ GPa, $\sigma_r = 20$ MPa). Moreover, $\Delta t = 0.1$ s. At the fifth step, the term H_{22} becomes negative and the domain of attraction of the Newton process is reached not until after eight iterations.

5 NUMERICAL EXAMPLE

A compressed specimen ($(x, y, z) \in \Omega = (0, L) \times (0, B) \times (0, H)$, $L = 200$ mm, $B = 100$ mm, $H = 1$ mm) is analysed under plane strain condition, as shown in fig. 4.

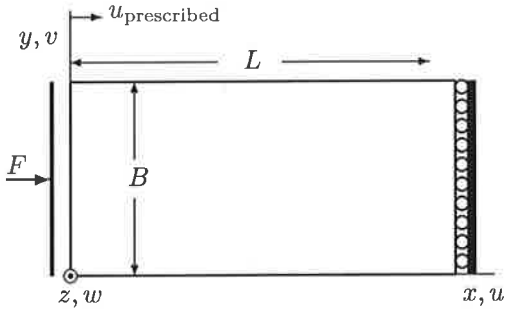


Figure 4: Problem description.

A strain localisation into a shear band is expected to take place due to damage induced strain-softening. The horizontal displacement at the left-hand side edge is prescribed at constant rate $\dot{u}_{\text{prescibed}}$ and constrained to remain straight. A von-Mises type viscoplastic solid is used, i.e. $\bar{\sigma} = \sigma_{\text{eff}}$. The constitutive parameters have the following values: Young's modulus $E = 40$ GPa, Poisson's ratio $\nu = 0.3$, reference stress $\sigma_{\text{r}} = 20$ MPa, the viscoplastic relaxation time $\tau_{\text{vp}} = 1000$ s, the transition strain rate $\eta = 10^{-3}$ s^{-1}. All the exponents p, r and n have the value of 4.

Eight-node-trilinear elements with mean dilatation formulation [12] were used in the computations, which were carried out for two different meshes, a coarse mesh of 12×6 elements and a finer mesh of 48×24 elements. To trigger the unstable localisation, an imperfection via a patch of elements was introduced by reducing the reference stress by 5 %.

Figure 5 shows the load-displacement curves calculated for three different loading rates (on the upper left) and four different damage relaxation times (on the upper right) using the coarse mesh, and for both meshes considering that $\tau_{\text{d}}\eta = 10^{-3}$ and $\dot{\epsilon}_0/\eta = 10$ (at the bottom). The average strain rate is defined as $\dot{\epsilon}_0 = \dot{u}_{\text{prescribed}}/L$. In comparison to the results of pure material behaviour (fig. 2a), the softening behaviour of the structure is much more rapid due to the localisation band.

As explained in the preceding section, a large number of time-step reductions, due to diminished convergence of local iterations, had to be done during the computations, especially in the computations for the highest loading rate.

Damage distribution is shown in fig. 6. It can be observed that damage bands are approximately at ±45° angles as in the classical strain-softening von-Mises type elastoplasticity.

6 CONCLUDING REMARKS

A phenomenological constitutive model for modelling the ductile-to-brittle transition due to increased strain-rate is presented. In the present model, the dissipation potential is additively split into damage and visco-plastic parts and the transition behaviour is obtained using a stress dependent damage potential. In this preliminary study, only isotropic damage and von-Mises type viscoplastic flow are considered. However, the chosen approach allows easily an extension to more advanced damage models. Further investigations will be focused on the study of material length scale.

The numerical implementation is also discussed. Due to the unstable nature of damage,

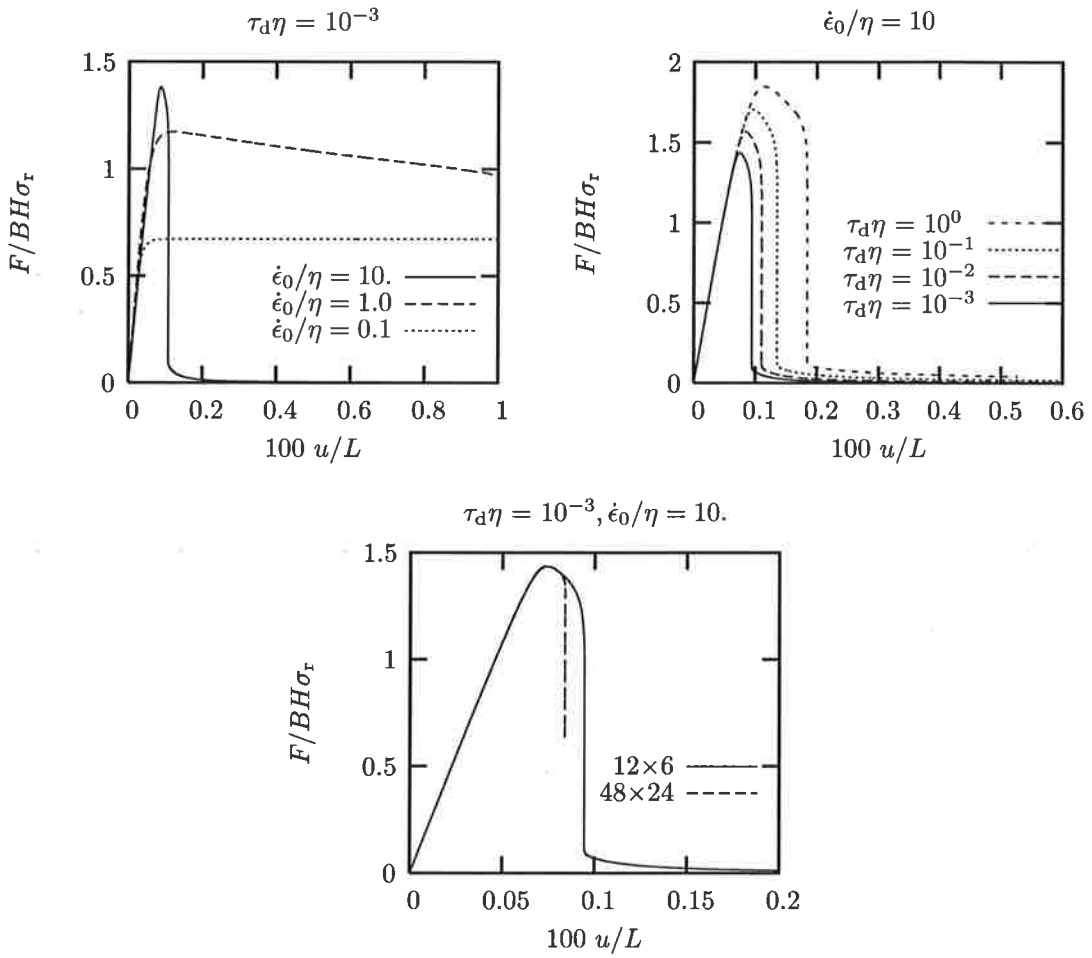


Figure 5: Load-displacement curves. A mesh of 12×6 elements in the upper figures.

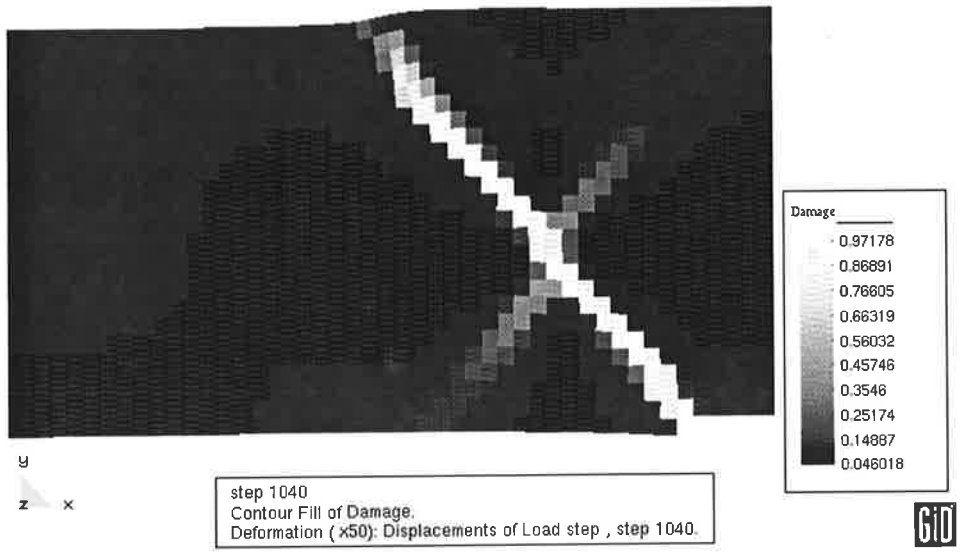


Figure 6: Damage D distribution for $\dot{\epsilon} = 10\eta$ and $\tau_{\mathrm{d}}\eta = 10^{-3}$ at the end of the computation ($F = 0.618BH\sigma_{\mathrm{r}}$). A mesh of 48×24 elements. Displacements magnified by 50 times.

the conventional backward Euler method does not perform well. Oscillations in the damage variable can result in convergence problems in the local Newton iteration at the integration point level. Therefore, further studies will be directed to develop a robust integration scheme for damage evolution.

ACKNOWLEDGMENTS

This research has been supported in part by Tekes - the National Technology Agency of Finland (project KOMASI, decision number 40288/05) and in part by the Academy of Finland, (decision number 210622).

REFERENCES

[1] A. Needleman. *Material rate dependence and mesh sensitivity in localization problems*, Comp. Meth. Appl. Mech. Engng., **67**, 69–85 (1988).

[2] A.C. Eringen. *Microcontinuum Field Theories, I. Foundations and Solids*, Springer, New-York, 1999.

[3] L.J. Sluys. *Wawe propagation, localisation and dispersion in softening solids*, PhD thesis, Department of Civil Engineering, Delft University of Technology, 1992.

[4] W. Wang. *Stationary and propagative instabilities in metals – a computational point of view*, PhD thesis, Department of Civil Engineering, Delft University of Technology, 1997.

[5] M. Frémond. *Non-Smooth Thermomechanics*, Springer, Berlin, 2002.

[6] P. Perzyna. *Fundamental problems in viscoplasticity*, Advances in Applied Mechanics, **9**, 243–377 (1966).

[7] G. Duvaut and L.J. Lions. *Inequalities in Mechanics and Physics*, Springer, Berlin, 1972.

[8] M. Ristinmaa and N.S. Ottosen. *Consequences of dynamic yield surface in viscoplasticity*, Int. J. Solids Structures, **37**, 4601–4622 (2000).

[9] J. Lemaitre, J.-L. Chaboche. *Mechanics of Solid Materials*, Cambridge University Press, 1990.

[10] J. Lemaitre. *A Course on Damage Mechanics*, Springer-Verlag, Berlin, 1992.

[11] R. Kouhia, P. Marjamäki, J. Kivilahti. *On the implicit integration of rate-dependent inelastic constitutive models*, Int. J. Numer. Meth. Engng., **62**, 1832–1856 (2005).

[12] T.J.R. Hughes. *The Finite Element Method - Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall, 1987.

MUISTIMETALLI-IMPLANTTIEN MALLINTAMINEN LÄÄKETIETEELLISISSÄ SOVELLUKSISSA

A. VUOTIKKA [a,*], S.-G. SJÖLIND [a], T. JÄMSÄ [b]

[a] Konetekniikan Osasto
[b] Lääketieteen tekniikan laitos
FIN-90014 Oulun Yliopisto
[*] e-mail: Antti-Jussi.Vuotikka@oulu.fi

TIIVISTELMÄ

NiTi-metalliseoksissa ilmenee ominaisuuksia, joita kutsutaan superelastisuudeksi ja muisti-ilmiöksi. Ne perustuvat materiaalin raetasolla tapahtuvaan termoelastiseen martensiitti-austeniitti -faasimuutokseen. Seoksen yksi tärkeimmistä ominaisuuksista on bioyhteensopivuus, joka mahdollistaa lääketieteellisen käytön.

Laskennallisen työkalun tarpeen kasvaessa materiaalin käyttäytymisen mallintamista varten on kehitetty erilaisia malleja. Laskennallisen mallin kehittämisen eri osa-alueista on tehty kirjallisuustutkimusta. Laboratoriossa on aiemmin tehty kokeita NiTi-langoille ja implanteille. Elementtimenetelmäohjelmistoilla on pyritty kuvaamaan testien tuloksia plastisuus- ja superelastisuusmallien avulla.

1. JOHDANTO

Eräillä metalliseoksilla ilmenee mielenkiintoisia ominaisuuksia, joita kutsutaan superelastisuudeksi ja muisti-ilmiöksi. Kaupallisesti eniten hyödynnetyt muistimetalliseokset pohjautuvat nikkeli-titaani- ja kupari-sinkkiseoksiin. Muistimetalliseosten ominaisuudet perustuvat materiaalin raetasolla tapahtuvaan termoelastiseen martensiitti-austeniitti -faasimuutokseen. Luvussa 2 on selitetty syvällisemmin fysikaaliset perusteet superelastisuudelle sekä muisti-ilmiölle lähteiden [1, 3, 5] mukaan.

Laskennallisen työkalun tarve kasvaa muistimetalleista valmistettavien rakenteiden kehittyessä yhä monimutkaisemmiksi. Vaativin osuus tällaisen työkalun valmistamisessa on materiaalin konstitutiivisen mallin muodostaminen. Materiaalimallien implementointi elementtimenetelmäohjelmistoihin on toinen tärkeä kohta mallien hyödynnettävyyttä ajatellen. Muistimetallimateriaalin käyttäytymisen mallintamisesta on tehty lähteiden [1, 2, 4, 5] pohjalta kirjallisuustutkimusta, jossa esitetään elementtimenetelmän keinot superelastisuuden ja muistimetalliefektin mallintamiseksi.

Lääketieteellisessä käytössä, kun materiaalia käytetään kehon sisällä implanttina, yksi tärkeimmistä ominaisuuksista on bioyhteensopivuus [5]. Muita vaatimuksia ovat hyvät väsymisominaisuudet, kyky tuottaa tasaista voimaa ja luonnonmateriaaleja lähellä olevat mekaaniset ominaisuudet. NiTi-seos pystyy täyttämään nämä vaatimukset hyvin ja siten sitä käytetään lääketieteessä implanttimateriaalina. NiTi-seoksen faasimuutosten vaikutusta bioyhteensopivuuteen on myös tutkittu [6].

Laboratoriossa on aiemmin tehty veto- ja taivutuskokeita NiTi-materiaalista valmistetuille langoille sekä avautumiskokeita lääketieteessä käytetylle implantille [7, 8]. Kyseisen materiaalin

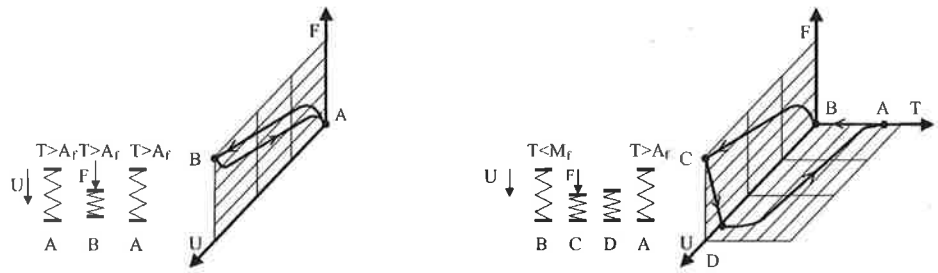
ominaisuuksia on testattu eri lämpötiloissa, joissa metalliseos muuttuu martensiittisesta faasista austeniittiseksi.

Elementtimenetelmäohjelmistolla on edellisissä tutkimuksissa [7, 8] mallinnettu materiaalin superelastisuutta plastisuusteorian avulla käyttäen bilineaarista materiaalimallia, 2D-palkkielementtiä ja suurten siirtymien teoriaa. Kaupalliseen elementtimenetelmäohjelmaan on tullut ensimmäinen materiaalin superelastisuutta kuvaava malli [9]. Tässä tutkimuksessa testataan tämän mallin mahdollisuuksia kuvata tehtyjä materiaalitestejä ja verrataan tuloksia plastisuusteorialla saatuihin tuloksiin.

Seuraavina tavoitteina on mallintaa muistimetallista valmistetun implantin vaikutusta luun kasvuvirheiden korjaamiseen. Laboratoriossamme on jo mallinnettu plastisuusmallien avulla ydinnaulan aiheuttamaa taivutusta luuhun [10]. Tavoitteena on testata valmiin superelastisuusmallin soveltuvuutta kyseiseen ongelmaan sekä mallintaa luun murtumien korjaamiseen käytettävää muistimetalli-implanttia.

2. MUISTI-ILMIÖN JA SUPERELASTISUUDEN FYSIKAALINEN PERUSTE

Jännityksen ja lämpötilan muutokset aiheuttavat metalliseoksissa martensiitti-austeniitti -faasimuutoksia. Faasimuutokset jaetaan yleensä kahteen ryhmään: termoelastisiin ja ei-termoelastisiin muutoksiin. Jälkimmäinen muutos esiintyy useimmissa materiaaleissa, eikä se ole kidetasolla palautuva, kun taas termoelastinen faasimuutos on palautuva. Muistimetalliseoksien tärkeimpien ominaisuuksien toiminta perustuu tähän termoelastisesti palautuvaan faasimuutokseen. Muistimetallimateriaalin muista metallimateriaaleista erottavat perusominaisuudet ovat superelastisuus ja muisti-ilmiö, jotka on esitetty kaavamaisesti kuvassa 1. Jousen liikkeellä on havainnollistettu voima-siirtymä-lämpötilakuvaajia.

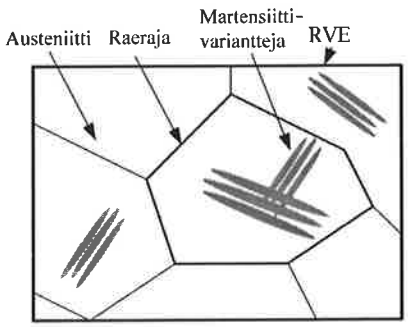


Kuva 1: Vasemmanpuoleisessa kuvassa on esitetty jousen superelastinen käyttäytyminen, ja oikeanpuoleisessa kuvassa on esitetty jousen yksisuuntainen muisti-ilmiö.

Muistimetalliseos on metalliseos, jossa esiintyy martensiitti-austeniitti -faasimuutos termomekaanisten kuormien vaikutuksesta. Tilamuutos jakautuu jännitystilan aiheuttamaan muutokseen, johon perustuu superelastisuus, ja lämpötilan aiheuttamaan muutokseen, joka on muisti-ilmiön perustana. Korkeammissa lämpötiloissa muistimetalliseoksessa esiintyy austeniittinen kantafaasi, jolla on korkea symmetria. Laskettaessa materiaalin lämpötilaa tai kohdistettaessa siihen mekaaninen kuormitus austeniittifaasi alkaa muuttua eri martensiittifaaseiksi, joilla on alhaisempi symmetria kuin kantafaasilla. Martensiitti-austeniitti -faasimuutos on diffuusioton leikkauksen dominoiva faasimuutos, joka ilmenee martensiittifaasin ydintymisenä ja kasvuna austeniittisessa kantafaasissa. Tyypillinen tilavuuselementti (RVE) muistimetalliseoksen mikrorakenteesta on esitettu kuvassa 2.

Muisti-ilmiössä lähdetään liikkeelle korkean symmetrian, kuutiohilaisesta austeniitista. Muistimetalliseoksen käydessä läpi martensiittisen tilamuutoksen se muuttuu alhaisen symmetrian martensiittifaasiin, kuten NiTi-muistimetalleilla monokliinisiksi martensiittifaasien varianteiksi. Kuormituksen puuttuessa martensiittifaasin eri variantit järjestyvät keskenään itsemukautuvalla

tavalla kaksostumalla, mikä ei aiheuta makroskooppisesti havaittavia muutoksia. Kuorman lisääminen pakottaa martensiittifaasin variantit suuntautumaan, mikä johtaa suuriin makroskooppisiin epäelastisiin venymiin. Kun materiaali lämmitetään tietyn lämpötilan yläpuolelle, martensiittifaasi palautuu austeniittifaasiksi ja epäelastiset venymät palautuvat.



Kuva 2: Tyypillinen tilavuuselementti (RVE) muistimetalliseoksen mikrorakenteesta.

Superelastisuus havaitaan, kun martensiitin faasimuutos aiheutetaan mekaanisesti kuormittamalla austeniittifaasia isotermisesti, jolloin suuntautunut martensiittifaasi muodostuu suoraan. Tällöin syntyy suuria epäelastisia venymiä, jotka palautuvat kuormituksen poistamisen yhteydessä käänteisen faasimuutoksen takia. Materiaali kokee suuria venymiä, jotka voivat olla täysin palautuvia jopa 8%:n venymiin saakka.

Materiaaliominaisuudet ovat erilaiset austeniitti- ja martensiittitilassa, joten lämpötila on otettava mukaan materiaalimalliin, kun halutaan kuvata käyttäytymistä eri lämpötiloissa. Tilamuutoksen aikana muistimetallin fysikaaliset ja mekaaniset ominaisuudet eivät pysy samana, jolloin materiaalin termomekaaninen käyttäytyminen muuttuu epälineaariseksi. NiTi-seosten superelastisuus ja palautuvan venymän suuruus riippuvat kuormituksen lisäksi myös voimakkaasti materiaalin valmistusprosesseista ja materiaalin seossuhteista. Muita tyypillisiä NiTi-materiaalin ominaisuuksia ovat kriittisten jännitysten eroavuus puristus- ja vetojännityksen alaisena sekä muutokset jännitys-venymäsuhteessa syklisen kuormituksen vaikutuksesta.

3. MUISTI-ILMIÖN JA SUPERELASTISUUDEN KUVAAMINEN

Muistimetallin martensiitti-austeniitti -faasimuutos vaatii monimutkaisemman materiaalimallin verrattuna perinteisiin metallin kuvaamiseen tarvittaviin materiaalimalleihin. Viimeisen kahdenkymmenen vuoden aikana monikiteisen muistimetalliseoksen konstitutiivista mallintamista on kehitetty eteenpäin huimin askelin. Suurin osa kirjallisuudessa esitetyistä konstitutiivisista malleista voidaan jakaa pääpiirteittäin kahteen ryhmään: fenomenologisiin eli makroskooppisiin malleihin ja mikromekaniikkaan perustuviin malleihin.

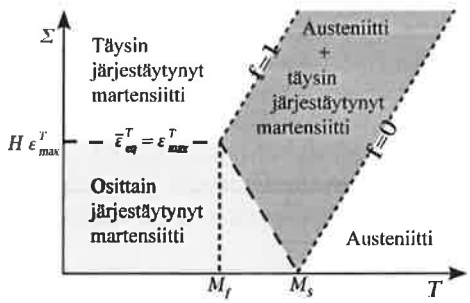
3.1. Makroskooppinen eli fenomenologinen malli

Fenomenologiset mallit monikiteisille muistimetalleille perustuvat kontinuumi- ja termomekaniikkaan, jossa käytetään sisäisiä muuttujia kuvaamaan faasimuutoksien aiheuttamia muutoksia mikrorakenteessa. Sisäisinä muuttujina voidaan käyttää eri martensiittivarianttien tilavuusosuuksia ja epäelastisia venymäkomponentteja. Sisäisille muuttujille tarvitaan myös lisäyhtälöitä, evoluutioyhtälöitä, jotka kuvaavat muuttujien kehitystä kuormitushistorian funktiona. Sisäisillä muuttujilla kuvataan faasimuutoksessa tapahtuvat mikrorakenteen muutokset, ja niiden avulla saadaan huomioitua mikrorakenteen keskimääräinen käyttäytyminen. Toinen vaihtoehto perustuu plastisuuden ja superelastisuuden näennäiseen samankaltaisuuteen. Austeniitti muuttuu martensiitiksi rajajännityksen yläpuolella, ja prosessi loppuu, kun koko materiaali on muuttunut

martensiitiksi. Isotermistä kuormitusta käytettäessä tämä malli on helposti sovitettavissa elementtimenetelmään.

Edellä esitetyt materiaalimallit ovat nopeudesta riippumattomia, joissa jännitys-venymävaste riippuu vain kuormitustiestä, mutta myös nopeudesta riippuvia konstitutiivisia muistimetallimalleja on jonkin verran tutkittu. Fysikaaliseen ilmiöön perustuvat mallit voivat olla hyvin yksinkertaisia, esimerkiksi mallinnettaessa yksiakselista muistimetallin superelastisuutta paloittain jatkuvilla funktioilla, tai erittäin monimutkaisia 3D-malleja, jotka sisältävät huomattavan määrän materiaaliparametrejä.

Fenomenologisen makroskooppisen mallin kehittäminen alkaa jännitys-lämpötila-faasimuutosdiagrammin (kuva 3) ja globaalin makroskooppisen energiafunktion eli kineettisen lain muodostamisesta. Ne kuvaavat metalliseoksen tilamuutoskäyttäytymistä. Ne voidaan muodostaa materiaalitestien tulosten pohjalta, kun testit on tehty useita lämpömekaanisia kuormituksia käyttäen. Toinen vaihtoehto faasinmuutosdiagrammin muodostamiselle on tehdä termodynaamisia analyysejä martensiittimuutokselle. Tämän jälkeen voidaan muodostaa evoluutioyhtälöt sisäisille muuttujille, jolloin saadaan matemaattinen kuvaus ennustamaan martensiitin faasimuutosta mielivaltaisella termomekaanisella kuormituksella. Makroskooppisella energiafunktiolla ja evoluutioyhtälöillä oletetaan olevan jokin tietty funktionaalinen muoto, jonka täytyy toteuttaa termodynamiikan toinen pääsääntö. Kaikkien fenomenologisten mallien perusrakenne on samanlainen, ja sen takia ne voidaan luokitella sisäisten tilamuuttujien konstitutiivisiksi malleiksi. Tuloksena saatavat fenomenologiset mallit eivät suoranaisesti riipu mikroskooppisen tason materiaaliparametreistä vaan makroskooppisen tason parametreistä, jotka pitää selvittää kattavilla materiaalitesteillä ja joilla ei useinkaan ole selvää fysikaalista tulkintaa.



Kuva 3: Jännitys-lämpötila-faasimuutosdiagrammi.

Ensimmäisiä fenomenologisia malleja, jotka pohjautuivat plastisuusmalleihin, on käytetty muistimetallin superelastisuuden kuvaamiseen. Niiden suurin eroavaisuus on lujittumisfunktion valinnassa, jolla mallinnetaan jännitys-venymävastetta jännityksen aiheuttaman martensiitti-austeniitti -faasimuutoksen vaikutuksesta. Tämän jälkeen malliin liitettiin mahdollisuus huomioida epäsymmetrinen vaste, koska muistimetallin käyttäytyminen eroaa vedon ja puristuksen välillä. Aikaisessa vaiheessa sisäisten muuttujien avulla huomioitiin martensiitin kaksostuminen faasidiagrammin pohjalta. Siinä käytettiin kahta sisäistä muuttujaa, jotta martensiitin eri tilavuusosuudet ja martensiittivariantin suuntautuminen voitiin mallintaa erillään. Termodynamiikkaan perustuva malli kaksostumisen muutoksista kehitettiin myöhemmin. Tähän liittyen martensiittifaasin eri variaatioiden vaihtelu kuormituksen vaihdellessa voidaan huomioida esimerkiksi ei-assosiatiivisella myötösäännöllä. Martensiitin tilavuusosuus ja faasimuutosvenymä voidaan käsitellä myös omina sisäisinä muuttujina. Jotkin näistä malleista on johdettu superelastisuuden simuloimista varten ja toiset muisti-ilmiön kuvaamiseen. Mallien kehitys on pääsääntöisesti

lähtenyt yksiulotteisiin ongelmiin, kuten palkkiin tai lankaan, soveltuvista konstitutiivisista malleista, ja aikaa myöten ne on laajennettu useampaan ulottuvuuteen.

Muistimetallien syklisen kuormituksen sekä faasimuutoksessa syntyvän plastisuuden huomioiminen materiaalimallissa on tällä hetkellä muistimetallin mallinnuksessa suurimman mielenkiinnon kohteena maailmalla. Plastisuus voidaan huomioida esimerkiksi siten, että tietty osuus martensiitista ei palaudu syklisen kuormituksen aikana. Tällöin rakenteeseen jää havaittavaa palautumatonta venymää mallinnettaessa. Syklisen kuormituksen jatkuessa plastisuuden osuus pienenee lopulta olemattomaksi. Toisaalta jännityksen aiheuttama tilamuutos voidaan mallintaa myös sallimalla molempien tilamuutos- ja plastisten venymien muodostuminen yhtä aikaa kuormituksen vaikutuksesta.

3.2. Mikromekaniikkaan eli rakeen käyttäytymiseen pohjautuva malli

Toinen haara kirjallisuudessa seuraa mikromekaanista lähestymistapaa. Mikromekaaniset mallit yrittävät ennustaa muistimetallin vastetta ottamalla huomioon monikiteisen muistimetallin raerakenteen. Mikromekaanisen mallin kehittely lähtee siis martensiittivarianttien tasolta. Tässä kokoluokassa transformaatiovenymä määritellään materiaalin kidetasolla. NiTi-muistimetalleilla esiintyy 24 erilaista martensiittivarianttien asentoa. Jokaisen variantin tilavuusjakauma valitaan sisäiseksi muuttujaksi. Vapaan energian funktio selvitetään siten kidetasolla, ja sen avulla esitetään yhden kiteen käyttäytyminen. Useampi kiteisen kappaleen käyttäytyminen saadaan skaalaamalla eri menetelmiä käyttäen.

Jos mikrorakenne on hyvin tunnettu, teoreettisella tasolla on mahdollista käyttää yksikiteisen muistimetallin hyvin tunnettua käyttäytymistä apuna ratkaistaessa reuna-arvo-ongelmaa monikiteiselle muistimetallille. Käytännössä ei ole mahdollista saada tarkkaa esitystä seoksen mikrorakenteesta, jolloin pelkästään huomioimalla seoksen raerakenne ei ole mahdollista ratkaista numeerisia ongelmia kolmessa ulottuvuudessa. Tämän takia tarvitaan homogenisointitekniikoita, jotta saadaan monikiteistä materiaalia edustavat termomekaaniset ominaisuudet esitettyä. Useimmat monikiteisen muistimetallin mikromekaaniset mallit perustuvat self-consistent -tyyppiseen keskiarvoistusmenetelmään, joka on kehitetty heterogeenisten materiaalien elastoplastiselle käyttäytymiselle.

Tämän tyyppiset mikromekaaniset lähestymistavat ovat lupaavia. Mikromekaaninen malli pystyy esittämään superelastisuuden, muistimetalli-ilmiön ja martensiittivariantin uudelleen järjestäytymisen täysin kolmiulotteisesti. Menetelmän haittapuolena on laskennan hitaus, varsinkin käytännön sovelluksissa, koska mallin käsittely vaatii suunnattoman ison laskentatyön. Tämän takia mallin käyttö rajoittuu yleensä superelastisuuden mallintamiseen, varsinkin jos homogenisointitekniikoita ei hyödynnetä.

4. IMPLEMENTOINTI ELEMENTTIMENETELMÄÄN

Konsepteja muistimetallimateriaalin konstitutiivisten mallien kehittämiseksi on siis tutkittu aktiivisesti. Konstitutiivisten mallien kehittäminen on kuitenkin jättänyt taka-alalle näiden mallien numeerisen implementoinnin elementtimenetelmään. Tärkeimpiä kohtia numeerisessa implementoinnissa on aikaintegroinnin tehokkuus, jota tarvitaan sisäisten muuttujien, kuten muutosvenymän ja martensiitin tilavuusjakauman, laskemiseen. Myöskään sopiva elementtiteknologia ei ole ollut kiinnostuksen keskipisteessä. Onnistunut simulaatio riippuu usein tehokkaasta rakenteen käyttäytymisen mallintamisesta kuin pikkutarkasta materiaalin yksityiskohtaisesta käyttäytymisen mallintamisesta. Varsinkin fenomenologisilla malleilla voidaan joustavasti tehdä kompromisseja tarkkuuden ja monimutkaisuuden välillä, kunhan mallille on tehty tehokas implementointi elementtimenetelmään.

Fenomenologisten mallien numeerinen implementointi on ollut kohtuullisen aktiivisen tutkimuksen alla, koska laskennallisilla menetelmillä on tässä tapauksessa juuret plastisuudessa käytettävissä algoritmeissa. Muistimetallin monivivahteinen käyttäytyminen vaatii kuitenkin ilmiöön paremmin soveltuvien algoritmien kehittämistä.

Muistimetallin materiaalikäyttäytymisen sovituksessa käytetään tyypillisesti return mapping -algoritmeja. Koska ne ovat venymäohjattuja, voidaan niitä soveltaa suoraan siirtymiin perustuvaan elementtimenetelmään. Muistimetallin materiaalikäyttäytyminen riippuu kuormitushistoriasta, jolloin se voidaan määrittää jännitys-venymäsuhteella. Jännitys-venymäsuhteeseen voidaan sisällyttää tilamuutosvenymä sisäisenä muuttujana, differentiaalinen evoluutioyhtälö tilamuutosvenymälle ja muunnosyhtälö määrittämään faasimuunnoksen syntymistä. Myös muita numeerisia algoritmeja, jotka on kehitetty venymänopeudesta riippumattomaan elastisplastiseen materiaalikäyttäytymisen mallintamiseen, voidaan käyttää muodostaessa muistimetallin termomekaanista mallia.

Return mapping -algoritmeja on tutkittu ja kehitetty laajasti viimeisten vuosien aikana muodostettaessa konstitutiivisia elastoplastisia yhtälöitä. Niitä kutsutaan myös elastinen ennustaja–plastinen korjaaja -algoritmeiksi. Algoritmi aloittaa täysin elastisella yriteosalla, jota seuraa plastinen korjaajavaihe. Korjaajavaiheen tarkoitus on iteratiivisesti toteuttaa myötöehto. Algoritmit eroavat toisistaan lähinnä evoluutioyhtälöiden numeerisessa muodostamisessa käytettyjen diskretisointien sekä käytetyn iteratiivisen numeerisen proseduurin ratkaisusta korjaajaosassa syntyneiden yhtälöiden osalta. Näitä algoritmeja ovat esimerkiksi backward Euleriin pohjautuva radial return -algoritmi ja keskipistesääntöön perustuva keskinormaali algoritmi.

Suurimmassa osassa näistä algoritmesta käytetään integrointimenetelmänä trapetsi- ja keskipistesäännön erikoistapauksia sopivasti yleistettyinä plastisen konsistenssiehdon toteutumisen helpottamiseksi. Integrointimenetelmien stabiilius epäelastisissa ongelmissa on osoitettu näille algoritmeille. Eksplisiittisiä menetelmiä käytettäessä rajoitetaan aika-askeleen valintaa stabiilisuusongelmien takia, kun taas implisiittisissä menetelmissä inkrementaalisen ratkaisun stabiiliuus on parempi. Implisiittisissä menetelmissä ylimääräinen tarkkuus saadaan keskipiste-integroinnilla sekä epäelastisen venymän korkeampiasteisella approksimaatiolla, esimerkiksi Runge-Kutta -menetelmillä.

5. BIOYHTEENSOPIVUUS

Bioyhteensopivuudella tarkoitetaan materiaalin kykyä tulla toimeen elävän kudoksen kanssa. Fysiologinen väliaine on erittäin korrodoiva, joten metalliseoksista valmistettujen implanttien ongelmana on metalli-ionien liukeneminen ja niiden vaikutus kudoksiin. Useat tutkimukset ovat osoittaneet NiTi-materiaalien olevan hyvin bioyhteensopivia. Materiaalin korkea nikkelipitoisuus ja korroosion aihettama liukeneminen ovat kuitenkin tarkemman kiinnostuksen kohteena. Tutkimuksen alla on implanttimateriaalin pinnan karkeuden, kemiallisen koostumuksen ja suojaavan pinnoitteen vaikutus bioyhteensopivuuteen.

Faasimuutoksen vaikutusta bioyhteensopivuuteen on myös tutkittu. Yleisesti tiedetään, että useamman eri faasin läsnäolo samassa seoksessa heikentää materiaalin korroosion kestävyyttä. NiTi-pohjaisissa muistimetalleissa jännityksen aiheuttaman martensiittifaasin on todettu heikentävän materiaalin bioyhteensopivuutta. Heikentymisen voimakkuus riippuu martensiitin ominaisuuksista ja rakenteen jäännösjännityksistä kantafaasissa.

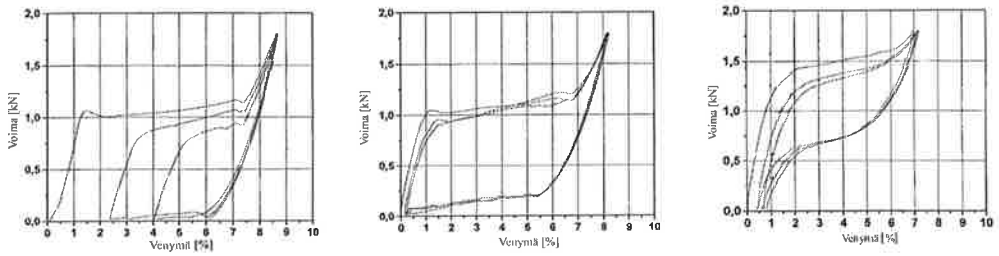
6. MATERIAALITESTIT

Laboratoriossamme on tehty useita erilaisia testejä NiTi-pohjaisille muistimetalliseoksille, joista osa on esitetty tässä työssä. Materiaalitestejä tehtiin erilaisille muistimetallilangoille sekä muistimetallilangasta valmistetuille implanteille. Kokeet on tehty Instron-aineenkoetuslaitteella tilassa, jonka lämpötilaa pystyttiin kontrolloimaan 0 – 40 °C:een välillä. 0 °C:een lämpötilassa materiaali on täysin martensiittinen ja 40 °C:een lämpötilassa materiaali on kuormittamattomana austeniittinen. Voiman, venymän ja siirtymän mittaus tapahtui Insronilla. Koekappaleen lämpötilaa mitattiin koekappaleeseen kiinnitetyllä lämpöanturilla.

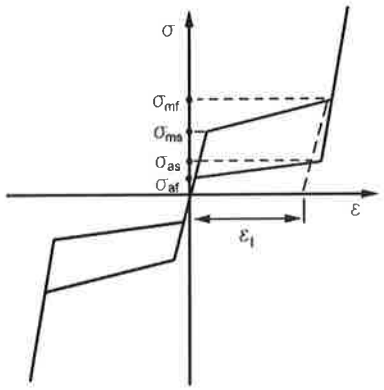
NiTi-langalle, jonka paksuus on 1.8 mm:ä, on tehty vetokuormitussyklejä 0 °C:een, 23 °C:een ja 40 °C:een lämpötiloissa. Syklien määrä rajattiin kolmeen kappaleeseen. Lanka kiinnitettiin siten, että voima-venymäyhteys voitiin mitata koko testin ajan. Vetokokeista saadut testitulokset

löytyvät kuvasta 4 siten, että vasemmalla esitetyssä kuvassa on käytetty testauslämpötilana 0 °C:tta, keskellä 23 °C:tta ja oikealla 40 °C:tta. Kuvista voidaan päätellä, että langan ollessa martensiittinen muodonmuutos ei ole täysin palautuva. Lämpötilan noustessa materiaalista tulee täysin superelastinen. Samalla voidaan havaita, että hystereesi pienenee lämpötilan kasvaessa ja kriittiset jännitykset kasvavat. Näistä kuvista voidaan poimia mallinnuksessa tarvittavat arvot taulukkoon 1. Taulukossa esitetyt arvot on selvitetty kunkin lämpötilan ensimmäisen kuormitussyklin voima-venymäkäyrästä. Tarvittavia ominaisuuksia ovat kyseisen muistimetalliseoksen kimmokertoimen arvo vetotapauksessa E_V, kriittiset jännitykset faasimuutosten eri vaiheille σ_{ms}, σ_{mf}, σ_{as}, σ_{af} sekä tilamuutosvenymän ε_l arvo. Kriittiset jännitykset ja tilamuutosvenymä on määritelty kuvassa 5.

Vastaavalle NiTi-langalle tehtiin myös kolmipistetaivutuskokeita. Kyseiset testit tehtiin samoissa lämpötiloissa kuin langan vetokokeetkin. Koejärjestely oli kuitenkin sellainen, että sillä saatiin vain voima-venymäyhteys voimaa kasvatettaessa. Tällöin hystereesiä ei saada esiin. Tämän testin avulla saatiin kimmokertoimen arvo taivutustapauksessa E_{VP}, joka eroaa vetokimmokertoimesta, koska materiaalin veto- ja puristuskäyttäytyminen eroavat toisistaan. Kimmokertoimen arvot eri lämpötiloissa on esitetty taulukossa 1.



Kuva 4: Muistimetallilangan vetokokeista saadut kolmen kuormitussyklin voima-venymäkäyrät 0 °C:een, 23 °C:een ja 40 °C:een lämpötiloille.



Kuva 5: Kuvassa on määritelty Ansyksen superelastisuusmallin tarvitsemat kriittiset jännitykset faasimuutosten eri vaiheille σ_{ms}, σ_{mf}, σ_{as}, σ_{af} sekä tilamuutosvenymä ε_l.

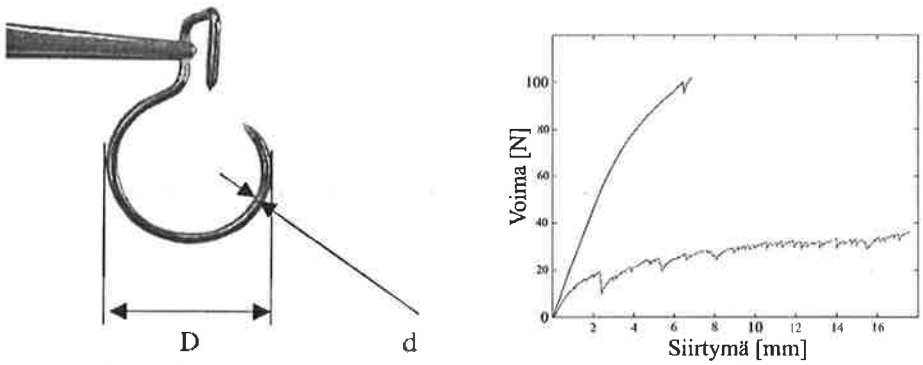
Taulukossa 1 on esitetty myös likimääräinen tangenttimoduli E_T, jota tarvitaan plastisuusmallissa. Faasimuutosten kriittiset arvot on esitetty jännityksinä, joka saadaan mittauksissa käytetystä voimasta jakamalla langan pinta-alalla.

NiTi-langasta valmistetulle koukkuimplantille on myös tehty materiaalikokeita. Implantti on esitetty kuvassa 6. Implantti on valmistettu eri materiaalista kuin edellä testattu NiTi-lanka. Koukkuimplantti kiinnitettiin yläosastaan vetokoneeseen, ja kokeen avaava voima kohdistui koukun alimpaan kohtaan. Yläosa kiinnitettiin jäykästi kuvassa näkyvien pinsettien kohdalta, joten

implantin muodonmuutos syntyi pelkästään koukun avautumisesta. Koukun avautumistesti tehtiin ensin 0 °C:een lämpötilassa, minkä jälkeen koukku lämmitettiin 37 °C:een muodonmuutoksen poistamiseksi. Tämän jälkeen koukulle tehtiin uusi avautumistesti 37 °C:een lämpötilassa. Implantin koukkuosan halkaisija D on 22 mm:ä, ja implantin langan halkaisija d on 2.3 mm:ä. Kokeen tulokset on esitetty kuvassa 6. Kuvasta nähdään koukun jäykkyyden kasvavan huomattavasti lämpötilan muuttuessa materiaalin austeniittiselle alueelle. Käyrien epätasaisuudet johtuvat kitkan vaikutuksesta koukun ja vetotapin välissä.

Ominaisuus	0 °C	23 °C	40 °C
E_{VP} [MPa]	48 000	67 000	71 000
E_V [MPa]	37 960	39 660	40 110
E_T [MPa]	1 480	1 510	2 640
σ_{ms} [MPa]	380	390	520
σ_{mf} [MPa]	510	570	688
σ_{as} [MPa]	78	197	354
σ_{af} [MPa]	-	39	197
ε_l [%]	6	5.5	5

Taulukko 1: Materiaalitesteistä poimitut arvot mallinnuksessa tarvittaville parametreille.



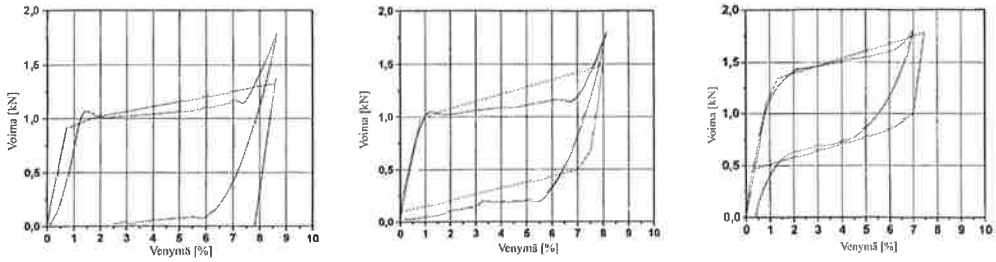
Kuva 6: Muistimetallilangasta valmistettu koukku ja sen avautumiskokeista saatu voima-siirtymäkäyrät.

7. MALLINNUS

Materiaalitesteistä saatuja tuloksia verrattiin staattisella FEM-analyysillä saatuihin tuloksiin. FEM-analyysit on tehty ANSYS-ohjelmistolla. Muistimetallilangan vetokokeista saatuja materiaaliarvoja käyttäen on tehty superelastisuusmallilla mallinnus langan vetokokeesta. Superelastisuusmallilla saatuja voima-venymäkäyriä on verrattu materiaalitestien tuloksiin kuvassa 7. Superelastisuusmallin käyttö vaatii solidielementtien käyttöä, joten langasta tehtiin solidimalli. Elementiksi valittiin SOLID186 tetraedri-muoto. Materiaaliarvot valittiin taulukon 1 mukaisesti ja 0 °C:een σ_{af}:n arvoksi valittiin pieni positiivinen luku. Laskenta suoritettiin täydellä Newton-Raphson-menetelmällä.

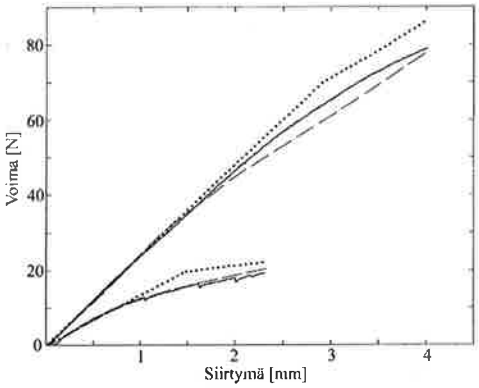
Kuvista nähdään, että superelastisuusmallilla voidaan hyvin kuvata yksinkertaisen rakenteen superelastisuuskäyttäytymistä. Pienellä kriittisten jännitysten arvojen ja tilanmuutosvenymän säätämisellä päästäisiin vielä hieman parempiin tuloksiin. Ainoa ongelma mallissa on se, että siinä ei voi määritellä superelastisuusalueen jälkeistä kimmokerrointa. Tämän vaikutuksen huomaa siinä, että voiman pienentyessä voima-venymäkäyrien välille tulee selvää eroavaisuutta, jota ei voida välttää. Toinen mallin ongelma on siinä, että rakenteeseen ei voi jäädä venymiä, mikäli

kuorma poistuu. Matalimman lämpötilan mallista saadussa voima-venymäkäyrässä jäännösvenymä saatiin aikaan siten, että kriittiset jännitykset valittiin pieniksi ja kuormitusta ei poistettu kokonaan laskennan lopussa.



Kuva 7: Voima-venymäkuvaajat muistimetallilangan vetokokeiden sekä superelastisuus-mallilla saatujen tulosten mukaisesti 0 °C:een, 23 °C:een ja 40 °C:een lämpötiloille.

Koukkuimplantin mallinnus tehtiin plastisuusteorian bilineaarisella materiaalimallilla sekä superelastisuusmallilla. Koukku mallinnetaan ympyrän kaarena, jonka toinen pää on kiinnitetty jäykästi. Koukkua avaava voima saadaan liittämällä koukun alimpaan kohtaan pistevoima. Plastisuusmallissa käytettiin 2D-palkkielementtiä suurten siirtymien teorian mukaan, koska langan paksuus on hyvin pieni rakenteen muihin dimensioihin nähden. Epälineaarisen mallin laskenta-aika on ollut myös kriittinen tekijä plastisuusmallilla mallinnettaessa. Elementiksi on valittu plastinen 2D-elementti BEAM23, jolla saadaan superelastisuuden alkuosa mallinnettua. NiTi-materiaalin jännitys-venymäyhteyden kuvaamiseen superelastisella alueella käytetään bilineaarista kinemaattista lujittumiskriteeriä. Parempi valinta olisi anisotrooppinen muokkauslujittumiskriteeri, mutta tarkkoja puristuspuolen materiaaliarvoja ei ole käytettävissä. Laskenta suoritetaan täydellä Newton-Raphson -menetelmällä. Superelastisuusmallia käytettiin kuten langan vetokoetta mallinnettaessa. Mallinnustulokset ja mittaustulokset on esitetty kuvassa 8.



Kuva 8: Koukkuimplantin vetokokeiden tulokset sekä plastisuus- ja superelastisuusmallien vastaavat tulokset. Plastisuusmallin tulos on pisteviivalla ja superelastisuusmallin tulokset on katkoviivalla.

Tuloksia vertaillessa huomataan, että implantin avautumiskokeessa voima muuttaa paikkansa koukun avautuessa. Laskentamallissa voima pysyy siinä solmussa, mihin se on alun perin määritelty. Tästä aiheutuu virhettä muodonmuutoksen kasvaessa. Toinen ongelmakohta on siinä, että edellä tehdyt materiaalitestit on tehty eri materiaalille, kuin mistä implantti on valmistettu. Avautumiskokeiden mittaustuloksista on vaikea selvittää mallinnuksessa tarvittavat materiaaliarvot, sillä implantin geometria vaikeuttaa voima-siirtymäkäyrän tulkitsemista.

8. YHTEENVETO

Ansys-ohjelmistossa oleva superelastisuusmalli osoittautui pienistä puutteistaan huolimatta kohtuullisen hyväksi malliksi superelastisuuden kuvaamiseen. Aiemmin superelastisuutta kuvaamaan käytetyn plastisuusmallin se voittaa helposti, koska sillä voidaan kuvata voima-venymäsuhdetta myös kuormitusta pienennettäessä. Suurimpina ongelmina oli jäännösvenymän puute ja superelastisuusalueen jälkeisen kimmokertoimen valintamahdollisuuden puute. Jäännösvenymän puute johtuu siitä, että Ansyksen malli sisälsi vain yhden sisäisen muuttujan, joka kuvasi martensiitti-austeniitti -faasimuutosta. Mallin yksinkertaisuuden huomasi siinä, että tässä työssä käytetyissä implanttirakenteissa laskenta-aika pysyi lyhyehkönä.

Seuraavana vaiheena on muisti-ilmiön ohjelmointi elementtiohjelmistoon. Sillä voidaan tuoda mallintamiseen mukaan myös lämpötilan aiheuttamat muodonmuutokset. Tätä tarvitaan implantin mallinnuksessa, kun koukku ensin avataan kylmässä ja asetetaan kehoon. Tämän jälkeen koukun lämpötilan kohoaminen kehon lämpöiseksi poistaa koukun muodonmuutokset.

LÄHTEET

[1] B.Peultier, T.Zineb, E.Patoor. *Macroscopic constitutive law of shape memory alloy thermomechanical behaviour*. Application to structure computation by FEM. Mechanics of Materials, 38, 2006, 510-524.

[2] S.Reese and D.Christ. *FEM optimization with Nitinol*. Min Invas Ther & Allied Technol, 13(4), 2004, 228-239.

[3] E.Patoor, D.Lagoudas, P.Entchev, L.Brinson, X.Gao. *Shape memory alloys, Part I: General properties and modelling of single crystals*. Mechanics of Materials, 38, 2006, 391-429

[4] D.Lagoudas, P.Entchev, P.Popov, E. Patoor, L.Brinson, X.Gao. *Shape memory alloys, Part II: Modeling of polycrystals*. Mechanics of Materials, 38, 2006, 430-462.

[5] Jussi Anttonen. Diplomityön käsikirjoitus, 2005

[6] A.Danilov, A.Kapanen, S.Kujala, J.Saaranen, J.Ryhänen, A.Pramila, T.Jämsä and J.Tuukkanen. *Biocompatibility of austenite and martensite phases in NiTi-based alloys*. Journal de Physique IV, 112, 2003, 1117-1120.

[7] Jouni Märsylä. Diplomityö, 1999.

[8] J.Ryhänen, J.Märsylä, E.Niemelä, J.Tuukkanen, T.Jämsä, A.Pramila. *Acromio-clavicular dislocation repair with NiTi AC-hook implant: Clinical pilot study, AFM and FEM analysis of the implant*. SMST Europe, Antwerp. (1999)

[9] Ansys Inc. *Theory Reference*. Release 10.0.

[10] S.Kujala, J.Tuukkanen, T.Jämsä, A.Danilov, A.Pramila, J.Ryhänen. *Comparison of the bone modelling effects caused by curved and straight nickel-titanium intramedullary nails*. Journal of Material Science: Materials in Medicine, 13, 2002, 1157-1161.

A COUPLED MAGNETOELASTIC MODEL FOR FERROMAGNETIC MATERIALS

ANOUAR BELAHCEN[1], KATARZYNA FONTEYN[1], STEFANIA FORTINO[2], REIJO KOUHIA[2]

[1] Laboratory of Electromechanics, Helsinki University of Technology
P.O. Box 3000, 02015 HUT, Finland
e-mail: anouar.belahcen@tkk.fi, katarzyna.fonteyn@tkk.fi

[2]Laboratory of Structural Mechanics, Helsinki University of Technology,
P.O. Box 2100, 02015 HUT, Finland
e-mail: stefania.fortino@tkk.fi, reijo.kouhia@tkk.fi

ABSTRACT

This paper presents a coupled magnetoelastic model for isotropic ferromagnetic materials used in electrical machines. As proposed by Dorfmann *et al.* for general nonlinear magnetoelastic solids, the constitutive equations of the model are written on the basis of the Helmholtz free energy for which the strain tensor and the magnetic induction vector are chosen as the basic variables. As a result of the method, a suitable form of the free energy is chosen by comparing the proposed constitutive relations and the corresponding experimental data under several external magnetic fields and pre-stresses.

1 INTRODUCTION

The ferromagnetic materials can show the well known phenomenon of magnetostriction and are characterized by a coupled magnetomechanical field. In the traditional models for magnetostriction the constitutive equations of the material (linear or nonlinear) are decoupled [1].

Nowadays the so–called magnetoelastic coupling is widely used for modeling the reciprocal effect between the magnetic and the elastic field. In particular, the linear elastic behaviour is usually considered, while the magnetic properties can be linear or nonlinear. Very simple magnetoelastic models are defined by using the magnetic forces as loads for the elastic field (weak coupling, see refs in [2]). In these coupling models the magnetic equations and the mechanical equations are solved separately. More accurate models, based on the so–called strong coupling ([3], [4], [5]), solve simultaneously the governing equations of the problem in the following coupled cases: *i*) linear magnetic - linear elastic fields, *ii*) nonlinear magnetic - linear elastic fields and *iii*) nonlinear magnetic - nonlinear elastic fields. The recent literature concerning the development of coupling magnetomechanical methods, pointed out that there is still a lack of both theoretical and experimental work in the development of constitutive relations of ferromagnetic materials (see [6] and [7]). As observed by Belahcen in [2], the knowledge of the coupled constitutive equations in general magnetostrictive materials is not possible without measurements needed to provide the required material parameters.

In this paper, starting from the model proposed by Dorfmann *et al.* ([8], [9],[10]) for general nonlinear magnetoelastic solids, the constitutive equations of isotropic ferromagnetic materials are written on the basis of the Helmholtz free energy. The strain tensor and the magnetic induction vector are chosen as the basic variables. Following Dorfmann *et al.*, since the Cauchy stress tensor is not in general symmetric, the so–called total stress tensor, symmetric and defined as a function of the Cauchy stress tensor and of the magnetic field vector, is introduced. In the general case of isotropic magnetoelastic solids, the Helmholtz free energy depends on the six invariants, forming the integrity

basis of an isotropic tensor function depending on a symmetric second order tensor and a vector. In this work a suitable form of the Helmholtz free energy for isotropic ferromagnetic materials used in electrical machines is then proposed by fitting the experimental data obtained by means of a simple but sufficiently accurate measurement device. The measurement set-up enables measurements of both magnetization and magnetostriction as functions of the externally applied stress. As numerical results, the magnetostriction curves for uniaxial problems under several external magnetic fields and pre-stresses are shown.

2 Governing equations for ferromagnetic materials

Let us consider a general electromagnetic body of domain Ω and boundary S subjected to body forces $\mathbf{f}$ in Ω and to surface forces $\mathbf{t}$ on S. Following Maugin [1], the mechanical balance laws of the body are:

$$\dot{\rho} + \rho \operatorname{div} \dot{\mathbf{u}} = 0 \tag{1}$$

$$\rho \ddot{\mathbf{u}} = \operatorname{div} \boldsymbol{\sigma} + \rho \mathbf{f} + \mathbf{f}_{em} \tag{2}$$

where ρ is the mass density, $\mathbf{u}$ the displacement vector, $(\dot{\cdot})$ represents the total time derivative operator $(\frac{\mathrm{d}}{\mathrm{d}t})$, $\boldsymbol{\sigma}$ is the Cauchy stress tensor and $\mathbf{f}_{em}$ the electromagnetic force (per unit volume). Equations (1) and (2) represent the conservation of mass and the balance of linear momentum equation, respectively.

The integral form of the energy balance equation in the isothermal case is written as

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \left(\frac{1}{2} \rho \dot{\mathbf{u}} \cdot \dot{\mathbf{u}} + \rho U \right) \mathrm{d}\Omega = \int_\Omega (\dot{\Phi} + \mathbf{f} \cdot \dot{\mathbf{u}}) \, \mathrm{d}\Omega + \int_S \mathbf{t} \cdot \dot{\mathbf{u}} \, \mathrm{dS} \tag{3}$$

where U represents the internal energy density per unit mass and Φ is the electromagnetic energy density. In particular:

$$\Phi = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H} \tag{4}$$

where $\mathbf{E}$ is the electric field vector, $\mathbf{D}$ the electric displacement field vector, $\mathbf{B}$ the magnetic induction field vector and $\mathbf{H}$ the magnetic field intensity vector. The vector fields $\mathbf{E}, \mathbf{D}, \mathbf{B}$ and $\mathbf{H}$ obey to the Maxwell equations for general electromagnetic bodies (see [1]). Furthermore, $\mathbf{B}$ and $\mathbf{H}$ are related through

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \tag{5}$$

where μ_0 is the magnetic permeability in the vacuum and $\mathbf{M}$ represents the magnetization. By using the mechanical balance laws (see Maugin [1]), equation (3) can be simplified as follows:

$$\int_\Omega \rho \dot{U} \, \mathrm{d}\Omega = \int_\Omega (\dot{\Phi} + \dot{\boldsymbol{\varepsilon}} : \boldsymbol{\sigma}) \, d\Omega \tag{6}$$

where $\boldsymbol{\varepsilon}$ is the infinitesimal strain tensor.

2.1 Case of quasi–magnetostatic of insulators

Let us now consider the assumption of quasi–magnetostatic of insulators (see [1]), which reduces the Maxwell equations for general electromagnetic bodies to

$$\operatorname{div} \mathbf{B} = 0 \tag{7}$$

$$\operatorname{curl} \mathbf{H} = \mathbf{0} \tag{8}$$

In particular, (7) describes the conservation of magnetic flux and (8) the Ampère law in the stationary case and null current density.

The assumption of quasi–magnetostatic of insulators reduces the balance equation (6) to

$$\int_\Omega \rho \dot{U} \, d\Omega = \int_\Omega (\dot{\mathbf{B}} \cdot \mathbf{H} + \dot{\boldsymbol{\varepsilon}} : \boldsymbol{\sigma}) \, d\Omega \tag{9}$$

Following Dorfmann *et al.*, let us further assume the existence of a Helmholtz free energy $\psi = \psi(\boldsymbol{\varepsilon}, \mathbf{B})$ for which the strain tensor $\boldsymbol{\varepsilon}$ and the magnetic induction vector $\mathbf{B}$ are chosen as the basic variables. In particular, in the case of isotropic ferromagnetic materials, ψ depends on the set of invariants

$$I_1 = \operatorname{tr}\boldsymbol{\varepsilon},\ I_2 = \operatorname{tr}\boldsymbol{\varepsilon}^2,\ I_4 = \mathbf{B}\cdot\mathbf{B},\ I_5 = \mathbf{B}\cdot\boldsymbol{\varepsilon}\cdot\mathbf{B},\ I_6 = \mathbf{B}\cdot\boldsymbol{\varepsilon}^2\cdot\mathbf{B} \tag{10}$$

By using the Clausius-Duhem inequality (see [1]), the constitutive equations of the material are expressed as

$$\boldsymbol{\sigma} = \rho\frac{\partial\psi}{\partial\boldsymbol{\varepsilon}} \tag{11}$$

$$\mathbf{M} = -\rho\frac{\partial\psi}{\partial\mathbf{B}} \tag{12}$$

It is worth to note that the Cauchy tensor $\boldsymbol{\sigma}$ is symmetric only when the magnetization $\mathbf{M}$ is everywhere parallel to $\mathbf{B}$ (see [1], [8], [9], [10]). This is the case of the linear isotropic bodies, for which equation (12) reduces to

$$\mathbf{M} = \gamma\mathbf{B} \tag{13}$$

with $\gamma = (\mu - 1)/(\mu\mu_0)$ where $\mu \geq 1$ represents the magnetic permeability. The proportionality between $\mathbf{M}$ and $\mathbf{B}$ implies the independency of the free energy on the invariants I_5 and I_6. Actually in real ferromagnetic bodies, the magnetic permeability of the material depends not only on the magnetic field strength but also on the mechanical stress, i.e. on the solution of the elastic field. Then, in general cases, γ can be defined as a function of the invariants I_4, I_5 and I_6. Furthermore, to avoid the use of a non-symmetric stress tensor, the so–called total stress tensor is defined:

$$\boldsymbol{\tau} = \boldsymbol{\sigma} + \mu_0^{-1}\left[\mathbf{BB} - \frac{1}{2}(\mathbf{B}\cdot\mathbf{B})\mathbf{I}\right] + (\mathbf{M}\cdot\mathbf{B})\mathbf{I} - \mathbf{BM} \tag{14}$$

where the dyad $\mathbf{BB}$ represents the open product, expressed in the component form as B_iB_j. By using (5) and in the absence of mechanical volume forces, equation (2) becomes:

$$\operatorname{div}\boldsymbol{\tau} = \mathbf{0} \tag{15}$$

This expression is obtained by incorporating the electromagnetic force $\mathbf{f}_{em}$ into the Cauchy stress tensor as a Maxwell stress (see [10]).

3 Proposed form of the Helmholtz free energy

Let us write the Helmholtz free energy ψ in terms of the invariants I_1, I_2, I_4, I_5 and I_6 in the following form:

$$\rho\psi = \frac{1}{2}\lambda I_1^2 + \mu I_2 + \frac{1}{2}\gamma_4 I_4 + \frac{1}{2}\gamma_5 I_5 + \frac{1}{2}\gamma_6 I_6 \tag{16}$$

where λ and μ represent the Lamé coefficients, the coefficient γ_4 is a polynomial expansion of the invariant I_4

$$\gamma_4 = \gamma_4^{(0)} + \frac{1}{2}\gamma_4^{(1)}I_4 + \frac{1}{3}\gamma_4^{(2)}I_4^2 + \frac{1}{4}\gamma_4^{(3)}I_4^3 + \frac{1}{5}\gamma_4^{(4)}I_4^4 + \ldots \tag{17}$$

while the coefficients γ_5 and γ_6 are constants.

The constitutive equations of the model are then:

$$\boldsymbol{\sigma} = \lambda I_1\mathbf{I} + 2\mu\boldsymbol{\varepsilon} + \frac{1}{2}\gamma_5\,\mathbf{B}\,\mathbf{B} + \frac{1}{2}\gamma_6(\mathbf{B}\,\boldsymbol{\varepsilon}\cdot\mathbf{B} + \mathbf{B}\cdot\boldsymbol{\varepsilon}\,\mathbf{B}) \tag{18}$$

$$\mathbf{M} = -\gamma_4'\mathbf{B} - \gamma_5\,\boldsymbol{\varepsilon}\cdot\mathbf{B} - \gamma_6\boldsymbol{\varepsilon}^2\cdot\mathbf{B} \tag{19}$$

where

$$\gamma_4' = \gamma_4^{(0)} + \gamma_4^{(1)}I_4 + \gamma_4^{(2)}I_4^2 + \gamma_4^{(3)}I_4^3 + \gamma_4^{(4)}I_4^4 + \ldots \tag{20}$$

Then, the total stress tensor (15) has the following form:

$$\begin{aligned}\boldsymbol{\tau} &= (\lambda I_1 - \tfrac{1}{2}\mu_0^{-1}I_4 - \gamma_4 I_4 - \gamma_5 I_5 - \gamma_6 I_6)\mathbf{I} + 2\mu\boldsymbol{\varepsilon} + \left(\mu_0^{-1} + \gamma_4' + \tfrac{1}{2}\gamma_5\right)\mathbf{B}\,\mathbf{B} + \\ &+ \tfrac{1}{2}(\gamma_5 + \gamma_6)\,(\mathbf{B}\,\boldsymbol{\varepsilon}\cdot\mathbf{B} + \mathbf{B}\cdot\boldsymbol{\varepsilon}\,\mathbf{B}) + \tfrac{1}{2}\gamma_6\,(\mathbf{B}\,\boldsymbol{\varepsilon}^2\cdot\mathbf{B} + \mathbf{B}\cdot\boldsymbol{\varepsilon}^2\,\mathbf{B})\end{aligned} \tag{21}$$

3.1 Uniaxial case

In this work we consider the uniaxial case, such that

$$\begin{aligned} \tau_{11} &= \tau; \text{ otherwise } \tau_{ij} = 0 \\ \varepsilon_{11} &= \varepsilon_1; \ \varepsilon_{22} = \varepsilon_{33} = \varepsilon_2; \text{ otherwise } \varepsilon_{ij} = 0 \qquad (22) \\ \mathbf{B} &= [B_1 \quad 0 \quad 0]^{\mathrm{T}} \qquad (23) \end{aligned}$$

where the indexes $1, 2, 3$ refer to the axes x_1, x_2, x_3, respectively.

The uniaxial model furnishes the following system of equations quadratic in ε_1:

$$\begin{cases} \tau &= \lambda(\varepsilon_1 + 2\varepsilon_2) + 2\mu\varepsilon_1 + \left(\frac{1}{2}\mu_0^{-1} + \frac{1}{2}\gamma_5 + \gamma_6\varepsilon_1\right) B_1^2 \\ 0 &= \lambda(\varepsilon_1 + 2\varepsilon_2) + 2\mu\varepsilon_2 - \left(\frac{1}{2}\mu_0^{-1} + \gamma_4' + \gamma_5\varepsilon_1 + \gamma_6\varepsilon_1^2\right) B_1^2 \end{cases} \qquad (24)$$

One of the solutions of the system represents the expression of the magnetostrictive strain in the direction of the x_1 axis which, after some manipulations, can be written as

$$\varepsilon_1 = -\frac{1}{4\nu\xi_6 b_1}\left\{\beta\left[1 - \left[1 + \frac{8\nu\xi_6 b_1}{\beta^2}\left(\frac{\tau}{E} - \frac{1}{2}(1 + 2\nu + 4\nu\xi_4 + \xi_5)b_1\right)\right]^{1/2}\right]\right\} \qquad (25)$$

with $\beta = 1 + (2\nu\xi_5 + \xi_6)b_1$ where $b_1 = \mu_0^{-1}B_1^2/E$. Furthermore $\xi_4 = \xi_4^{(0)} + \xi_4^{(1)}I_4 + \xi_4^{(2)}I_4^2 + \xi_4^{(3)}I_4^3 + \gamma_4^{(4)}I_4^4$. Note that $\xi_4^{(i)} = \gamma_4^{(i)}\mu_0$ $(i = 0, 1, 2, 3, 4)$, $\xi_5 = \gamma_5\mu_0$ and $\xi_6 = \gamma_6\mu_0$ are dimensionless parameters to be evaluated by fitting the experimental data in terms of magnetostrictive strains, magnetic field values and magnetic induction field values. Finally, ν is the Poisson's ratio and E the Young's modulus.

Note that, by neglecting the dependence of the Helmholtz free energy on the invariant I_6, system (24) becomes linear in ε_1 and the expression of the magnetostrictive strain in the direction of the x_1 axis reduces to:

$$\varepsilon_1 = \frac{1}{1 + 2\nu\xi_5 b_1}\left[\frac{\tau}{E} - \frac{1}{2}(1 + 2\nu + 4\nu\xi_4 + \xi_5)b_1\right] \qquad (26)$$

4 Some results

4.1 Measurements of magnetostriction

The experiments for measuring the magnetization and the magnetostriction data were conducted at the Laboratory of Electromechanics (Helsinki University of Technology, Finland). To obtain the data as function of both the magnetic induction field and the external mechanical stress, a slightly modified version of the 25 cm Epstein frame was used (see Figure 1). The modifications mainly involve the dimensions of the test specimen and the use of extra insulation between strips (see [2] for the details). The screw system allowed for the application of both tensile and compressive mechanical stresses. The use of plastic strips between the iron strips prevented them bending when compressive stress was applied. The measurement of magnetostriction was conducted by using the same set-up as for the measurement of magnetization. The only difference is a piezoelectric force transducer introduced between the load cell and the specimen holder. The transducer was used for the measurement of the force due to magnetostriction in the direction of the magnetic field.

As described in [2], the device for measuring magnetostriction furnishes a force from which the magnetostriction and the magnetostrictive stress need to be calculated. In particular, one strip of the measurement device with its supports and holders, can be schematized as shown in Figure (2). This part is mechanically equivalent to the system of springs shown in the same figure. The force measured by the piezoelectric force transducer is the force in the system of springs at equilibrium after the magnetostriction takes place. Since the measurement is made at a frequency of 5 Hz, the mass effect can be ignored. The spring constants of the iron strip, the force transducer, the load cell and the

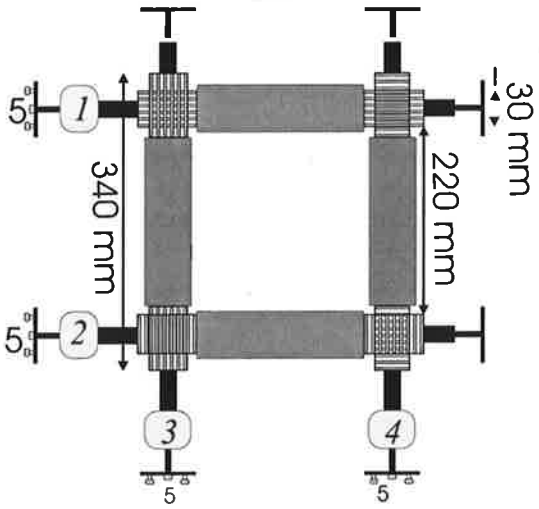


Figure 1: Epstein frame. Cross section of the iron strip: $S = 30\,\text{mm} \times 7.5\,\text{mm}$.

support are, respectively, $k_{iron} = 1.7 \times 10^8 \text{N/m}$, $k_{FT} = 1.0 \times 10^8 \text{N/m}$, $k_{LC} = 5.0 \times 10^8 \text{N/m}$ and $k_{support} = 5.0 \times 10^6 \text{N/m}$.

The elongation of each spring at equilibrium is calculated from the measured force and the spring constants. Then, the elongation Δu of the iron due to the magnetostriction is the sum of the elongation of each spring. To produce the same magnetostrictive elongation Δu in the iron, a magnetostrictive force

$$F_{MS} = -k_{iron} \Delta u \tag{27}$$

is needed. The magnetostrictive stress is calculated as

$$\sigma_{MS} = F_{MS}/S \tag{28}$$

where S is the cross section of the iron strip while the magnetostrictive strain will be:

$$\varepsilon_{MS} = \Delta u/l \tag{29}$$

where l is the length of the iron specimen.

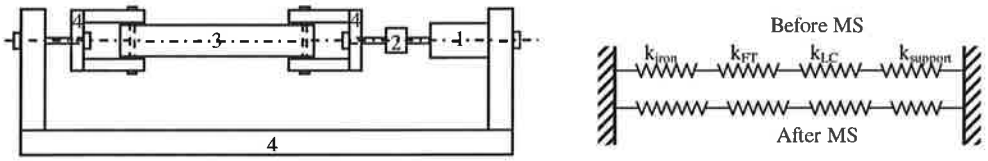


Figure 2: Schematic of the measurement device and its equivalent representation. Left: 1: load cell, 2: force transducer, 3: specimen, 4: supports. Right: equivalent representation.

4.2 Validation of the theoretical model

In the following, the magnetostriction curves obtained by fitting the experimental data for both cases of compressive (Figures 4 and 6) and tensile (Figures 5 and 7) mechanical pre-stresses, are shown. The used valued of the Young's modulus, the Poisson's ratio and the permeability in the

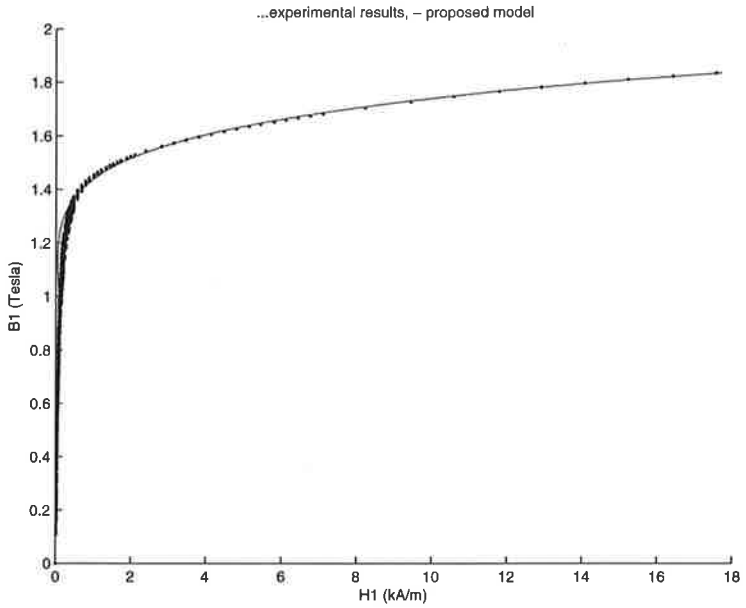


Figure 3: Magnetization curves: magnetic field (H1) vs. magnetic induction field (B1) in the presence of compressive pre-stresses; dependence of the free energy on the invariants I_1, I_2, I_4, I_5, I_6.

	compressive stress (MPa)	tensile stress (MPa)
1	0.0	0.0
2	4.360×10^{-2}	4.360×10^{-2}
3	8.720×10^{-1}	8.720×10^{-1}
4	1.744	1.308
5	2.616	1.744
6	3.488	2.180
7	4.360	2.616
8	5.232	3.052
9	6.104	3.488
10	6.976	3.924

Table 1: Magnetostriction curves: values of the pre-stresses.

	dependence of ψ on I_1, I_2, I_4, I_5	dependence of ψ on I_1, I_2, I_4, I_5, I_6
$\xi_4^{(0)}$	-0.99974858	-0.99974858
$\xi_4^{(1)}$	0.00076054	0.00076054
$\xi_4^{(2)}$	-0.00089881	-0.00089881
$\xi_4^{(3)}$	0.00008964	0.00008964
$\xi_4^{(4)}$	0.00012688	0.00012688
ξ_5	-0.34426261	-0.32974548
ξ_6		-7457.27

Table 2: Dimensionless parameters.

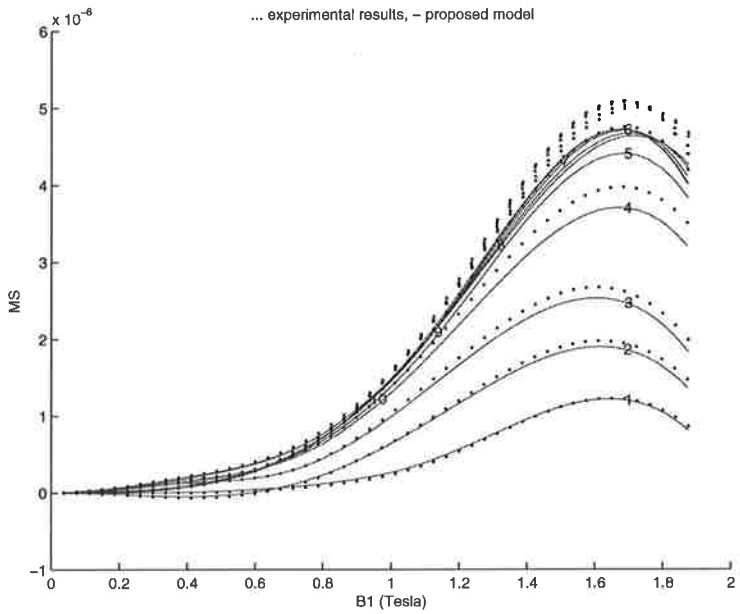


Figure 4: Magnetostriction (MS) vs. magnetic field (B1) in the presence of compressive pre-stresses (values reported in Table 1); dependence of the free energy on the invariants I_1, I_2, I_4, I_5.

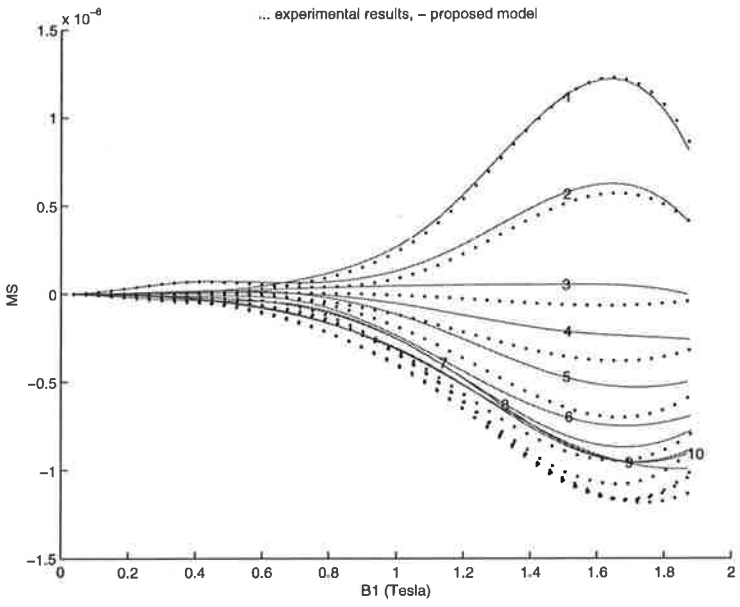


Figure 5: Magnetostriction (MS) vs. magnetic field (B1) in the presence of tensile pre-stresses (values reported in Table 1); dependence of the free energy on the invariants I_1, I_2, I_4, I_5.

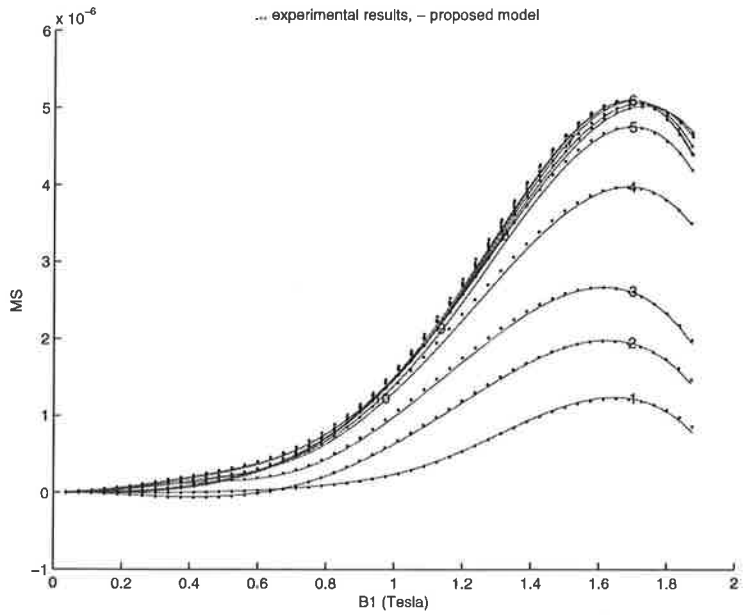


Figure 6: Magnetostriction (MS) vs. magnetic field (B1) in the presence of compressive pre-stresses (values reported in Table 1); dependence of the free energy on the invariants I_1, I_2, I_4, I_5, I_6.

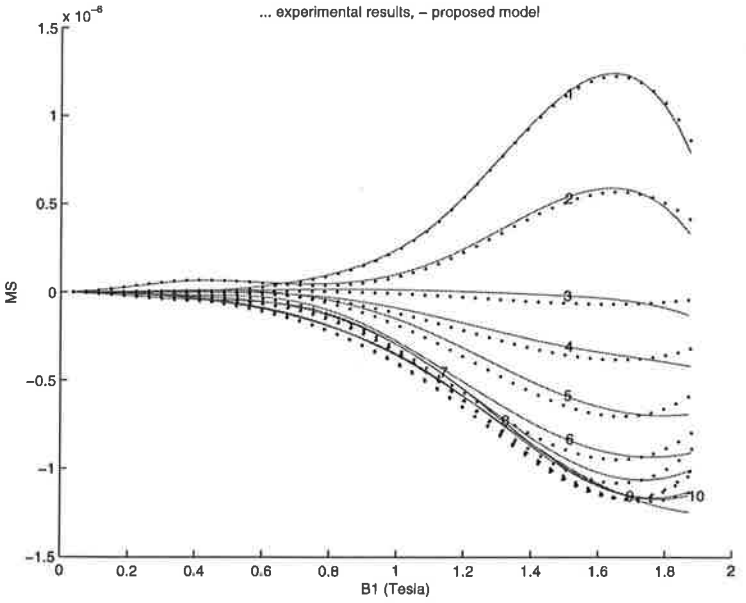


Figure 7: Magnetostriction (MS) vs. magnetic field (B1) in the presence of tensile pre-stresses (values reported in Table 1); dependence of the free energy on the invariants I_1, I_2, I_4, I_5, I_6.

vacuum are $E = 183.62\,\mathrm{GPa}$, $\nu = 0.34$ and $\mu_0 = 4\pi\times10^{-7}\,\mathrm{N/A^2}$, respectively. The values of the pre-stresses are reported in Table 1 while Table 2 report the estimated values obtained for the dimensionless parameters.

The results show that the proposed model is suitable to describe the phenomenon of magnetostriction. It is worth to note that the definition of the coefficient γ_4 through the fine polynomial expansion (17), permitted to correctly describe the effect of the magnetic saturation of the magnetostrictive strain under high magnetic fields. Furthermore, the dependence of the Helmholtz free energy on the invariant I_5 is necessary in order to furnish the correct sign of magnetostriction. Finally, the results show that the dependence of the free energy on the invariant I_6 is important to accurately predict the experimental results in the presence of external mechanical pre–stresses. Although the proposed model is simple, the obtained results are very good when compared with those of the standard square model used in [6] for magnetostrictive materials and based on the expansion in series of the Gibbs free energy. In that paper, the authors had to introduce more complicated models in order to better describe the magnetic saturation at high magnetic fields and the whole phenomenon of magnetostriction even at low values of pre-stresses.

The magnetization curves drawn in Figure 3 for the case of compressive pre-stresses show that the results predicted by the theoretical model agree well with the experimental data.

5 CONCLUSIONS

This paper presented a coupled magnetoelastic model for ferromagnetic materials. As proposed by Dorfmann *et al.* in ([8], [9],[10]), the constitutive equations are written on the basis of the Helmholtz free energy. In particular, the strain tensor and the magnetic induction vector are chosen as the basic variables. A form of the free energy suitable for describing the phenomenon of magnetostriction in ferromagnetic materials used in electrical machines is presented. The constitutive equations derived from the proposed free energy depend on the invariants I_1, I_2, I_4, I_5 and I_6 which are functions of both the magnetic and the mechanical field. The dimensionless parameters of the model were evaluated by fitting the experimental data of magnetization and magnetostriction obtained by means of a modified Epstein device. The results show a good agreement between the magnetostriction curves obtained from the theoretical model and the experimental results. Since both the magnetic saturation under growing magnetic fields and the magnetostriction in the presence of different mechanical external pre-stresses are described with accuracy, the proposed model is competitive with respect to the recent magnetoelastic coupled models proposed in literature (see [6]).

Acknowledgments

The present research was funded by TEKES, the National Technology Agency of Finland, in the context of the project KOMASI (decision number 40288/05).

REFERENCES

[1] Maugin GA. Continuum Mechanics of Electromagnetic Solids. North–Holland, Amsterdam, 1988.

[2] Belahcen A. Magnetoelasticity, Magnetic Forces and Magnetostriction in Electrical Machines. Doctoral Thesis. Helsinki University of Technology, Laboratory of Electromechanics. Report 72 (2004).

[3] Besbes M., Ren Z, Razek A. Finite Element Analysis of magneto-Mechanical Coupled Phenomena in Magnetostrictive Materials. IEEE Transaction on Magnetics 1996, 32 (3): 1058–1061.

[4] Besbes M, Ren Z, Razek A. A generalized FiniteElment Model of Magnetostriction Phenomena IEEE Transaction on Magnetics 2001; 37 (5): 3324–3328.

[5] Pérez-Aparicio J.L., Sosa H. A continuum three-dimensional, fully coupled, dynamic, non-linear finite element formulation for magnetostrictive materials. Smart Mater. Struct. 2004, 13:493–502.

[6] Wan Y, Fang D, Hwang KC. Non–linear constitutive relations for magnetostrictive materials. Int J Engr Mech 2003; 38: 1053–1065.

[7] Fang DN, Feng X, Hwang KC. Study of magnetomechanical non-linear deformation of ferro-magnetic materials. Theory and experiment. Proc. Instn Mech Engrs 2004; 218, Part C: J. Mechanical Engineering Science: 1405–1410.

[8] Brigadnov I.A., Dorfmann A. Mathematical modelling of magneto-sensitive elastomers. Int. J. of Solids and Struct. 2003; 40:4659–4674.

[9] Dorfmann A., Ogden R.J. Magnetoelastic modelling of elastomers Eur. J. of Mech. 2003; 22:497–507.

[10] Dorfmann A., Ogden R.W., Saccomandi G. Universal relations for non-linear magnetoelastic solids. Int. J. Non–Linear Mech. 2004; 39:1699–1708.

A MODEL FOR CONCRETE AT HIGH TEMPERATURES

Juha HARTIKAINEN, Reijo KOUHIA, Timo MANNINEN

Helsinki University of Technology, Structural Mechanics, P.O. Box 2100, FIN-02015 TKK

ABSTRACT

Prediction of concrete performance at high temperatures is of great importance, in particular for safety evaluation of concrete structures in fires in tunnels and high buildings or in thermal hazard situations like hypothetical core disruptive nuclear accidents. High temperatures induce strong physical and chemical changes in the micro-structure of concrete that affect its mechanical behaviour.

In the present paper, a preliminary version of a fully coupled non-linear multi-phase mathematical model based on continuum thermodynamics is presented for simulation of hygro-thermo-mechanical behaviour of moist, partially saturated concrete at high temperatures. The hygro-thermal behaviour of concrete is highly complex phenomenon which is influenced by cracking of concrete. The couplings between hygro-thermo-mechanical responses are dealt with by a consistent constitutive approach choosing proper expressions for the Helmholtz free energy and the dissipation potential.

1 INTRODUCTION

In many modern engineering applications concrete structures are exposed to temperatures above ambient conditions (temperatures higher that 50°C) such as pressure furnaces, chimneys and nuclear reactors. An extremely important area is the safety evaluation of structures under severe accidental situations like fires. In particular, the evaluation of the residual strength of structural components that have been damaged has become imperative for modern civil engineering.

In the literature, there are relatively few publications related to mathematical modelling of concrete structures at high temperatures [1]. This is probably due to the complexity of the problem [2, 3, 4, 5, 6].

Concrete is a composite material consisting mainly of a mineral aggregate bound by a matrix of hardened cement paste. The paste is highly porous and normally contains a large amount of free evaporable water. The chemical composition of the cement can vary considerably, e.g. the Portland cement manufactured mainly from a calcareous material has typically the following oxide composition: CaO, 60 - 67 %; SiO_2, 17 - 25 %; Al_2O_3, 3 - 8 %; and Fe_2O_3, 0.5 - 6 %. At any stage of hydration the cement paste includes gel (made of hydrates of the various compounds of cement), anhydrous cement, gel pores and capillary pores [1].

When concrete is exposed to high temperature, changes in chemical composition, physical structure and water content occur mainly in the hardened cement paste. Heating causes a gradual loss of evaporable water leading to dehydration of the hardened cement paste. Also, the calcium hydroxide in Portland cement paste starts to convert into calcium oxide, in which case the chemically bound water is gradually released to become free, evaporable

water. Heating can also induce changes in porosity and pore structure of the hardened cement paste.

The coefficient of thermal expansion of the binder is different from that of the aggregate. Therefore mere temperature increase induces microcracking and changes in the pore structure of concrete thus reducing its strength and stiffness [7]. Further mechanical degradation in heated concrete is induced by build-up pore pressure due to evaporation of pore water as well by shrinkage of the binder due to water loss. The reduction of strength and elastic stiffness due to different chemical, hydro- and thermomechanical processes is termed damage. The branch of continuum mechanics, which deals with behaviour of damaged and damaging materials, is called continuum (or continuous) damage mechanics (CDM) [8, 9].

In the present paper, a thermodynamically consistent model to describe the behaviour of concrete at elevated temperatures is presented. The model is based on the theory of mixtures and the principles of continuum mechanics and macroscopic thermodynamics [10, 11, 12, 13, 14, 15]. In the model, the following phenomena are taken into account:

- the reversible thermoelastic behaviour of concrete,
- damaging of concrete,
- transport of water, water vapour and air in the porous cement paste,
- adsorption of water into the cement gel,
- the phase change between water and water vapour and
- diffusion of water vapour in the gaseous component.

2 THERMOMECHANICAL THEORY

2.1 Basic definitions and kinematics

In the continuum theory of mixtures each constituent is assumed to be spread over the spatial domain in a continuous manner and are able to coexist at any point of the region. As a consequence of the smoothing the variables and functions are continuous and differentiable describing the state of the material in some average sense.

The concrete is considered as a multiconstituent system consisting of a solid skeleton (s), liquid water (l) and a gaseous component (g) of water vapour (v) and air (a).

The volume fractions of the solid, liquid and gaseous components are defined as

$$\beta_k = \frac{\mathrm{d}V_k}{\mathrm{d}V}, \qquad k \in \{\mathrm{s}, \mathrm{l}, \mathrm{g}\}, \tag{1}$$

where $\mathrm{d}V_k$ is the volume of component k and $\mathrm{d}V$ the reference volume. Since the water vapor and air belong to the same gaseous component, their relative proportions are measured via the molar fractions ζ_v and ζ_a such that

$$\zeta_\mathrm{v} \equiv \zeta = \frac{n_\mathrm{v}}{n_\mathrm{v} + n_\mathrm{a}}, \qquad \zeta_\mathrm{a} \equiv 1 - \zeta = \frac{n_\mathrm{a}}{n_\mathrm{v} + n_\mathrm{a}}, \tag{2}$$

where n_k the mole number of constituent k. The molar volume fractions, ξ_k, are defined as

$$\xi = \beta, \quad \xi_\mathrm{l} = \beta_\mathrm{l}, \quad \xi_\mathrm{v} = \zeta_\mathrm{v}\beta_\mathrm{g} \equiv \zeta\beta_\mathrm{g}, \quad \xi_\mathrm{a} = \zeta_\mathrm{a}\beta_\mathrm{g} \equiv (1-\zeta)\beta_\mathrm{g}. \tag{3}$$

Apparently the molar volume fractions satisfy the constraints

$$\sum_k \xi_k = 1, \quad \xi_k \geq 0, \quad k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}. \tag{4}$$

The molar volume fractions ξ_k relate the the apparent densities ρ_k to the intrinsic (bulk) densities $\bar{\rho}_k$ according to

$$\rho_k = \xi_k \bar{\rho}_k, \quad k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}. \tag{5}$$

In addition, it is useful to introduce the porosity, η, and the water saturation, χ, as follows

$$\eta = 1 - \xi\,, \quad \chi = \frac{\xi_{\mathrm{l}}}{1-\xi}. \tag{6}$$

The state of motion of constituent k at an arbitrary instant of time t is described by a velocity field $\boldsymbol{v}_k(\boldsymbol{x},t)$, where $\boldsymbol{x}$ is the vector of spatial coordinates. The motion the solid component can be described more conveniently by its displacement field $\boldsymbol{u} \equiv \boldsymbol{u}(\boldsymbol{x},t)$.

The deformations are described either by the rate of deformation

$$\boldsymbol{d}_k = \tfrac{1}{2}\left[\nabla \boldsymbol{v}_k + (\nabla \boldsymbol{v}_k)^{\mathsf{T}}\right], \quad k \in \{\mathrm{s},\mathrm{l},\mathrm{v},\mathrm{a}\} \tag{7}$$

or by the strain

$$\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon} = \tfrac{1}{2}\left[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathsf{T}}\right]. \tag{8}$$

The material time derivative of a quantity following the movement of constituent k is determined by the operator

$$\frac{\mathrm{d}_k}{\mathrm{d}t} = \frac{\partial}{\partial t} + \boldsymbol{v}_k \cdot \nabla. \tag{9}$$

The velocity $\boldsymbol{v}$ is then obviously the material time derivative $\mathrm{d}\,\boldsymbol{u}/\mathrm{d}t$.

Because constituents have in general different velocities at the same macroscopic point of the mixture, a reference velocity field $\boldsymbol{v}_*(\boldsymbol{x},t)$ is introduced in order to establish the fundamental principles for the mixture. The material time derivative with respect to the reference movement can be expressed as follows

$$\frac{\mathrm{d}_k}{\mathrm{d}t} = \frac{\mathrm{d}_*}{\mathrm{d}t} + \boldsymbol{v}_{k*} \cdot \nabla, \tag{10}$$

where $\boldsymbol{v}_{k*}$ is the relative velocity of the constituent k with respect to the reference velocity: $\boldsymbol{v}_{k*} = \boldsymbol{v}_k - \boldsymbol{v}_*$. In the classical mixture theory [16], the barycentric velocity is used as a reference, whereas in the models based on Biot's theory [17], the movement of the solid is taken as the reference movement. In this paper, the latter approach is chosen, i.e. $\boldsymbol{v}_* = \boldsymbol{v}$.

It is also convenient to define the generalised Darcian fluxes, $\boldsymbol{J}_{\mathrm{l}}$ and $\boldsymbol{J}_{\mathrm{g}}$, of liquid and gaseous components with respect to the motion of the solid component as well as the generalised Fickian flux, $\boldsymbol{j}_{\mathrm{v}}$, of vapour constituent with respect to the motion of the gaseous component such that

$$\boldsymbol{J}_{\mathrm{l}} = \beta_{\mathrm{l}}(\boldsymbol{v}_{\mathrm{l}} - \boldsymbol{v}\ ), \quad \boldsymbol{J}_{\mathrm{g}} = \beta_{\mathrm{g}}(\boldsymbol{v}_{\mathrm{g}} - \boldsymbol{v}\ ), \quad \boldsymbol{j}_{\mathrm{v}} = \zeta(\boldsymbol{v}_{\mathrm{v}} - \boldsymbol{v}_{\mathrm{g}}), \tag{11}$$

where the molar-weighted velocity of the gaseous component $\boldsymbol{v}_{\mathrm{g}}$ is defined as

$$\boldsymbol{v}_{\mathrm{g}} = \zeta \boldsymbol{v}_{\mathrm{v}} + (1-\zeta)\boldsymbol{v}_{\mathrm{a}}. \tag{12}$$

2.2 Conservation laws and entropy inequality

Introducing the Cauchy stress tensor $\boldsymbol{\sigma}_k$, the specific internal energy e_k the heat flux vector $\boldsymbol{q}_k$ and the external energy supply r_k of constituents $k \in \{\mathrm{s},\mathrm{l},\mathrm{v},\mathrm{a}\}$ as well as the acceleration of gravity $\boldsymbol{g}$, the conservation laws of mass, linear momentum and energy for the concrete can be expressed as follows [12]:

$$\theta = 0, \tag{13}$$

$$\theta_{\mathrm{l}} + \theta_{\mathrm{v}} = 0, \tag{14}$$

$$\theta_{\mathrm{a}} = 0, \tag{15}$$

$$\boldsymbol{m} + \boldsymbol{m}_{\mathrm{l}} + \boldsymbol{m}_{\mathrm{v}} + \boldsymbol{m}_{\mathrm{a}} = 0, \tag{16}$$

$$\ell + \ell_{\mathrm{l}} + \ell_{\mathrm{v}} + \ell_{\mathrm{a}} = 0, \tag{17}$$

where the mass production rate θ_k, the linear momentum production rate $\boldsymbol{m}_k$ and the energy production rate ℓ_k for constituents $k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}$ are defined as

$$\theta_k = \frac{\partial \rho_k}{\partial t} + \nabla\cdot(\rho_k \boldsymbol{v}_k), \tag{18}$$

$$\boldsymbol{m}_k = \rho_k \frac{\mathrm{d}_k \boldsymbol{v}_k}{\mathrm{d}t} + \theta_k \boldsymbol{v}_k - \nabla\cdot\boldsymbol{\sigma}_k - \rho_k \boldsymbol{g}, \tag{19}$$

$$\ell_k = \rho_k \frac{\mathrm{d}_k e_k}{\mathrm{d}t} + (e_k - \tfrac{1}{2}\boldsymbol{v}_k\cdot\boldsymbol{v}_k)\theta_k - \boldsymbol{\sigma}_k : \boldsymbol{d}_k + \boldsymbol{m}_k\cdot\boldsymbol{v}_{k*} + \nabla\cdot\boldsymbol{q}_k - r_k. \tag{20}$$

According to the second principle of thermodynamics the entropy production should always be positive. Introducing the absolute temperature T and the specific entropy s_k of constituent k, the entropy inequality for the concrete can be stated as

$$T(\gamma\ + \gamma_\mathrm{l} + \gamma_\mathrm{v} + \gamma_\mathrm{a}) \geq 0, \tag{21}$$

where the entropy production rate γ_k of constituent k is defined as follows

$$\gamma_k = \rho_k \frac{\mathrm{d}_k s_k}{\mathrm{d}t} + s_k\theta_k + \nabla\cdot\left(\frac{\boldsymbol{q}_k}{T}\right) - \frac{r_k}{T}. \tag{22}$$

2.3 Constitutive relations

The thermodynamic state and the material behaviour are defined in terms of variables of state and dissipation through the Helmholz free energies and the dissipation potential. The variables defining the thermodynamic state are the absolute temperature T, which is assumed to be uniform for all constituents, the strain tensor $\boldsymbol{\varepsilon}$ and the damage tensor $\boldsymbol{D}$, which takes into account microfracturing of the solid component, as well as the intrinsic densities $\bar{\rho}$, $\bar{\rho}_\mathrm{l}$, $\bar{\rho}_\mathrm{v}$ and $\bar{\rho}_\mathrm{a}$ and the molar volume fractions ξ , ξ_l, ξ_v and ξ_a. The variables defining the dissipation behaviour in turn are the heat flux, $\boldsymbol{q} = \boldsymbol{q}\ + \boldsymbol{q}_\mathrm{l} + \boldsymbol{q}_\mathrm{v} + \boldsymbol{q}_\mathrm{a}$, the rate of damage, $\dot{\boldsymbol{D}} \equiv \mathrm{d}\ \boldsymbol{D}/\mathrm{d}t$, and the relative velocities $\boldsymbol{v}_\mathrm{l}$, $\boldsymbol{v}_\mathrm{v}$ and $\boldsymbol{v}_\mathrm{a}$. Reversible material behaviour is described by means of the Helmholz free energies

$$\begin{aligned} \psi\ &= \psi\ (T, \boldsymbol{\epsilon}, \boldsymbol{D}, \bar{\rho}\ , \xi\ , \xi_\mathrm{l}, \xi_\mathrm{v}, \xi_\mathrm{a}), \quad & \psi_\mathrm{l} &= \psi_\mathrm{l}(T, \bar{\rho}_\mathrm{l}, \xi\ , \xi_\mathrm{l}, \xi_\mathrm{v}, \xi_\mathrm{a}), \\ \psi_\mathrm{v} &= \psi_\mathrm{v}(T, \bar{\rho}_\mathrm{v}, \xi\ , \xi_\mathrm{l}, \xi_\mathrm{v}, \xi_\mathrm{a}), & \psi_\mathrm{a} &= \psi_\mathrm{a}(T, \bar{\rho}_\mathrm{a}, \xi\ , \xi_\mathrm{l}, \xi_\mathrm{v}, \xi_\mathrm{a}), \end{aligned} \tag{23}$$

whereas irreversible material behaviour is characterised through the dissipation potential

$$\phi = \phi(\boldsymbol{q}, \dot{\boldsymbol{D}}, \boldsymbol{v}_\mathrm{l}\ , \boldsymbol{v}_\mathrm{v}\ , \boldsymbol{v}_\mathrm{a}\ ;\ T, D, \bar{\rho}_\mathrm{v}, \xi\ , \xi_\mathrm{l}, \xi_\mathrm{v}, \xi_\mathrm{a}). \tag{24}$$

The thermodynamically admissible constitutive relations are derived from the Helmholz free energies and the dissipation potential by exploiting the entropy inequality (22) as follows. Introducing the Legendre transformations

$$T^{-1}\psi_k = T^{-1}e_k - s_k, \quad k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}, \tag{25}$$

postulating the representation of the power of dissipation

$$T(\gamma\ + \gamma_\mathrm{l} + \gamma_\mathrm{v} + \gamma_\mathrm{a}) = \frac{\partial\phi}{\partial\boldsymbol{q}}\cdot\boldsymbol{q} + \frac{\partial\phi}{\partial\dot{\boldsymbol{D}}} : \dot{\boldsymbol{D}} + \frac{\partial\phi}{\partial\boldsymbol{v}_\mathrm{l}}\cdot\boldsymbol{v}_\mathrm{l}\ + \frac{\partial\phi}{\partial\boldsymbol{v}_\mathrm{v}}\cdot\boldsymbol{v}_\mathrm{v}\ + \frac{\partial\phi}{\partial\boldsymbol{v}_\mathrm{a}}\cdot\boldsymbol{v}_\mathrm{a} \tag{26}$$

and taking into account the field equations (13)-(20) yield the following equation

$$
\begin{aligned}
&-\rho\left(s+\frac{\partial\psi}{\partial T}\right)\frac{\mathrm{d}\,T}{\mathrm{d}t}-\rho_{\mathrm{l}}\left(s_{\mathrm{l}}+\frac{\partial\psi_{\mathrm{l}}}{\partial T}\right)\frac{\mathrm{d}_{\mathrm{l}}T}{\mathrm{d}t}-\rho_{\mathrm{v}}\left(s_{\mathrm{v}}+\frac{\partial\psi_{\mathrm{v}}}{\partial T}\right)\frac{\mathrm{d}_{\mathrm{v}}T}{\mathrm{d}t}\\
&\qquad-\rho_{\mathrm{a}}\left(s_{\mathrm{a}}+\frac{\partial\psi_{\mathrm{a}}}{\partial T}\right)\frac{\mathrm{d}_{\mathrm{a}}T}{\mathrm{d}t}-\left(\frac{\nabla T}{T}+\frac{\partial\phi}{\partial\boldsymbol{q}}\right)\cdot\boldsymbol{q}\\
&\qquad-\left[\psi_{\mathrm{v}}+\frac{1}{\bar{\rho}_{\mathrm{v}}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{v}}}+\tfrac{1}{2}\boldsymbol{v}_{\mathrm{v}}\cdot\boldsymbol{v}_{\mathrm{v}}-\psi_{\mathrm{l}}-\frac{1}{\bar{\rho}_{\mathrm{l}}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{l}}}-\tfrac{1}{2}\boldsymbol{v}_{\mathrm{l}}\cdot\boldsymbol{v}_{\mathrm{l}}\right]\theta_{\mathrm{v}}\\
&\qquad+\left(\boldsymbol{\sigma}-\rho\frac{\partial\psi}{\partial\epsilon}+\xi\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi}\mathbf{I}\right):\boldsymbol{d}+\left(\boldsymbol{\sigma}_{\mathrm{l}}+\xi_{\mathrm{l}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{l}}}\mathbf{I}\right):\boldsymbol{d}_{\mathrm{l}}\\
&\qquad+\left(\boldsymbol{\sigma}_{\mathrm{v}}+\xi_{\mathrm{v}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{v}}}\mathbf{I}\right):\boldsymbol{d}_{\mathrm{v}}+\left(\boldsymbol{\sigma}_{\mathrm{a}}+\xi_{\mathrm{a}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{a}}}\mathbf{I}\right):\boldsymbol{d}_{\mathrm{a}}-\left(\rho\frac{\partial\psi}{\partial\boldsymbol{D}}+\frac{\partial\phi}{\partial\dot{\boldsymbol{D}}}\right):\dot{\boldsymbol{D}}\\
&\qquad-\left(\rho\frac{\partial\psi}{\partial\bar{\rho}}-\frac{\xi}{\bar{\rho}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi}\right)\frac{\mathrm{d}\,\bar{\rho}}{\mathrm{d}t}-\left(\rho_{\mathrm{l}}\frac{\partial\psi_{\mathrm{l}}}{\partial\bar{\rho}_{\mathrm{l}}}-\frac{\xi_{\mathrm{l}}}{\bar{\rho}_{\mathrm{l}}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{l}}}\right)\frac{\mathrm{d}_{\mathrm{l}}\bar{\rho}_{\mathrm{l}}}{\mathrm{d}t}\\
&\qquad-\left(\rho_{\mathrm{v}}\frac{\partial\psi_{\mathrm{v}}}{\partial\bar{\rho}_{\mathrm{v}}}-\frac{\xi_{\mathrm{v}}}{\bar{\rho}_{\mathrm{v}}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{v}}}\right)\frac{\mathrm{d}_{\mathrm{v}}\bar{\rho}_{\mathrm{v}}}{\mathrm{d}t}-\left(\rho_{\mathrm{a}}\frac{\partial\psi_{\mathrm{a}}}{\partial\bar{\rho}_{\mathrm{a}}}-\frac{\xi_{\mathrm{a}}}{\bar{\rho}_{\mathrm{a}}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{a}}}\right)\frac{\mathrm{d}_{\mathrm{a}}\bar{\rho}_{\mathrm{a}}}{\mathrm{d}t}\\
&\qquad-\left[\boldsymbol{m}_{\mathrm{l}}+\sum_j\left(\rho_{\mathrm{l}}\frac{\partial\psi_{\mathrm{l}}}{\partial\xi_j}\nabla\xi_j-\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{l}}}\nabla\xi_{\mathrm{l}}\right)+\frac{\partial\phi}{\partial\boldsymbol{v}_{\mathrm{l}}}\right]\cdot\boldsymbol{v}_{\mathrm{l}}\\
&\qquad-\left[\boldsymbol{m}_{\mathrm{v}}+\sum_j\left(\rho_{\mathrm{v}}\frac{\partial\psi_{\mathrm{v}}}{\partial\xi_j}\nabla\xi_j-\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{v}}}\nabla\xi_{\mathrm{v}}\right)+\frac{\partial\phi}{\partial\boldsymbol{v}_{\mathrm{v}}}\right]\cdot\boldsymbol{v}_{\mathrm{v}}\\
&\qquad-\left[\boldsymbol{m}_{\mathrm{a}}+\sum_j\left(\rho_{\mathrm{a}}\frac{\partial\psi_{\mathrm{a}}}{\partial\xi_j}\nabla\xi_j-\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{a}}}\nabla\xi_{\mathrm{a}}\right)+\frac{\partial\phi}{\partial\boldsymbol{v}_{\mathrm{a}}}\right]\cdot\boldsymbol{v}_{\mathrm{a}}=0. \qquad (27)
\end{aligned}
$$

Requiring eq. (27) to hold for any evolution of $\mathrm{d}\,T/\mathrm{d}t$, $\mathrm{d}_{\mathrm{l}}T/\mathrm{d}t$, $\mathrm{d}_{\mathrm{v}}T/\mathrm{d}t$, $\mathrm{d}_{\mathrm{a}}T/\mathrm{d}t$, $\boldsymbol{q}$, θ_{v}, $\boldsymbol{d}$, $\boldsymbol{d}_{\mathrm{l}}$, $\boldsymbol{d}_{\mathrm{v}}$, $\boldsymbol{d}_{\mathrm{a}}$, $\dot{\boldsymbol{D}}$, $\mathrm{d}\,\bar{\rho}\,/\mathrm{d}t$, $\mathrm{d}_{\mathrm{l}}\bar{\rho}_{\mathrm{l}}/\mathrm{d}t$, $\mathrm{d}_{\mathrm{v}}\bar{\rho}_{\mathrm{v}}/\mathrm{d}t$, $\mathrm{d}_{\mathrm{a}}\bar{\rho}_{\mathrm{a}}/\mathrm{d}t$, $\boldsymbol{v}_{\mathrm{l}}$, $\boldsymbol{v}_{\mathrm{v}}$ and $\boldsymbol{v}_{\mathrm{a}}$ result in the relevant constitutive relations

$$
s=-\frac{\partial\psi}{\partial T},\quad s_{\mathrm{l}}=-\frac{\partial\psi_{\mathrm{l}}}{\partial T},\quad s_{\mathrm{v}}=-\frac{\partial\psi_{\mathrm{v}}}{\partial T},\quad s_{\mathrm{a}}=-\frac{\partial\psi_{\mathrm{a}}}{\partial T}, \qquad (28)
$$

$$
\frac{\nabla T}{T}=-\frac{\partial\phi}{\partial\boldsymbol{q}}, \qquad (29)
$$

$$
\begin{aligned}
&\boldsymbol{\sigma}=\rho\frac{\partial\psi}{\partial\epsilon}-\xi\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi}\mathbf{I},\quad \boldsymbol{\sigma}_{\mathrm{l}}=-\xi_{\mathrm{l}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{l}}}\mathbf{I},\\
&\boldsymbol{\sigma}_{\mathrm{v}}=-\xi_{\mathrm{v}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{v}}}\mathbf{I},\qquad \boldsymbol{\sigma}_{\mathrm{a}}=-\xi_{\mathrm{a}}\sum_j\rho_j\frac{\partial\psi_j}{\partial\xi_{\mathrm{a}}}\mathbf{I},
\end{aligned} \qquad (30)
$$

$$\rho \frac{\partial \psi}{\partial \boldsymbol{D}} + \frac{\partial \phi}{\partial \dot{\boldsymbol{D}}} = \boldsymbol{0}, \tag{31}$$

$$\begin{aligned} \rho \frac{\partial \psi}{\partial \bar{\rho}} - \frac{\xi}{\bar{\rho}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi} = 0, \quad \rho_{\mathrm{l}} \frac{\partial \psi_{\mathrm{l}}}{\partial \bar{\rho}_{\mathrm{l}}} - \frac{\xi_{\mathrm{l}}}{\bar{\rho}_{\mathrm{l}}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{l}}} = 0, \\ \rho_{\mathrm{v}} \frac{\partial \psi_{\mathrm{v}}}{\partial \bar{\rho}_{\mathrm{v}}} - \frac{\xi_{\mathrm{v}}}{\bar{\rho}_{\mathrm{v}}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{v}}} = 0, \quad \rho_{\mathrm{a}} \frac{\partial \psi_{\mathrm{a}}}{\partial \bar{\rho}_{\mathrm{a}}} - \frac{\xi_{\mathrm{a}}}{\bar{\rho}_{\mathrm{a}}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{a}}} = 0, \end{aligned} \tag{32}$$

$$\psi_{\mathrm{v}} + \frac{1}{\bar{\rho}_{\mathrm{v}}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{v}}} + \tfrac{1}{2} \boldsymbol{v}_{\mathrm{v}} \cdot \boldsymbol{v}_{\mathrm{v}} - \psi_{\mathrm{l}} - \frac{1}{\bar{\rho}_{\mathrm{l}}} \sum_j \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{l}}} - \tfrac{1}{2} \boldsymbol{v}_{\mathrm{l}} \cdot \boldsymbol{v}_{\mathrm{l}} = 0, \tag{33}$$

$$\begin{aligned} \boldsymbol{m}_{\mathrm{l}} &= -\sum_j \left(\rho_{\mathrm{l}} \frac{\partial \psi_{\mathrm{l}}}{\partial \xi_j} \nabla \xi_j - \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{l}}} \nabla \xi_{\mathrm{l}} \right) - \frac{\partial \phi}{\partial \boldsymbol{v}_{\mathrm{l}}}, \\ \boldsymbol{m}_{\mathrm{v}} &= -\sum_j \left(\rho_{\mathrm{v}} \frac{\partial \psi_{\mathrm{v}}}{\partial \xi_j} \nabla \xi_j - \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{v}}} \nabla \xi_{\mathrm{v}} \right) - \frac{\partial \phi}{\partial \boldsymbol{v}_{\mathrm{v}}}, \\ \boldsymbol{m}_{\mathrm{a}} &= -\sum_j \left(\rho_{\mathrm{a}} \frac{\partial \psi_{\mathrm{a}}}{\partial \xi_j} \nabla \xi_j - \rho_j \frac{\partial \psi_j}{\partial \xi_{\mathrm{a}}} \nabla \xi_{\mathrm{a}} \right) - \frac{\partial \phi}{\partial \boldsymbol{v}_{\mathrm{a}}}. \end{aligned} \tag{34}$$

3 SPECIFIC MODEL

The mathematical model consisting in the conservation laws (13)-(20) and the constitutive relations (28)-(34) is specified next by choosing particular expressions for the thermodynamic potentials ψ_k, $k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}$ and ϕ.

The Helmholz free energies are composed of thermal (T), mechanical (M), interaction (I) and constraint (C) parts as

$$\psi_k = \psi_k^{\mathrm{T}} + \psi_k^{\mathrm{M}} + \psi_k^{\mathrm{I}} + \psi_s^{\mathrm{C}}, \quad k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}, \tag{35}$$

and the partial expressions are defined with respect to a prescribed reference state, in which $T = T_0$, $\boldsymbol{\epsilon} = \boldsymbol{0}$, $\boldsymbol{D} = \boldsymbol{0}$, and $\psi_k^{\mathrm{T}} = \psi_{k,0}^{\mathrm{T}}$, $\psi_k^{\mathrm{M}} = \psi_{k,0}^{\mathrm{M}}$, $\psi_k^{\mathrm{I}} = \psi_{k,0}^{\mathrm{I}}$, $\psi_k^{\mathrm{C}} = \psi_{k,0}^{\mathrm{C}}$, $\xi_k = \xi_{k,0}$ and $\bar{\rho}_k = \bar{\rho}_{k,0}$ for $k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}$.

The expressions for thermal contributions given as follows

$$\psi^{\mathrm{T}} = \psi_{,0}^{\mathrm{T}} - c_{,0} \left[T \ln \frac{T}{T_0} - (T - T_0) \right], \tag{36}$$

$$\psi_{\mathrm{l}}^{\mathrm{T}} = \psi_{\mathrm{l},0}^{\mathrm{T}} - c_{\mathrm{l},0} \left[T \ln \frac{T}{T_0} - (T - T_0) \right] - l_{\mathrm{l},0} \frac{T - T_0}{T_0}, \tag{37}$$

$$\psi_{\mathrm{v}}^{\mathrm{T}} = \psi_{\mathrm{v},0}^{\mathrm{T}} - c_{\mathrm{v},0} \left[T \ln \frac{T}{T_0} - (T - T_0) \right] - l_{\mathrm{v},0} \frac{T - T_0}{T_0}, \tag{38}$$

$$\psi_{\mathrm{a}}^{\mathrm{T}} = \psi_{\mathrm{a},0}^{\mathrm{T}} - c_{\mathrm{a},0} \left[T \ln \frac{T}{T_0} - (T - T_0) \right], \tag{39}$$

determine the caloric capacity of the system in terms of the specific heat capacities $c_{k,0}$ of constituents $k \in \{\mathrm{s}, \mathrm{l}, \mathrm{v}, \mathrm{a}\}$ and the fusion heat of melting $l_{\mathrm{l},0}$ and the fusion heat of sublimation $l_{\mathrm{v},0}$ of water.

The expressions for the mechanical parts

$$\psi^{\mathrm{M}} = \psi^{\mathrm{M}}_{,0} - \frac{\bar{K}_{,0}}{\bar{\rho}}\left(\ln\frac{\bar{\rho}}{\bar{\rho}_{,0}} + 1\right) + \frac{1}{\bar{\rho}}\left[\frac{\nu G}{1-2\nu}(\operatorname{tr}\boldsymbol{\varepsilon})^2 + G\operatorname{tr}(\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}) - \gamma(T-T_0)\operatorname{tr}\boldsymbol{\epsilon}\right.$$
$$\left. + a_1 \operatorname{tr}\boldsymbol{D}(\operatorname{tr}\boldsymbol{\varepsilon})^2 + a_2 \operatorname{tr}\boldsymbol{D}\operatorname{tr}(\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}) + a_3 \operatorname{tr}\boldsymbol{\varepsilon}\operatorname{tr}(\boldsymbol{\varepsilon}\cdot\boldsymbol{D}) + a_4 \operatorname{tr}(\boldsymbol{\varepsilon}\cdot\boldsymbol{D}\cdot\boldsymbol{\varepsilon})\right], \tag{40}$$

$$\psi^{\mathrm{M}}_{\mathrm{l}} = \psi^{\mathrm{M}}_{\mathrm{l},0} - \frac{\bar{K}_{\mathrm{l},0}}{\bar{\rho}_{\mathrm{l}}}\left(\ln\frac{\bar{\rho}_{\mathrm{l}}}{\bar{\rho}_{\mathrm{l},0}} + 1\right), \tag{41}$$

$$\psi^{\mathrm{M}}_{\mathrm{v}} = \psi^{\mathrm{M}}_{\mathrm{v},0} + \frac{RT}{M_{\mathrm{v}}}\ln\frac{\bar{\rho}_{\mathrm{v}}}{\bar{\rho}_{\mathrm{v},0}}, \tag{42}$$

$$\psi^{\mathrm{M}}_{\mathrm{a}} = \psi^{\mathrm{M}}_{\mathrm{a},0} + \frac{RT}{M_{\mathrm{a}}}\ln\frac{\bar{\rho}_{\mathrm{a}}}{\bar{\rho}_{\mathrm{a},0}} \tag{43}$$

take into account compressibility of each component and damaging of the thermoelastic solid component [18, 19]. $\bar{K}_{k,0}$ denotes the intrinsic bulk modulus of constituent k and G is the shear modulus, ν the Poisson's ratio and γ the thermal expansion parameter of the solid component at the reference state. Considering isotropic thermal expansion, the additional material parameters a_1, a_2, a_3 and a_4 are constrained by the relations

$$2a_1 + \tfrac{2}{3}a_2 + \tfrac{1}{3}a_3 = 0, \quad a_3 + \tfrac{2}{3}a_4 = 0. \tag{44}$$

Furthermore, M_k is the molecular weight of constituent k and R the universal gas constant.

The the adsorption of pore water and the interactions between vapour and air are characterised by the molar volume fractions through the following expressions

$$\psi^{\mathrm{I}} = \psi^{\mathrm{I}}_{,0}, \tag{45}$$

$$\psi^{\mathrm{I}}_{\mathrm{l}} = \psi^{\mathrm{I}}_{\mathrm{l},0} + \frac{RT}{M_{\mathrm{l}}}a\left(\frac{\xi_{\mathrm{l}}}{\xi} - \frac{\xi_{\mathrm{l},0}}{\xi_{,0}}\right)^b, \tag{46}$$

$$\psi^{\mathrm{I}}_{\mathrm{v}} = \psi^{\mathrm{I}}_{\mathrm{v},0} + \frac{RT}{M_{\mathrm{v}}}\ln\frac{\xi_{\mathrm{v}}}{\xi_{\mathrm{v}}+\xi_{\mathrm{a}}}, \tag{47}$$

$$\psi^{\mathrm{I}}_{\mathrm{a}} = \psi^{\mathrm{I}}_{\mathrm{a},0} + \frac{RT}{M_{\mathrm{a}}}\ln\frac{\xi_{\mathrm{a}}}{\xi_{\mathrm{v}}+\xi_{\mathrm{a}}}, \tag{48}$$

where the parameters a and b depend mainly on the specific surface area of the porous medium.

The constraints (4) pertaining to the molar volume fractions are dealt with an indicator function, $\mathcal{I}$, such that [15]

$$\psi^{\mathrm{c}}_k = \frac{T}{\bar{\rho}_k}\mathcal{I}(\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}}), \qquad \mathcal{I}(\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}}) = \begin{cases} 0 & \text{if } (\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}}) \in C, \\ +\infty & \text{otherwise,} \end{cases} \tag{49}$$

where

$$C = \left\{(\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}}) \in \mathbb{R}^4 \mid \textstyle\sum_k \xi_k = 1,\ \xi_k \geq 0,\ k \in \{\mathrm{s},\mathrm{l},\mathrm{v},\mathrm{a}\}\right\}. \tag{50}$$

The dissipation potential involves contributions from heat transfer, evolution of damage, motions of liquid and gas and diffusion of vapour as follows

$$\phi = \tfrac{1}{2}T^{-1}\boldsymbol{q}\cdot\boldsymbol{\lambda}^{-1}\cdot\boldsymbol{q} + \tfrac{1}{2}\dot{\boldsymbol{D}}\cdot\boldsymbol{B}^{-1}\cdot\dot{\boldsymbol{D}} + \tfrac{1}{2}\mu_{\mathrm{l}}\boldsymbol{J}_{\mathrm{l}}\cdot\boldsymbol{\kappa}_{\mathrm{l}}^{-1}\cdot\boldsymbol{J}_{\mathrm{l}} + \tfrac{1}{2}\mu_{\mathrm{g}}\boldsymbol{J}_{\mathrm{g}}\cdot\boldsymbol{\kappa}_{\mathrm{g}}^{-1}\cdot\boldsymbol{J}_{\mathrm{g}} + \tfrac{1}{2}D_{\mathrm{v}}^{-1}\boldsymbol{j}_{\mathrm{v}}\cdot\boldsymbol{j}_{\mathrm{v}}, \tag{51}$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(T,\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}})$ is the thermal conductivity tensor and $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{D})$ the damage evolution tensor of the concrete. In addition, $\boldsymbol{\kappa}_{\mathrm{l}} = \boldsymbol{\kappa}_{\mathrm{l}}(\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}})$ and $\boldsymbol{\kappa}_{\mathrm{g}} = \boldsymbol{\kappa}_{\mathrm{g}}(\xi_{\mathrm{s}},\xi_{\mathrm{l}},\xi_{\mathrm{v}},\xi_{\mathrm{a}})$

are the permeability tensors of concrete for the liquid and gaseous components while $\mu_l = \mu_l(T)$ and $\mu_g = \mu_g(T, \xi_v, \xi_a)$ are their dynamic viscosities, respectively. Finally, $D_v = D_v(T, \bar{\rho}_v, \xi_v, \xi_a)$ denotes the diffusivity parameter of vapour.

4 CONCLUDING REMARKS

A rather general formulation for the analysis of hygro-thermo-mechanical behaviour of concrete at elevated temperatures is presented. The resulting model is capable of describing the following phenomena: the thermoelastic and damaging behaviour of concrete, transport of water, water vapour and air in the porous cement paste, adsorption of water into the cement gel and diffusion of water vapour in the pores. In addition, the model takes into account the phase change between liquid water and water vapour.

Further improvements to the model will include:

- shrinkage of concrete due to the loss of adsorbed and chemically bound water during heating,
- inclusion of a model for transient thermal creep and
- development of a model capable to describe plastic behaviour of concrete, which is mainly due to the relative sliding of the microdamaged material surfaces.

Moreover, numerical implementation of the model is under development.

ACKNOWLEDGEMENTS

This research has been supported by the Academy of Finland, (decision number 210622).

REFERENCES

[1] Z.P. Bažant, M.F. Kaplan, *Concrete at high temperature*, Longman, 1996.

[2] G. Khoury, C. Majorana, P. Kalifa, B. Aarup, M. Giannuzzi, F. Corsi, *The Hiteco Brite project on the Effect of Heat on High Performance Concrete*,

[3] C.E. Majorana, V. Salomoni, B.A. Schrefler, *Hygrothermal and mechanical model of concrete at high temperature*, Materials and Structures, vol 31, 1998, pp. 378–386.

[4] D. Gawin, C.E. Majorana, B.A. Schrefler, *Numerical analysis of hygro-thermal behaviour and damage of concrete at high temperature*, Mechanics of Cohesive-Frictional Materials, vol 4, 1999, pp. 37–74.

[5] D. Gawin, F.Pesavento, B.A. Schrefler, *Modelling of hygro-thermal behaviour of concrete at high temperature with thermo-chemical and mechanical material degradation*, Computer Methods in Applied Mechanics and Engineering, vol 192, 2003, pp. 1731–1771.

[6] K. Maekawa, T. Ishida, T. Kishi, *Multi-scale modeling of concrete performance, integrated material and structural mechanics*, Journal of Advanced Concrete Technology, vol 1, No 2, 2003, pp. 91–126.

[7] U. Diederichs, U.-M. Jumppanen, V. Penttala, *Behaviour of high strength concrete at high temperatures*, Helsinki University of Technology, Division of Structural Engineering, report 92, 1989, 76 p.

[8] J. Lemaitre, J.-L. Chaboche. *Mechanics of Solid Materials*, Cambridge University Press, 1990.

[9] J. Lemaitre. *A Course on Damage Mechanics*, Springer-Verlag, Berlin, 1992.

[10] M. Frèmond, M. Mikkola, *Thermomechanical modelling of freezing soil*, in Ground Freezing 91, Volume 1, Proceedings of the sixth international symposium on ground freezing, 10-12.9. 1991, Beijing, China, pp. 17–24.

[11] J. Hartikainen, M. Mikkola, *General thermomechanical model of freezing soil with numerical application*, in Proceedings of the International Symposium on Ground Freezing and Frost Action in Soils, Luleå, Sweden, April 15–17, 1997, pp. 101–105.

[12] J. Hartikainen, M. Mikkola, *Thermomechanical model of freezing soil by use of the theory of mixtures*, in Proceedings of the Finnish Mechanics Days, J. Aalto, T. Salmi (eds.), Oulu, Finland, September 5–6, 1997, pp. 1–23.

[13] M. Mikkola, J. Hartikainen. *Mathematical model of soil freezing and its numerical implementation.* International Journal for Numerical Methods in Engineering. 2001. Vol. 52, s. 543-557.

[14] P. Jussila, *Thermomechanics of Porous Media - I : Thermohydraulic model for compacted bentonite.* Accepted for publication in Transport in Porous Media.

[15] M. Frémond. *Non-Smooth Thermomechanics*, Springer, Berlin, 2002.

[16] C. Truesdell, R. Toupin. *The classical field theories*, volume III/I of Encyclopedia of Physics, chapter Principles of Classical Mechanics and Field Theory. Springer-Verlag, Berlin, Göttingen, Heidelberg, 1960.

[17] M.A. Biot. *General theory of three-dimensional consolidation.* Journal of Applied Physics, vol 12, 1941, pp. 155–164.

[18] S. Murakami, K. Kamiya. *Constitutive and Damage Evolution Equations of Elastic-brittle Materials Based on Irreversible Thermodynamics.* International Journal of Mechanical Sciences. Vol. 39, 1997, pp. 473-286.

[19] N. Challamel, C. Lanos, C. Casandjian. *Strain-based anisotropic damage modelling and unilateral effects.* International Journal of Mechanical Sciences. Vol. 47, 2005, pp. 459-473.

ASEEN TULIPUTKEN LUJUUSTEKNISEN MITOITUKSEN OHJELMOINTI

Juha Toivola
Mekalyysi Oy
Myllymäentie 5
37960 Sotkia
toivolaj@surfeu.fi

Seppo Moilanen
Patria Weapon Systems Oy
PL 18, Vammaksentie
38201 Vammala
seppo.moilanen@patria.fi

TIIVISTELMÄ

Raskaiden aseiden tuliputkiin kohdistuu sisäballistisen vaiheen aikana pulssimainen painekuormitus. Perinteisesti tuliputken mitoitus on perustunut putkimateriaalin staattiseen lujuuteen. Vanhoilla aseilla ja ampumatarvikkeilla tuliputken käyttöikää on rajoittanut kuluminen. Uusissa asekonstruktioissa on käytetty lujempia materiaaleja ja lujuuden kasvu on myös hyödynnetty putken staattisessa mitoituksessa. Samoin putken sisäpinnan pinnoitusmenetelmillä on vähennetty kulumista. Tällöin putken väsyminen toistuvassa laukauskuormituksessa saattaa tulla määrääväksi tuliputken mitoitusperusteeksi. Putken kestoiälle on kyettävä määrittämään arvio sen väsymislujuuteen perustuvilla laskennallisilla tarkasteluilla. Työssä esitellään tuliputken lujuusteknisen mitoituksen teoreettista taustaa sekä mitoituksen ohjelmointia MATLAB-ympäristössä. Ohjelmaa on käytetty mm. AMOS-kranaatinheitinasejärjestelmän (Advanced MOrtar System) tuliputken mitoitukseen.

TAUSTAA

Raskaiden aseiden tuliputkiin kohdistuu sisäballistisen vaiheen aikana pulssimainen painekuormitus. Maksimipaine vaihtelee putkikoordinaatin (putken pituusakselin) suhteen tavallisimmin siten, että panoskammion painetaso on suurempi kuin paine putken suulla. Ruutikaasun tuottamalle paineelle saadaan arvio sisäballistisen vaiheen laskennallisista simulaatioista tai olemassa oleville aseille ja ampumatarvikkeille tehdyistä mittauksista.

Perinteisesti tuliputken mitoitus on perustunut putkimateriaalin staattiseen lujuuteen. Aseen sallittu käyttöpaine-alue on ollut kokemusperäisen turvamarginaalin (varmuusluvun) verran alhaisempi putken alkavaa myötäämistä vastaavaan paineeseen nähden. Vanhoilla aseilla ja ampumatarvikkeilla tuliputken käyttöikää on rajoittanut kuluminen ja siitä aiheutuva aseen suorituskyvyn heikkeneminen.

Uusissa asekonstruktioissa on käytetty lujempia putkimateriaaleja ja lujuuden kasvu on myös hyödynnetty putken staattisessa mitoituksessa. Samoin putken sisäpinnan pinnoitusmenetelmillä on

vähennetty kulumista. Tällöin putken väsyminen toistuvassa laukauskuormituksessa saattaa tulla määrääväksi tuliputken mitoitusperusteeksi. Putken kestoiälle on kyettävä määrittämään arvio sen väsymislujuuteen perustuvilla laskennallisilla tarkasteluilla kuten esimerkiksi standardin STANAG 4110 painemääritykset edellyttävät (Fatigue Design Pressure, FDP-curve) /1/. Asejärjestelmien käyttöturvallisuuden todentamiseksi nykyisin pyritään testaamaan aseiden kestävyys ja erityisesti väsymiskestävyys laskennallisten tarkastelujen tulosten varmistamiseksi. Aseen konstruktiivisen lujuuden testaus on kuitenkin vain pieni osa käyttöturvallisuuteen liittyvästä koko asejärjestelmän testauksesta /2/.

TULIPUTKEN STAATTINEN LUJUUS

Aseiden tuliputket ovat kohtuullisen sileitä sylintereitä ja voimakkaita geometrisia poikkileikkauksen epäjatkuvuuskohtia on pyritty välttämään. Päistään avoimen (normaalivoima $N = 0$), sisäpuolisella paineella kuormitetun, vakio- ja paksuseinämäisen sylinteriputken kimmoteoriaan perustuvista yhtälöistä voidaan johtaa valitusta vertailujännityshypoteesista riippuen sisäpuoliselle myötöpaineelle sangen yksinkertaiset lausekkeet.

Tuliputkien tapauksessa poikkileikkauksen ulkohalkaisijan muutoksen pienuudesta johtuen "käsilaskentalausekkeet" antavat täysin FEM-laskentatuloksiin yhtyviä tuloksia. Suurempia virheitä aiheutuu todellisuudessa dynaamisen tilanteen käsittelemisestä staattisena sekä painekuormituksen laskennan epävarmuudesta. Painekuormituksen laskentatuloksia (sisäballistiikan ratkaisua) ei voi verrata mittaustuloksiin koko tuliputken matkalla mittaamiseen liittyvien ongelmien vuoksi.

Lineaarisen kimmoteorian mukaan sisäpuolisen paineen kuormittaman paksuseinäisen ja autofretoimattoman putken poikkileikkauksen jännitykset tasomuodonmuutostilassa ovat

$$\sigma_r = -p\frac{(d_u/d)^2 - 1}{(d_u/d_s)^2 - 1} \qquad \text{(radiaalisuuntainen jännitys)}$$

$$\sigma_t = p\frac{(d_u/d)^2 + 1}{(d_u/d_s)^2 - 1} \qquad \text{(kehän suuntainen jännitys)}$$

joissa p on sisäpuolinen paine, d_u putken ulkohalkaisija, d_s putken sisähalkaisija ja d laskentahalkaisija, eli tarkastelupistettä vastaava halkaisija. Kehän suuntainen jännitys, joka on aina vetoa, saa suurimman arvonsa putken sisäpinnalla. Radiaalisuuntaisen jännityksen itseisarvo on myös suurimmillaan putken sisäpinnalla.

TULIPUTKEN VÄSYMISTARKASTELU

Suurikaliiperisille aseille on tyypillistä kuormituskertojen alhainen lukumäärä, käyttöiän aikana tavallisesti satoja tai tuhansia laukauksia. Suurimmilla panoksilla rasitustaso on lähellä putken myötäämistä. Käytettäviä panoksia ja ammuksia on myös useita, jolloin aseen rasitustaso vaihtelee. Tällöin olisi tarpeen määrittää jokin mitoituksessa käytettävä peruslaukaisuyhdistelmä, johon muiden laukausyhdistelmien rasituksia verrataan, tai sitten tehdä mitoitus suurimman tyypillisesti käytettävän panoksen mukaan. Toisaalta staattisen mitoituksen vaatimukset saattavat johtaa riittävään kestoikään, mikä myös on pystyttävä toteamaan. Epävarmuutta putken rasitusten määrittämiseen aiheutuu mm. rihlojen muokkautumisesta ammuksen johtorenkaaseen, mikä saattaa kasvattaa putken laukausrasituksia /3/.

Laukauksessa esiintyy lämpökuormitus, joka aiheuttaa putken sisäpinnalle lämpösäröverkon jo varhaisessa vaiheessa, tyypillisesti muutamien kymmenien laukausten jälkeen. Lämpösäröverkon särösyvyys on n. (0,2...0,5) mm. Tätä pienempiä säröjä ei kannata etsiä putken valmistusvaiheen NDT-tarkastuksessa. Putken rihlaus aiheuttaa kehän suunnassa geometrisia epäjatkuvuuskohtia, jotka toimivat hyvänä kasvualustana säröille rihlatuilla tuliputkilla. Kirjallisuudesta löytyy runsaasti valokuvia, joiden mukaan putken suuntaiset säröt ovat alkaneet kasvaa juuri rihlauksen juuresta. Koska kehän suuntainen vetojännitys on suurimmillaan putken sisäpinnalla, väsymisen kannalta kriittisin kohta on putken sisäpinta, jos putken ulkopinnalla ei ole voimakkaita geometrisia epäjatkuvuuskohtia.

Tuliputkien materiaaleina käytetään taottuja erikoisteräksiä, joilla on suuri lujuus ja lujuuteensa nähden hyvät sitkeysarvot myös matalissa lämpötiloissa. Murtumisparametrien määrittäminen näille teräksille on vaikeaa, esim. murtumissitkeyden K_{IC} sijasta joudutaan tyytymään tasojännitystilassa (tai lähellä sitä) saatavaan K_Q -arvoon, koska testikappaleen koko ei täytä tasomuodonmuutostilan edellytyksiä. Tuliputkessa materiaalin ominaisuudet saattavat olla taonnasta johtuen suuntautuneita, jolloin materiaaliltaan lähinnä isotrooppisesta ja homogeenisesta koekappaleesta mitattujen materiaaliparametrien käyttämiseen liittyy epävarmuutta.

Perinteiset high cycle –alueella toimivat nimellisiä jännityksiä käyttävät väsymisanalyysimenetelmät ovat käyttökelpoisia, mikäli nimelliset jännitykset lovenvaikutuslukuineen kuvaavat kriittisen kohdan jännitystason oikein. Näin on, mikäli kriittisen kohdan geometria ei muutu väsymisen aikana. Väsymissärön synnyttyä geometria muuttuu (jännitysten kannalta), mutta mikäli väsymisiästä suurin osuus kuluu säröjen ydintymiseen, väsymisiän ennusteeseen tätä kautta tuleva virhe on pieni.

Mikäli rakenne sisältää valmiiksi alkusäröjä tai säröt ydintyvät jo varhaisessa vaiheessa, väsymisiän määrää särön kasvu kriittisen suuruiseksi. Sen määrittämiseksi on tällöin luontevinta käyttää murtumismekaniikan särön kasvun laskentaan perustuvia menetelmiä. Käytännön kannalta ongelmaksi muodostuu käytettävän särönkasvulain parametrien määrittäminen ja niiden arvojen luotettavuus sekä kriittisen särön koon määritys. Kestoikäennusteet ovat erittäin herkkiä särönkasvulain parametrien muutoksille ja toisaalta lineaarista murtumismekaniikkaa sovellettaessa murtumissitkeyden mittaaminen on sitkeille materiaaleille hankalaa. Särön kasvun laskenta on sinällään yksinkertaista ja nopeaa, jolloin erilaisten herkkyysanalyysien tekeminen materiaaliparametrien suhteen on helppoa.

Tuliputken tapauksessa syntyvät säröt ovat tyypillisesti tuliputken suuntaisia, jolloin särön kuormitus on peruskuormitustavan I mukainen. Koska tuliputki on pitkä, sen poikkileikkaukseen syntyy varsin tarkasti tasomuodonmuutostila (ja lisäksi palkkiteorialla varsin tarkasti laskettavissa oleva taivutustila). Tasomuodonmuutostilassa särön kärjen edessä olevan plastisen alueen koko on pienempi kuin tasojännitystilassa, jolloin lineaarisen murtumismekaniikan käyttö on perustellumpaa.

Todennäköisesti kaikkien palveluskäytössä olevien panosten painerasituksesta laskettu jännitysintensiteettikerroin K_I ylittää kynnysarvon ΔK_{TH}, jonka alapuolella särö ei kasva. Alkusäröjen kuormitus on laukausrasituksessa tykyttävää, ja mikäli särön kasvulain materiaaliparametrit mitataan tykyttävällä kuormituksella, keskijännityksen vaikutus on koetuloksissa mukana laukausrasitusta vastaavalla tavalla. Tällöin särön kasvun laskemiseen kannattanee käyttää Paris'n kaavaa /7/

$$\frac{da}{dN} = C(\Delta K_I)^m$$

Särön kasvun laskennan lopetuskriteerit ovat:

- intensiteettikerroin saavuttaa murtumissitkeyden arvon, mikä johtaa epästabiiliin särönkasvuun
- särön edessä oleva kannas on läpiplastisoitunut
- intensiteettikerrointa ei voida laskea

Kaksi ensimmäistä lopetuskriteeriä ovat itsestäänselviä, sen sijaan kolmas vaatii perusteluja. Kirjallisuudesta löytyvät intensiteettikertoimen käsikirjakaavat ovat tyypillisesti numeerisesti lasketuihin tuloksiin sovitettuja lausekkeita. Niiden ilmoitettu voimassaoloalue loppuu särön syvyyden ollessa n. (75...80)% seinämän paksuudesta. Laskennan jatkaminen tämän rajan ylitse on riskialtista, joten käytännön työssä, erityisesti käsitellyssä tapauksessa, on varminta lopettaa laskenta ilmoitetun voimassaoloalueen rajalle ja todeta, että särö on kasvanut (laskentakaavoihin nähden) kriittisen syvyiseksi.

Toinen kirjallisuudesta löytyvien intensiteettikertoimen käsikirjakaavojen ongelma on, että tuliputken tapauksessa halkaisijasuhde d_u/d_s osoittautui olevan usein suurempi kuin käsikirjakaavojen voimassaoloalue sallii. Ne on ilmeisesti laadittu aineen siirtoon esim. ydinvoimaloiden yhteydessä tarkoitetuille paksuhkoseinämisille putkille. Tuliputken seinämä on kuitenkin niihin nähden selvästi paksumpi erityisesti panoskammion läheisyydessä.

Intensiteettikertoimen laskentaan liittyvät ongelmat voinee ratkaista esim. käyttämällä intensiteettikertoimen laskentaan painofunktiomenetelmää /8, 9/. Tässä selostetussa työssä siihen ei kuitenkaan lähdetty.

TULIPUTKEN LUJUUSTEKNISEN MITOITUKSEN OHJELMOINTI

Tuliputken lujuustekninen mitoitus sisältää staattisen mitoituksen ja väsymismitoituksen. Koska mitoitus ei sisällä raskasta laskentaa särön kasvua ehkä lukuunottamatta, laskenta voitaisiin suorittaa perinteisenä käsilaskentana. Toisaalta uusien ampumatarvikkeiden käyttöönotto sekä herkkyysanalyysit materiaaliparametrien suhteen vaativat laskennan toistamista useita kertoja. Tuliputken geometriset tiedot pysyvät näissä laskennoissa samoina. Virheiden välttämiseksi, olemassa olevan tiedon hallitsemiseksi ja käyttömukavuuden vuoksi Patria Weapon Systems Oy:ssä katsottiin tarpeelliseksi ohjelmisto, jossa tuliputken geometria syötetään vain kerran ja laskenta voidaan toistaa erilaisilla lähtöarvoilla helposti. Patria Weapon Systems Oy:n ja Mekalyysi Oy:n välisissä neuvotteluissa päädyttiin mitoitusohjelmiston toteuttamiseen MATLAB-ympäristössä.

Aikana jolloin tietokoneita käytettiin tekstipohjaisilla päätteillä, kynnys tässä kuvatun mitoitusohjelmiston tekemiseen olisi ollut pienempi, koska ohjelmiston vaatima syöttödata olisi annettu (ulkopuolisen kannalta varsin kryptisen näköisenä) tekstitiedostona ja laskentatulokset olisi tulostettu tekstitietiedostoon tai päätteen ruutuun. Graafisten käyttöjärjestelmien yleistyessä kynnys pienehköjen ohjelmien tekoon on noussut, koska niiltäkin vaaditaan graafisia käyttöliittymiä ja graafisia tulostuksia. Teknisen laskennan yhteydessä näyttää siltä, että MATLAB on tällä hetkellä paras ratkaisu laskennan suorittamiseen, graafisen käyttöliittymän ohjelmoimiseen ja graafiseen tulostukseen, mikäli ohjelmoitava laskenta on selvärajainen ja laajuudeltaan pienehkö. MATLABia käyttäen ohjelmoijan ei tarvitse syventyä minkään ohjelmointikielen hienouksiin, vaan riittävän näyttävää ja samalla toimivaa jälkeä saa aikaan varsin nopeasti. Muitakin vaihtoehtoja tietysti on.

Tuliputken lujuustekniseen mitoitusohjelmaan annetaan syöttöarvoina:

- materiaaliparametrit: kimmomoduuli E, Poisson kerroin ν, murto- ja myötölujuus (R_m, R_e), murtumissitkeys K_{IC}, Paris'n kaavan materiaaliparametrit C ja m
- geometriset suureet: putkigeometria, alkusärön koko
- kuormituskertojen määrä N ja tunnettu paine
- aseen sisäballistisiin simulointeihin perustuva laskennallinen suunnittelupainekäyrä DP, jota ei käytetä lujuuslaskennassa, mutta ehdon DP < SMP tulee toteutua /1/.

Ohjelma laskee syöttöarvojen perusteella:

- tuliputken staattiset painerajat (SMP) materiaalin myötörajaan perustuen
- väsymispainerajat FDP(N) määrätylle kuormituskertojen (laukausten) lukumäärälle
- kuormituskerrat $N(p_i)$ tunnetulle painekuormalle p_i.

Väsymistarkastelut perustuvat lineaariseen murtumismekaniikkaan kuten edellä on esitetty ja tarkasteltavat putken sisäpinnan särötapaukset ovat:

- äärettömän pitkä aksiaalisärö (rajoitus putkigeometrialle $d_u < 1{,}20 d_s$)
- puolielliptinen särö (rajoitus putkigeometrialle $d_u < 1{,}25 d_s$ sekä särön muodolle $a < c < 20a$, a särön syvyys, c särön pituus)
- reunasärö puoliäärettömässä levyssä.

Viimeinen tapaus piti lisätä, koska käytetyn äärettömän pitkän aksiaalisärön jännitysintensiteettikertoimen laskentakaavan rajoitus oli liian rajoittava tuliputkien mitoitustarpeita ajatellen. Sen käyttö paksuseinäisen putken sisäpuolisen aksiaalisärön yhteydessä lienee perusteltua, kun särön syvyys suhteessa seinämän paksuuteen on pieni.

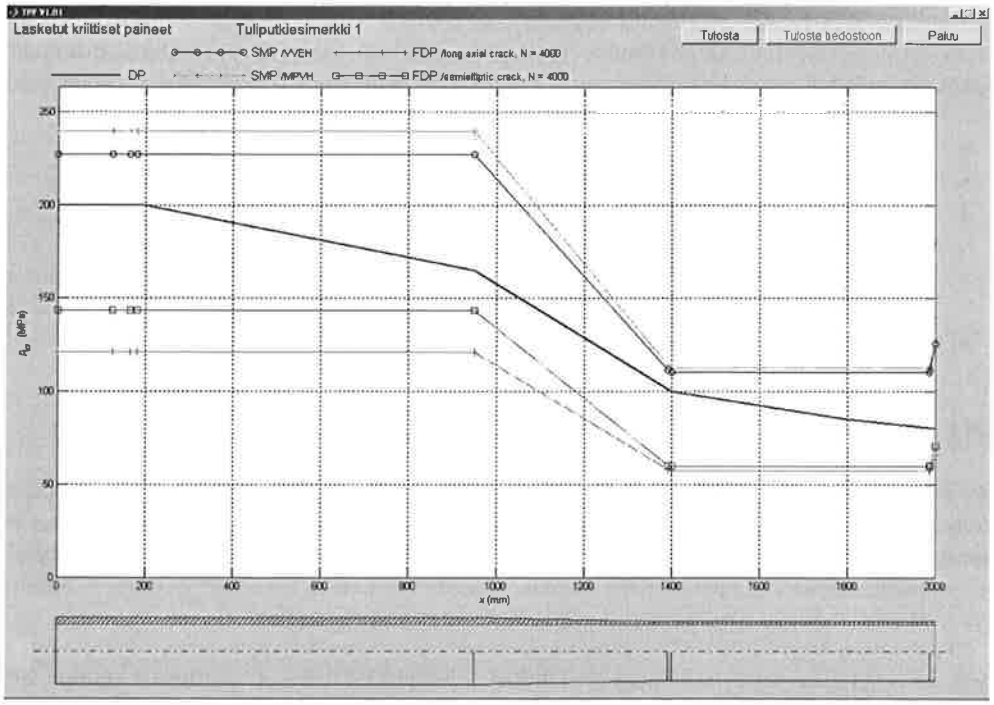
Laskennan tuloksina tulostetaan

- Stanag 4110 mukaiset painekäyrät putken pituudelle:
  - staattiset myötöpaineet (SMP) jännnityskriteerin mukaan putkimatkalle
  - väsymispainekäyrät (FDP) valitulle särömuodolle ja annetulle laukausmäärälle putkimatkalle
- kriittistä särökokoa vastaava kuormituskertojen määrä putkimatkalle ($N(p_i)$ nimellinen sallittu kuormituskertojen määrä annetulla paineella)
- särön syvyyden ja jännitysintensiteetin kehittyminen laukausmäärän funktiona halutuissa tarkastelupisteissä
- särön kasvun laskennan lopetuskriteerit.

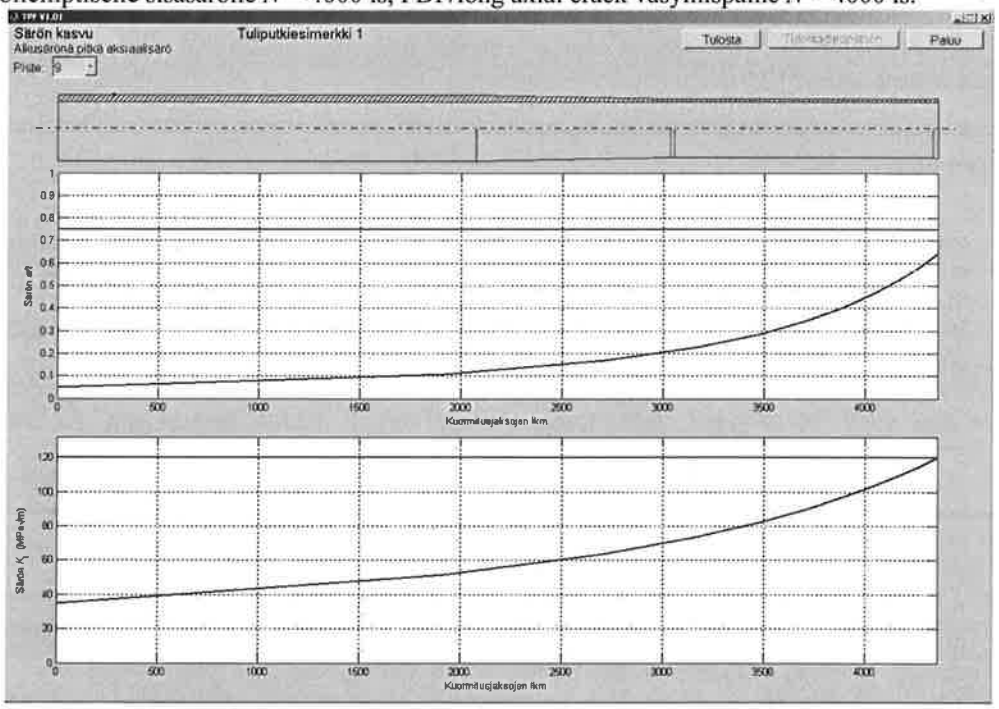
Kuvissa (Kuva 1 ja Kuva 2) on esitetty esimerkki tuliputken kriittisten paineiden sekä särön ja jännitysintensiteetin kasvun graafisesta tulostuksesta.

Alun perin ohjelmoinnin toteutusta ajatellen eniten työtä vaativaksi osuudeksi arveltiin graafisen käyttöliittymän tekemistä. Yllättäen särön kasvun laskennan toteuttaminen luotettavasti osoittautui työläimmäksi. Tässä yhteydessä ilmeni ainakin seuraavia ongelmia:

- Joissakin tapauksissa särön kasvu pysähtyi käytännöllisesti katsoen kokonaan, kun särön syvyys ylitti laskentakaavojen voimassaoloalueen rajan. Tästä johtuen ohjelman pitää tarkkailla useampaa laskennan lopetuskriteeriä, mikä monimutkaistaa ohjelmaa. Alun perin ajatuksena oli, että jännitysintensiteettikertoimen ja murtumissitkeyden vertailu riittää.
- Kirjallisuudessa esitetyt jännitysintensiteettikertoimien laskentalausekkeet ja niiden ilmoitetut voimassaolorajat osoittautuivat epäluotettaviksi. Laskentalausekkeiden ja niiden voimassaoloalueiden luotettavuuden varmistaminen vaati ylimääräistä työtä.
- Integroitaessa Paris'n kaavaa MATLABin automaattinen askelpituuden valinta aiheuttaa joskus puolielliptisen särön tapauksessa särön syvyyden kasvamisen sen pituutta suuremmaksi, jolloin jännitysintensiteettikerroin menee kompleksiseksi ja laskenta kaatuu.



Kuva 1. Tuliputken kriittisten painearvojen laskentaesimerkki. Kuvan painekäyrät ylhäältä alas: SMP/MPVH myötöpaine maksimipäävenymähypoteesi, SMP/VVEH myötöpaine vakiovääristymisenergiahypoteesi eli von Mises, DP suunnittelupaine (< SMP), FDP/semielliptic väsymispaine puolielliptiselle sisäsärölle $N = 4000$ ls, FDP/long axial crack väsymispaine $N = 4000$ ls.



Kuva 2. Särönsyvyyden ja jännitysintensiteetin kasvu annetulla painekuormalla p_i kuormitusjaksojen N (laukauksien) funktiona. Laskentapoikkileikkaus on osoitettu pallukalla putken kuvassa.

Tässä selostetussa työssä syntynyttä ohjelmaa on käytetty Patria Weapon Systems Oy:ssä mm. AMOS®-kranaatinheitinasejärjestelmän (Advanced MOrtar System) tuliputkien mitoitukseen. Käyttökokemuksien perusteella ohjelmasta on löytynyt joitakin jatkokehitystä vaativia piirteitä, mm:

- tietojen syöttöä voisi kehittää
- laskettavat tulokset olisi voitava valita kulloisenkin tarpeen mukaan
- särön edessä olevan kannaksen läpiplastisoituminen olisi lisättävä laskennan lopetuskriteeriksi
- geometrisista epäjatkuvuuskohdista, kuten rihlauksesta tai putken seinämän läpiporauksista, johtuvien jännityskeskittymien huomioonottaminen karkealla tasolla olisi lisättävä
- jännitysintensiteettikertoimen laskennan rajoitteista olisi päästävä eroon.

YHTEENVETO

Kokemuksina tässä selostetusta työstä esiin nousee kaksi asiaa. Ensinnäkin pienehköjen teknistä laskentaa suorittavien ohjelmien ohjelmointi MATLAB-ympäristössä on helppoa mukaan lukien graafinen käyttöliittymä ja graafinen tulostus. Toiseksi kriittinen suhtautuminen laskentatuloksiin, tässä tapauksessa jännitysintensiteettikertoimen arvoihin, on aina paikallaan. Erityisesti ohjelmointityössä laskentatuloksien oikeellisuuden takaaminen vaatii paljon työtä.

Ohjelman käyttäjän sekä laskennallisten tulosten loppukäyttäjän eli puolustusvoimien asealan asiantuntijoiden kannalta Mekalyysi Oy:n toimittaman ohjelman käyttöominaisuudet ja erityisesti laskentatuloksien selkeä "insinöörimäinen" esitystapa ylittivät alkuperäiset odotukset ja tavoitteet /6/. Ohjelmisto on osoittautunut helppokäyttöiseksi ja laskennalliset tulokset käyttökelpoisiksi uusien asejärjestelmien suunnittelussa ja testauksessa sekä vanhojen asejärjestelmien suorituskykyarvioineissa. Käyttäjän kannalta nopea ja luotettava laskenta on sekä taloudellisesti että laskentaresurssien käytön suhteen ylivoimaisen tehokasta etenkin nykysuuntauksen mukaiseen raskaaseen ja aikaa vievään särön FEM-mallinnukseen verrattuna.

VIITTEET

1. Stanag 4110. *Definition of pressure terms and their interrelationship for use in the design and proof of cannons or mortars and ammunition.* North Atlantic Treaty Organization, Standardization Agreement, Stanag.
2. Henttinen, S. *Aseiden ja ampumatarvikkeiden rakenteellinen turvallisuus.* Julkaisussa: MPDI 40v 1963-2003. Suomen Alfatec Oy. 2003. ss. 17,...,21.
3. Edwards, T.D. *Projectile driving band interactions with gun barrels.* In: Gun Tubes Conference 2005, 10-14 April, 2005. Keble College Oxford. RMCS Shrivenham. Presentation S7/P28. 7p.
4. Toivola, J. *Teoriaopas. Thick Pipe Fatigue TPF Versio 1.0.* Mekalyysi Oy. Sotkia. 2003. 14 s.
5. Toivola, J. *Käyttöohje. Thick Pipe Fatigue TPF Versio 1.0.* Mekalyysi Oy. Sotkia. 2003. 11 s.
6. Moilanen, S. *Tuliputken väsymisikälaskennasta.* Esitelmämuistio Patria Weapon Systems Oy, PDM 18900. Vammala. 2003. 17 s.
7. Ikonen, K., Kantola, K. *Murtumismekaniikka.* Otakustantamo. Helsinki. 1986.
8. Zheng, X.J., Glinka, G., Dubey, R.N. *Calculation of stress intensity factors for semielliptical cracks in a thick-wall cylinder.* Int. J. Pres. Ves. & Piping, Vol. 62, 1995, ss. 249-258.
9. Kiciak, A., Glinka, G., Burns, D.J. *Calculation of stress intensity factors and crack opening displacements for cracks subjected to complex stress fields.* In: Gun Tubes Conference 2002, 15-18 September, 2002. Keble College Oxford. RMCS Shrivenham. Presentation S4P23. 8p.

FRACTURE AND J-INTEGRAL ASSESSMENT OF A WELDED CFSHS K-JOINT TESTED AT -40 °C

T. BJÖRK, S. HEINILÄ and G. MARQUIS
Laboratory of Fatigue and Strength
Lappeenranta University of Technology
P.O. Box 20
FIN-53851 Lappeenranta, FINLAND

ABSTRACT: The deformation, ultimate load capacity and fracture behaviour of full-scale welded K-joints fabricated from cold-formed rectangular hollow sections has been studied both numerically laboratory testing at -40 °C. The experimental program revealed that the primary failure mode was ductile tearing of the chord flange at the weld toe of the tension brace flange. In some cases the ductile tearing led to brittle fracture. Finite element based *J*-integral assessment of an advancing crack was combined with J-Δa curves measured at -40 °C. Assessment revealed that crack advance, once initiated, would be expected to continue at nearly constant load. This assessment method was found to be effective in modelling the ultimate ductility of the joint.

1 INTRODUCTION

Rectangular hollow sections (RHS) are widely used in load-carrying structures due to their good load transfer behaviour and aesthetic form. Experimental tests and numerical analyses on the capacity and fracture of structural hollow section joints are regularly reported, however, most of these investigations involved joints fabricated using hot-formed structural hollow section members or testing has been at room or elevated temperatures [Wardenier 1982 and 2002, Packer et al. 1992, Packer & Henderson 1997, Puthli 1998, Vainio 2000, Zhao et al. 2005]. Full-scale sub-zero temperature tests are complex and relatively expensive, especially if a significant number of joint geometries and section profiles are of interest. Therefore, only limited test data is available for cold-formed structural hollow section joints at sub-zero temperatures [Niemi 1990, Niemi 1996].

Finite element (FE) modelling can be an effective tool for studying the nonlinear load deformation behaviour of RHS joints. For cold formed structural hollow sections (CFSHS), the yield strength of the corner is typically higher than that of the adjacent flat faces because of the greater degree of cold forming. True stress-strain curves determined by material tests from both the flat face and the corner region can be used to develop nonlinear material models for the FE analyses. It was found that the load-deformation behaviour predicted by FE analysis was in good agreement with that measured in the laboratory experiments, However, the ultimate strength or ultimate deformation capacity of a joint based on FE analysis is strongly dependent on the choice of the failure mode. Previous work showed that the simple von Mises theory over predicted the deformation at failure [Björk 2005, Björk et al. 2003a, Björk et al. 2005].

In practice, sufficient load carrying capacity means that the joint can withstand the required design load with a sufficiently low failure probability. The design load can be assessed using design guidance documents [EN 2003], where the capacity is based on the nominal yield strength of the parent metal. Even though load carrying capacity is the most significant design criterion, deformation capacity is an extremely relevant parameter when considering structural safety. In order to fulfil the requirements for the

deformation capacity, the joint must have sufficient plastic deformation before final failure. The concept of "sufficient deformation capacity" of the RHS-joint is not standardized, but a widely adopted practice is to ensure that the plastic deformation / chord width-ratio $\delta_p/b_0 \geq 0.5\%$ for joints where the widths of the brace and chord members are equal, $\beta = 1.0$, and $\delta_p/b_0 \geq 1.0\%$ for joints with $\beta < 1.0$ [Wardenier 1982].

The current study focuses on the use of nonlinear FE analysis of a K-joint fabricated from cold formed rectangular sections. Elastic-plastic analysis was used to model the load-deformation behaviour of the joint. Additionally, *J*-integral values at several crack depths were numerically computed. *J-Δa* crack resistance curves measured for the material near the weld in the through thickness direction at -40 °C were experimentally measured using standard procedures. *J*-integral values were used in conjunction with the *J-Δa* curve in order to asses both the ultimate load carrying capacity of the joint as well as the ultimate deformation capacity.

2 EXPERIMENTS

2.1 Full-scale tests

Quasi-static tests of full-scale symmetric K-joints at ambient and sub-zero temperatures have been performed. The gap type K-joint was used because it is very common joint type in truss and other types of structures. Joint geometries for which minimum deformation capacity was expected were chosen for testing. In order to restrict the deformation capacity of the joint, the gap between the brace members was chosen to be the minimum value allowed in the design recommendations. The material of test tubes had a nominal yield strength f_y = 355 MPa. A schematic of the test specimen is shown in Figure 1 and a typical test specimen is shown in Figure 2.

More that 30 sub-zero K-joint tests were performed. The joint chosen for detailed FE analysis had chord dimensions 100x100x6 (b_0 x h_0 x t_0 [mm]), brace dimension 50x50x5, a gap of 8 mm and an angle between chord and braces of 50°. The test temperature was -40 °C. The specimen chord was fixed in a frame and preloaded in tension to 460 kN by hydraulic jacks. Pinned boundary conditions were used for the brace ends. A schematic of the test frame is given in Figure 3. More full details of the test program are reported elsewhere [Björk 2005, Björk et al. 2003a]. The measured experimental ultimate load was 365 kN which was in good agreement to the design code value [EN 2003]. Measured joint deformation at fracture, δ_p/b_0, was 1.5%.

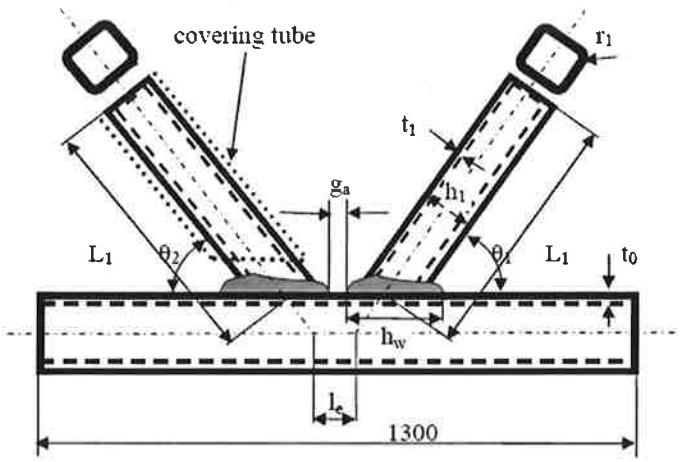


Figure 1. Schematic of K-joint test specimen.

Figure 2. A typical K-joint test specimen.

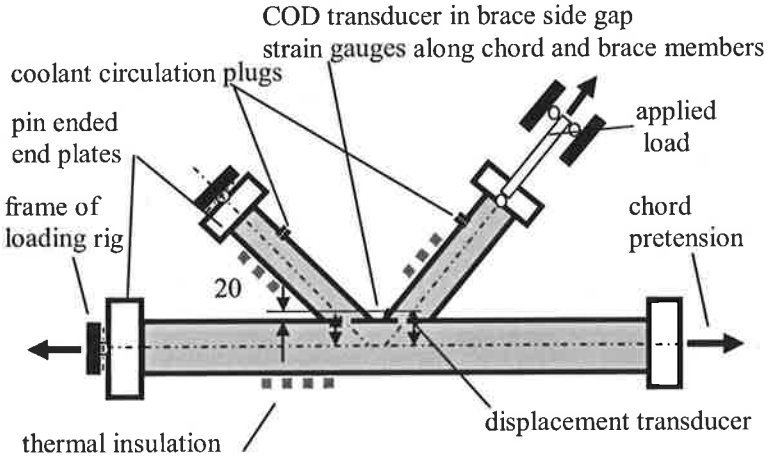


Figure 3. Schematic illustration of K-joint test arrangement.

2.2 Fracture surface evidence

The primary failure mode for many of the joints tested at sub-zero temperatures was ductile tearing of the chord flange followed by brittle fracture. Failure started from the weld of the tension brace at the chord flange side of the weld toe. The fracture initiated at the location where, according to FE analysis, the stresses of the gap were greatest. The failure surface in the chord flange was normal to the tension brace, i.e., at an angle of approximately 50° with respect to the longitudinal axis of the chord.

The fracture initiated by ductile tearing and then, as the crack size increased, the fracture type changed to brittle. This is seen in Figure 4. In the ductile zone the surface was rougher and appears as the darker region in the photograph. The brittle surface has a lighter shade and a typical chevron pattern. The failure modes observed using SEM agreed well with the macro-level observation concerning the change of failure mechanism.

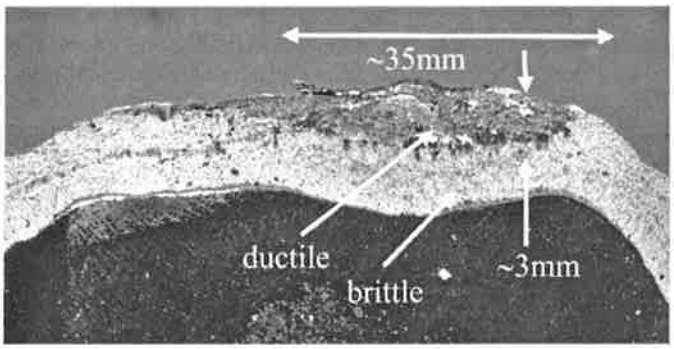


Figure 4. Failure surface of a K-joint at -40°C.

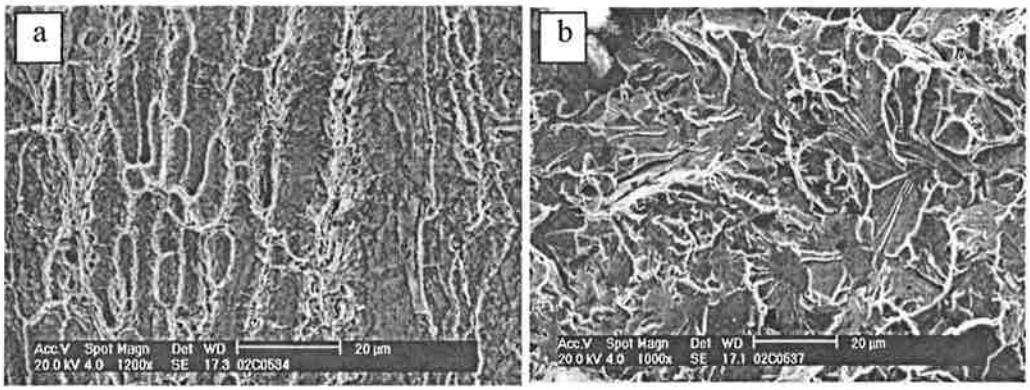


Figure 5. Ductile (a) and brittle (b) failures defined by SEM.

The failure progressed by ductile tearing until the crack depth through the wall of the chord was approximately 3 mm. The corresponding crack width along the chord surface was about 35 mm. After this, the fracture type changed to brittle and continued until complete separation of the chord member. Similar types of failures were observed in numerous K-joints. The crack length at which the fracture behaviour changed from ductile to brittle varied from one joint to another but the mechanism remained constant. In some cases the critical crack depth was only a fraction of the chord wall thickness while in other cases a full through thickness ductile crack was observed. In those cases where the critical crack was through thickness, the surface crack length was approximately equal to the width of the brace member [Björk 2005].

2.3 *J-Δa* tests

J-Δa crack resistance test specimens (5x10x40 mm) were cut from the same RHS from which the chord of the K-joint was taken. The position from which the specimens were cut is shown in Figure 6. In order to simulate the weld, a cladding operation was performed using a one-pass MAG process. The same weld parameters were used as those used during K-joint fabrication. The cladding was applied to the cut surface of the specimen. After cladding, a pre-notch was machined in the specimen followed by fatigue pre-cracking to achieve a sharp crack of the desired dimension.

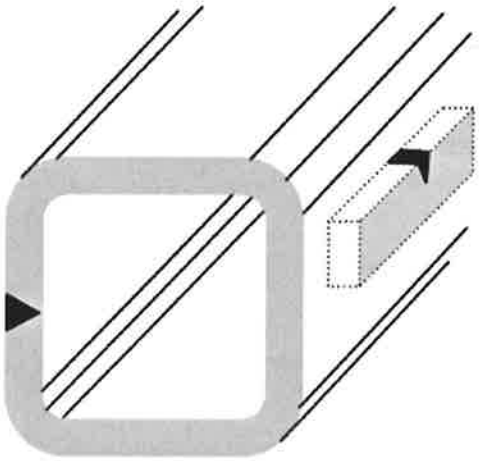


Figure 6. Location of J-Δa specimens taken from the RHS.

The tip of the fatigue pre-crack was about 1 mm from the fusion boundary as seen in Figure 7. The fusion boundary was located near the specimen surface at a depth similar to that observed in the welded K-joint. Specimens were pre-cracked at ambient temperature and tested at -40 °C according to ASTM E 1921-03 [ASTM 2002].

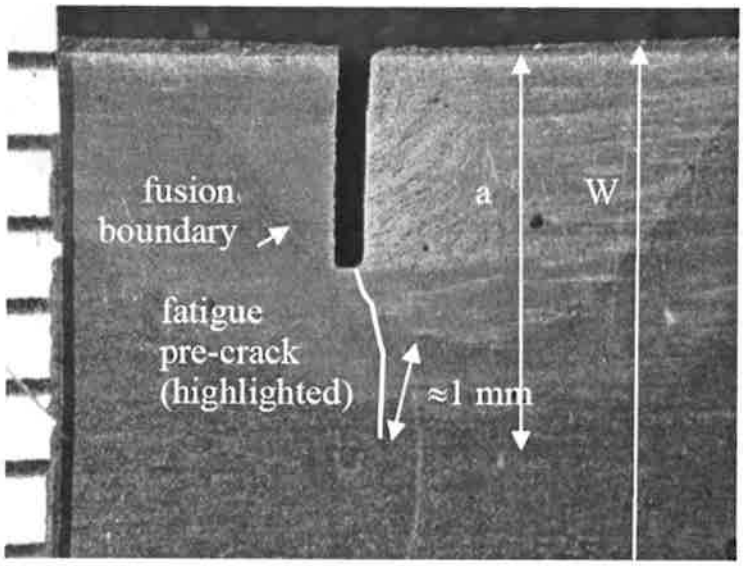


Figure 7. An example of a pre-cracked J-Δa test specimen.

3 THE J-INTEGRAL

Because the experimental program revealed that ductile tearing followed by brittle failure was observed for many welded K-joints at sub-zero temperatures, it was decided to include ductile crack growth in the analysis model using a J- integral fracture mechanics approach. The J-integral as developed by Rice for elastic materials generally describes the flow of energy into the crack tip region [Rice & Tracey 1969]. The theoretical load carrying capacity is based on J-integral value defined as

$$J = \int_0^{\varepsilon} W \mathrm{d}y + \int_s \left[t_x \frac{\mathrm{d}u_x}{\mathrm{d}x} + t_y \frac{\mathrm{d}u_y}{\mathrm{d}x} \right] \mathrm{d}s \tag{1}$$

where u_i are the nodal displacements, s is the path length and W is the strain energy density which can be divided into elastic and plastic component

$$W = W_e + W_{pl} = \int_0^{\varepsilon_{ij}} \sigma_{ij} \mathrm{d}\varepsilon_{ij} + W_{pl} \tag{2}$$

In the case of FEA, the plastic strain energy density W_{pl} can be obtained directly within the FE software. The tractions t_x and t_y are

$$t_x = \sigma_x \overline{n}_x + \sigma_{xy} \overline{n}_y \tag{3}$$

$$t_y = \sigma_{xy} \overline{n}_x + \sigma_y \overline{n}_y \tag{4}$$

where $\overline{n}_x$ and $\overline{n}_x$ are the normal vector x- and y-components.

The procedure for defining the critical load, F_{crit}, and subsequently the critical deformation capacity is based on the EPRI approach [Kumar & al.1984]. In this method the calculated elastic-plastic J-integral is compared to J_R –values obtained from material test results. The procedure is illustrated schematically in Figure 8.

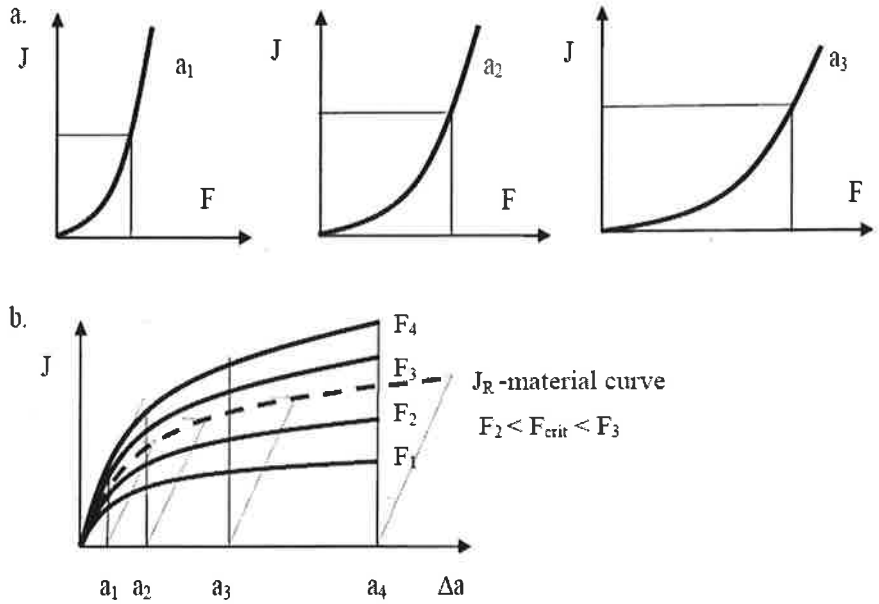


Figure 8. a) Numerically defined F and J–integral dependence for a structure and b) critical load F_{crit} assessment procedure.

The applied force F and J–integral dependence of a joint is defined using FE analysis for several crack sizes, a_1, a_2, a_3, etc., see Figure 8a. From these curves the J-Δa relationships for alternate force levels, F_1, F_2, F_3, F_4, etc., can be defined as shown in Figure 8b. By plotting the J_R curve from material tests on the same diagram, the critical load F_{crit} can be estimated.

It should be noted that the shape of the constant force J-Δa curves for a structure is highly dependent on the structure itself. A structure for which the crack growth is predominantly load-controlled will have a generally positive curvature. While the curve for a structure in which the crack growth is displacement controlled will have a negative curvature or even a negative slope.

4 FINITE ELEMENT CALCULATIONS

4.1 Load-deformation evaluation

The program MSC MARC was chosen for the FE load-deformation analysis. Eight node linear solid elements and 4-node linear tetra elements were used. A solid model was used in order to take into account the real joint geometry with the welds. A Ramberg-Osgood type true σ–ε material model was for the analysis. Details of this analysis have been reported previously [Björk 2005, Björk et al. 2003b]. A comparison of the experimentally and the numerically predicted load-deformation behaviour is shown in Figure 9.

The FE analysis provided good estimates of the elastic–plastic joint behaviour of the analysed joints. However, FE analysis tended to overestimate the ultimate deformation and load carrying capacity. This is mainly due to the von Mises failure hypotheses used in the analysis. This failure hypothesis does not take into consideration the void nucleation or crack growth that lead to ductile failure in a joint. Experimentally, joints failed before the local strains reached the measured critical strain value obtained from the coupon tests.

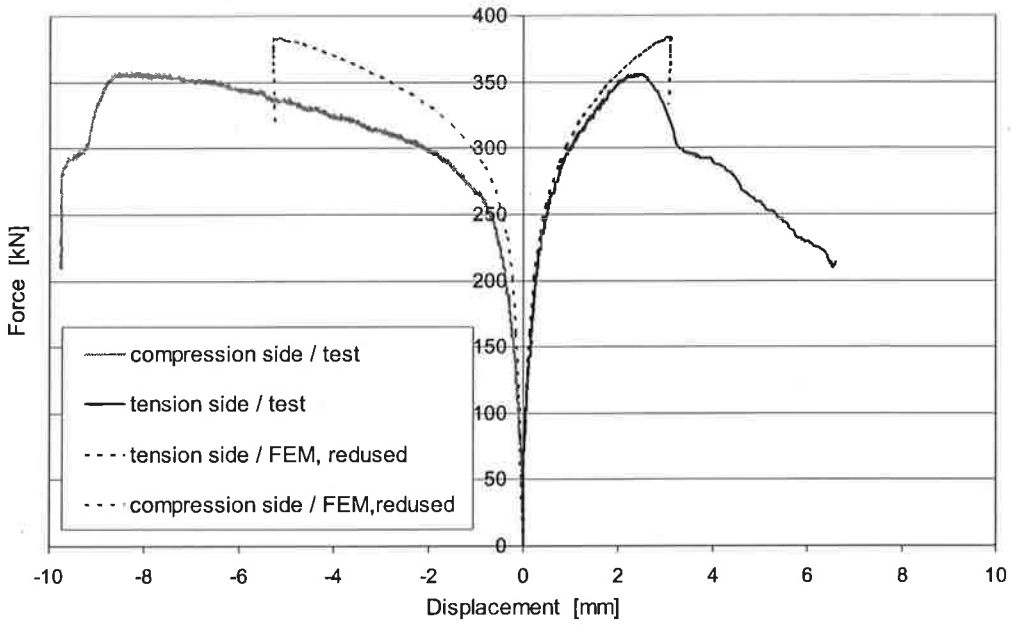


Figure 9. Measured and numerically simulated load displacement behaviour of a K-joint.

4.2 J-integral evaluation

Quasi-static tests of full-scale symmetric K-joints at ambient and sub-zero temperatures were simulated using ANSYS. Measured geometrical data from the laboratory tested K-joint was used in the model. The finite element model of a K-joint was solved for several different crack depths and the *J*-integral was computed at numerous load steps for each crack depth. Identical FE models were used except that the crack depth varied from 0.1 mm to 3 mm. Material properties for the brace and the chords were based on the measured material data. Elastic linear hardening material models were assumed.

Due to the geometric symmetric and the loading condition, ½ of the joint was modelled. The boundary conditions, applied loads and the element model used to define the *J*-integral in the critical region of the K-joint of interest are shown in Figure 10. Boundary conditions were assumed to represent the laboratory test rig boundary conditions. Modelling of the crack and crack tip required a very fine mesh so the FE models were large. For this reason only the regions near the crack were modelled using solid elements while shell elements were used elsewhere. The boundary conditions required some beam elements and spider web meshing on the end of one chord.

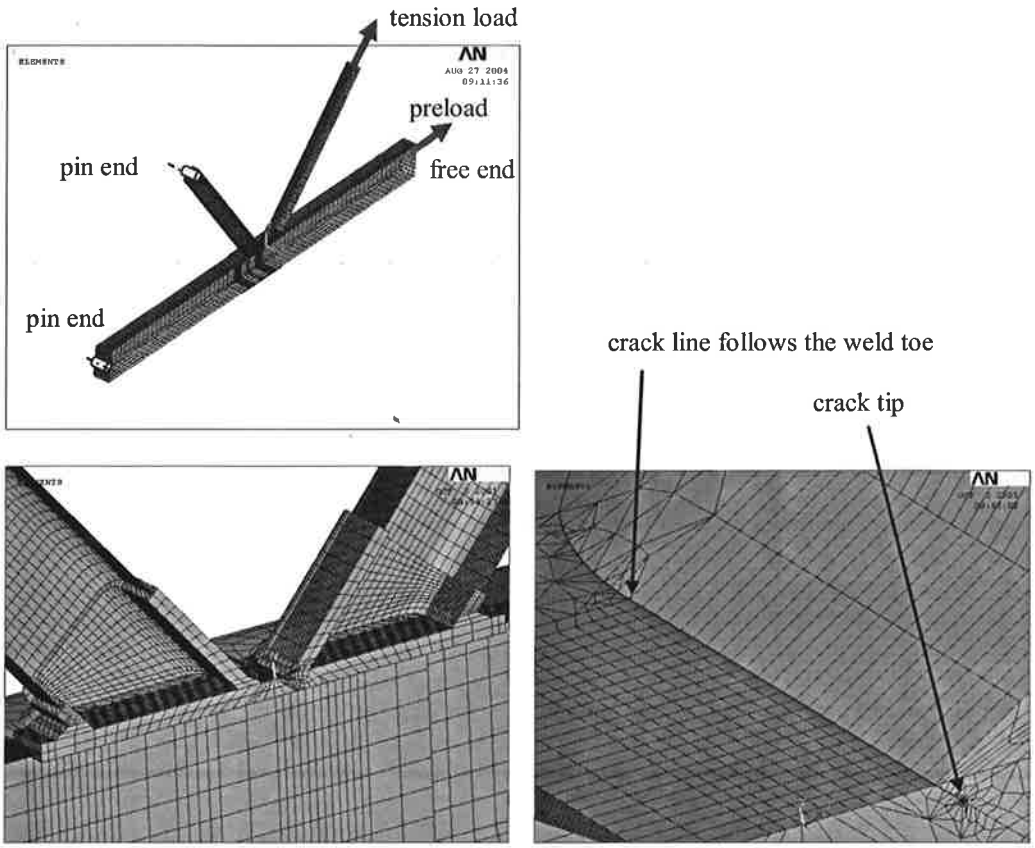


Figure 10. The FE model used for assessing the *J*-integral in a K-joint.

Regions near the crack were meshed with 20-node solid elements. Regions distant from the crack, but that were expected to experience nonlinear deformation, were modelled with 8-node solid elements. Pyramid

or tetrahedral versions of the 20-node solid elements were used in some difficult transition volumes near the crack. Connections between the parabolic and linear solids were done by reducing the mid-side nodes. Other regions of the structure, such as chords and the ends of the brace, were modelled with linear or parabolic shell elements depending on the connection to the solid elements.

The crack tip mesh consisted of two rows of crack tip elements. Only a limited number of crack tip elements around the crack tip were used because of the difficult joint geometry. The crack tip elements were also modelled as elastic crack tip elements having middle nodes ¼ distance from the crack tip and the nodes at the tip connected to each other. Each of the FE models had approximately 35 000 elements and 462 000 degrees of freedom.

The *J*-integral along the weld toe between the corner of the brace member and the centreline of the joint was assessed. For smaller crack depths and at lower loads, the *J*-integral reached a maximum value at a distance approximately 3.2 mm from the brace corner. This corresponds reasonably well with the location of the ductile crack from the fracture surface (see Fig 4.).

5 RESULTS AND DISCUSSION

The *J*-Δa crack resistance test results are summarized in Figure 11. These were generated using ASTM test procedures at -40 °C. The symbols LTY2, LTY8 and LTY12 refer to test specimen numbers. Specimen LTY2 was without cladding while specimens LTY8 and LTY12 included a cladding operation and the ductile crack growth was initiated in the HAZ approximately 1 mm from the fusion line (refer to Figure 7). In all cases the crack propagation direction represented a through-wall ductile crack extension.

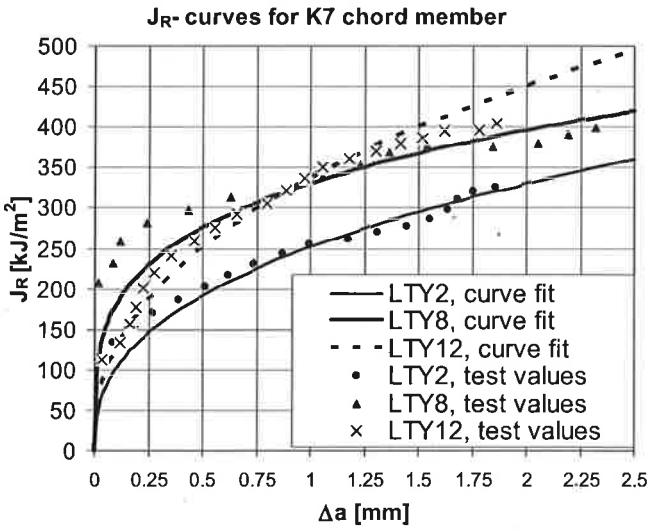


Figure 11. The measured crack resistance curves for specimens taken from RHS.

The *J*-integral calculation results for the K-joint of interest are shown in Figure 12 for crack depths of 0.1, 0.2, 0.5, 1.0 and 2.0 mm. The *J*-integral is calculated for a path situated 3.2 mm from the corner area of the brace where the highest elastic stresses were computed. This location corresponds well to the location where ductile fracture was observed to occur based on the fracture surface analyses.

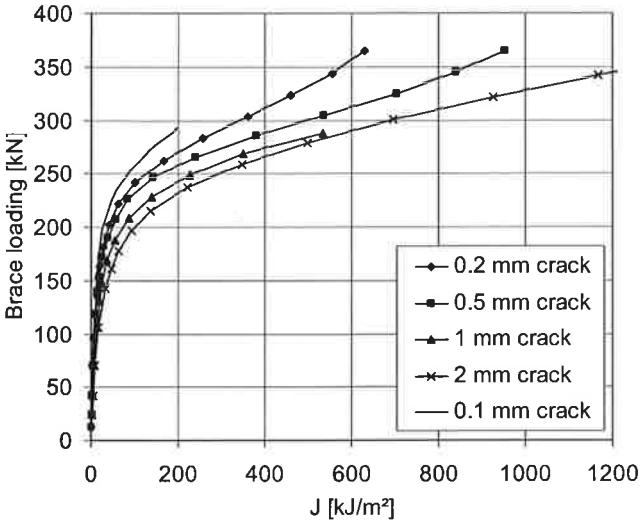


Figure 12. Computed *J*-integral values for the K-joint at a path 3.2 mm from the brace corner.

From the curves in Figure 12, the constant force curves (F = 200, 250, 300 and 350 kN) were derived. These are plotted in Figure 13. Experimental J-Δa crack resistance curves of the chord material from Figure 8 are shown in this same figure. The upper curve is for a crack initiating in the HAZ of the welded material while the lower curve is for a crack initiating and propagating in the base material. The upper crack resistance curve was assumed to be valid for crack depths up to 1 mm. For cracks deeper than this the base metal curve was considered more representative.

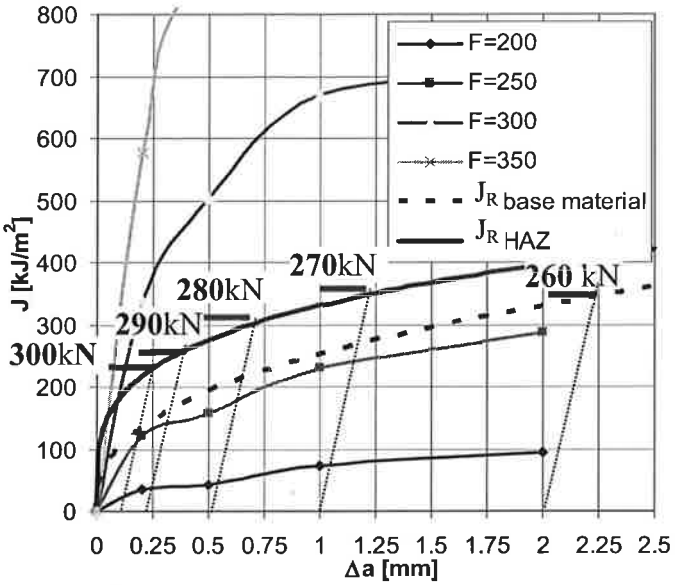


Figure 13. Assessment of the critical brace load to initiate ductile crack extension.

From Figure 13 it can be immediately seen that the 0.1 mm deep continuous crack-like defect along the weld toe would begin to grow by ductile fracture when the force in the brace member attains a force level between 250-300 kN. Further numerical analysis revealed that this value was approximately 300 kN. Once ductile fracture is initiated, this assessment shows that the crack

would continue to propagate even at slightly lower force levels. For example, a crack which has extended to a depth of 1 mm would continue to advance even if the load on the brace member is reduced to 270 kN.

When the crack has passed through the HAZ area, it was estimated that the force required for continued crack extension drops still further, i.e., about 260 kN for a 2 mm deep crack. The ductile crack extension for this K-joint is therefore combines both load- and displacement controlled mechanisms. Consequently, crack propagation does not become completely unstable, but the crack is predicted to advance at a relatively constant load. The estimated critical force is somewhat less than the value obtained from experimental test of this joint geometry, $F_u = 365$ kN. The numerical procedure, however, has included some assumptions that may need some refinement. For example, initial crack shapes other then the assumed continuous toe crack should be studied. The method does, however, provide a promising starting point for further studies. The method was a significant improvement when compared to estimating the ultimate deformation capacity using a more conventional von Mises type of failure criterion that tended to give overly optimistic assessment of joint ductility.

Once the FE model is constructed, it is possible to assess the temperature at which a joint will not exhibit sufficient ductility. This can be accomplished by substituting alternate J-Δa crack resistance curves measured at different temperatures. These small-specimen material tests are far less expensive that full joint tests. Also the influence of alternate welding processes involving different heat inputs can be assessed in this fashion.

6 CONCLUSIONS

The deformation capacity and ultimate strength of K-joints fabricated from cold-formed rectangular hollow sections at sub-zero temperatures has been studied both experimentally and numerically. Non-linear elastic-plastic finite element analysis was used to model the load-deformation behaviour of a K-joint. The true σ–ε material model used in the FE analysis was based on actual material tensile tests performed at -40 °C. The FE model geometry corresponded to the measured geometry of a joint tested in the laboratory, also at -40 °C.

The nonlinear FE analysis with solid elements could be used to accurately estimate the elastic-plastic load-displacement behaviour of K-joints, but did not provide an estimate of the deformation capacity or ultimate strength of the joint. The conventional von Mises type of failure criterion gave overly optimistic assessments of joint ductility and ultimate load-carrying capacity.

Elastic-plastic fracture mechanics analyses including the J-integral for one K-joint were combined with J-Δa crack resistance material data obtained using standardized test procedure. The ultimate load predicted using this procedure was slightly lower than the experimentally measured capacity of the joint. The method did, however, give an accurate representation of the location from which ductile fracture was expected to initiate. The method is promising and some the assumptions used in the analysis should be refined. Specifically, smaller crack sizes and initially semi-elliptical crack shapes should be investigated.

7 REFERENCES

[1.] ASTM E 1921-03. Standard test method for determination of reference temperature, T_0, for ferritic steels in the transition range. ASTM International, USA, 2002.

[2.] Björk, T., Ductility and ultimate strength of cold–formed rectangular hollow section joints at sub-zero temperatures, D. Sc. Thesis Lappeenranta University of Technology, *Acta Universitatis Lappeenrantaensis 233*, 2005

[3.] Björk, T., Kemppainen, R., Ilvonen, R. & Marquis, G., The capacity of cold-formed rectangular hollow section K-gap joints. *Proc. 10th symposium on tubular structures*, Edited by Jaurrieta, Alonso and Chica. ISBN 90 5809 552 5, Swets and Zeitlinger, Lisse, 2003.

[4.] Björk, T., Kemppainen, R., Ilvonen, R. & Marquis, G., On the definition of a material model for finite element analysis of cold-formed structural hollow section K-joints, *Proc. 10th symposium on tubular structures*, Edited by Jaurrieta, Alonso and Chica. ISBN 90 5809 552 5. Swets and Zeitlinger, Lisse, 2003.

[5.] Björk, T., Pellikka, V., Marquis, G. & Ilvonen, R., An experimental and numerical study on the fracture strength of welded structural hollow section X-joints. *Proc. 5th international ASTM/ESIS symposium of fatigue and fracture mechanics.* Reno, 2005.

[6.] Kumar, V. et al., *Advances in elastic-plastic fracture analysis. EPRI NP-3607.* Electric Power Research Institute, Palo Alto 1984.

[7.] Niemi, E., Behaviour of rectangular hollow section K-joints at low temperatures. *Proc. 3rd International symposium on tubular structures*, edited by E. Niemi and P. Mäkeläinen, Elsevier Applied Science, London, 1990.

[8.] Niemi, E., J., Use of cold-formed rectangular hollow sections in welded structures subjected to low ambient temperatures. CIDECT project 5AQ/2, *Cold-formed RHS in artic steel structures, final report 5AQ-5-96*, 1996.

[9.] Packer, J. A., Wardenier, J., Kurobane, Y., Dutta, D. & Yeomans, N., *Design Guide for rectangular hollow section (RHS) joints under predominately static loading.* CIDECT. ISBN 3-8249-0089-0 Verlag TÜV Rheinland GmbH, Köln 1992.

[10.] Packer, J. A. & Henderson, J. E., *Hollow structural section connections and trusses – design guide 2nd ed.*, ISBN 0-888111-086-3. Canadian Institute of Steel Construction, Universal Offset Limited, Allison, Canada, 1997.

[11.] Puthli, R., Hohlprofilkonstruktionen aus Stahl nach DIN ENV 1993 (EC3) und DIN 18800. ISBN 3-8041-2975-5, Verner Verlag, Düsseldorf, 1998.

[12.] Rice, J. R. & Tracey, D. M., On the ductile enlargement of voids in triaxial tress fields, *Journal of Mechanics and Physics of Solids*, Vol. 17, pp. 201-217, 1969.

[13.] Vainio, H., *Design Handbook for Rautaruukki structural hollow sections,* Rautaruukki Metform. ISBN 952-5010-47-3, Otava Book Printing Ldt, Keuruu 2000.

[14.] Wardenier, J., *Hollow sections joints.* ISBN 9062750842. Delft University Press, Delft 1982.

[15.] Wardenier, J., *Hollow sections in structural applications*, Delft University of Technology, Bouwen met Staal, ISBN 90-72830-39-3, 2002.

[16.] Zhao, X. L., Wilkinson, T. & Hancock, G. J., *Cold-Formed Tubular Members and Connections. Structural behaviour and Design.* ISBN 0-080-4410-17, Elsevier, Oxford 2005

[17.] EN1993-1-8. Design of steel structures, Part 1-8: Design of joints, European Committee for Standardization, 2003.

LUJITEMUOVIEN VÄSYTYSTESTAUS: MENETELMÄT, ONGELMIA JA ESIMERKKITAPAUS

M. KAJATSALO, T. KARTTUNEN, J. KOVANEN

Mikkelin ammattikorkeakoulu
YTI –tutkimuskeskus, materiaaliteknologian toimiala
PL 181, 50101 Mikkeli
e-mail: tero.karttunen@mikkeliamk.fi

TIIVISTELMÄ

Komposiittien mahdollisuudet konepajateollisuudessa (KOMPA) projektissa tutkittiin yhdessä Ahlstrom Glassfibren kanssa lasikuitulujitettujen muovien väsymiskäyttäytymistä. Tutkimuksen tavoitteina on ollut selvittää miten ja millä menetelmällä koekappaleiden väsytystestaus tulisi suorittaa, miten kappaleet tulisi valmistaa ja miten erilaiset kudoksen ominaisuudet vaikuttavat lujitemuovin väsymiskäyttäytymiseen. Tutkimus suoritettiin kahdessa vaiheessa. Ensimmäisessä vaiheessa suoritettiin lujitemuovien väsymiseen liittyen kattava kirjallisuustutkimus sekä haastateltiin useita komposiittialan ammattilaisia. Toisessa vaiheessa suoritettiin väsytystestit lasikuitulujitemuovista valmistetuille koekappaleille. Kokeet suoritettiin DLR:llä saksassa sekä Mikkelin ammattikorkeakoulun materiaalitekniikan laboratoriossa Mikkelissä. Kokeiden tarkoituksena oli verrata eri laitoksissa saatuja tuloksia keskenään sekä selvittää miten kudoksen ominaisuudet vaikuttavat lujitemuovin väsymiskäyttäytymiseen.

1. JOHDANTO

Lujitemuovituotteita käytetään nykyään mitä moninaisimmissa kohteissa kuten paperi- ja painokoneiden teloissa, paperi- ja muun prosessiteollisuuden putkistoissa ja säiliöissä, mastoissa ja lipputangoissa, urheiluvälineissä sekä kuljetusvälineissä. Käytössä komposiitit altistuvat tyypillisesti väsyttäville kuormituksille. Tällaiset kuormitukset aiheuttavat lujitemuovin materiaaliominaisuuksien muuttumista ja johtavat lopulta materiaalin vaurioitumiseen ja sen kantokyvyn pettämiseen. Tietoa lujitemuovimateriaalien väsymiskäyttäytymisestä saadaan väsytyskokeista. Lujitemuovituotteiden väsymistestaus on haastava tehtävä, jonka teorioiden ja periaatteiden perusteellinen tuntemus on tärkeää tulosten oikeellisuuden kannalta.

2. KIRJALLISUUSTUTKIMUS

Kirjallisuustutkimuksessa käsiteltiin laminaatin, testisauvojen, testausmenetelmän sekä ympäristön osuutta väsymisessä. Lujitemuovin käyttäytymiseen väsyttävän kuormituksen alaisena vaikuttavat mm. matriisin ja kuitujen ominaisuudet, kuitujen suuntaus, kuitutuotteen ominaisuudet, käytetty valmistustekniikka, kuidun ja matriisin välinen sidos, kuitujen tilavuusosuus rakenteessa sekä valmistuksen ammattitaito. Lujitemuovikomposiiteissa käytettyjä kuitutuotteita ovat irralliset kuidut sekä erilaiset kuiduista valmistetut puolivalmisteet, kuten katkokuitumatot, huovat, kudokset, yhdistelmätuotteet ja multiaksiaalit. Testisauvoille suoritettuihin testituloksiin vaikuttavat puolestaan laminaattien valmistus, koekappaleiden valmistus, koekappaleen muoto, käytetty menetelmä, testauslaitteisto, testaajan ammattitaito (kiinnitys koneeseen jne.), ympäristö sekä väsytystaajuus.

Väsytystestien suorittamiseksi etsittiin tarvittavaa tietoa kirjallisuudesta, komposiittialan osaajilta ja testauslaitoksilta. Tuloksista koostettiin raportti "Coupon Fatigue Testing of FRP laminates" (MAT05-1603-007). Raportti tarjoaa tietopaketin lujitemuovien väsymisilmiöstä ja väsytystestauksesta. Väsytystestauksessa keskitytään koesauvojen valmistuksen ja testauksen esittelemiseen, sillä oikeilla osilla tehtävät väsytystestit suunnitellaan usein tapauskohtaisesti. Raportti palvelee kuitenkin myös oikeiden osien testin suunnittelemista, sillä esimerkiksi laminaattien valmistus, kuormitustapaukset ja tulosten esittäminen ei riipu koekappaleen muodosta.

Lujitemuovien väsytystestauksesta on julkaistu paljon artikkeleita ja tutkimuksia mutta siitä huolimatta lujitemuoveja käsittelevät standardit (ASTM D 3479, ISO 13003) eivät ole sellaisenaan käytössä monessakaan testipaikassa, vaan tutkimuslaitokset ovat kehittäneet omia menetelmiä väsymisen testaamiseen. Tähän saattaa olla syynä testilaboratorion oma pitkä väsytystestaushistoria ja näkemys testin suorittamisesta (tulosten vertailtavuus). Raportin (MAT05-1603-007) liitteeksi on tehty lyhennelmä tuulivoimateollisuuden sertifikaatteja myöntävän Germanischer Lloydin tuulivoimaosaston määräyksistä testin ja testitulosten suhteen.

3. VÄSYTYSTESTIT

3.1 VÄSYTYSTESTIT YTI –TUTKIMUSKESKUKSESSA

YTI:n väsytystestit tehtiin 30.1. – 6.4.2006. Testeissä tehtiin 4kpl staattisia vetokokeita ja 37kpl väsytyskokeita. Väsytyskokeista onnistui 23kpl.

3.1.1 TESTIJÄRJESTELY

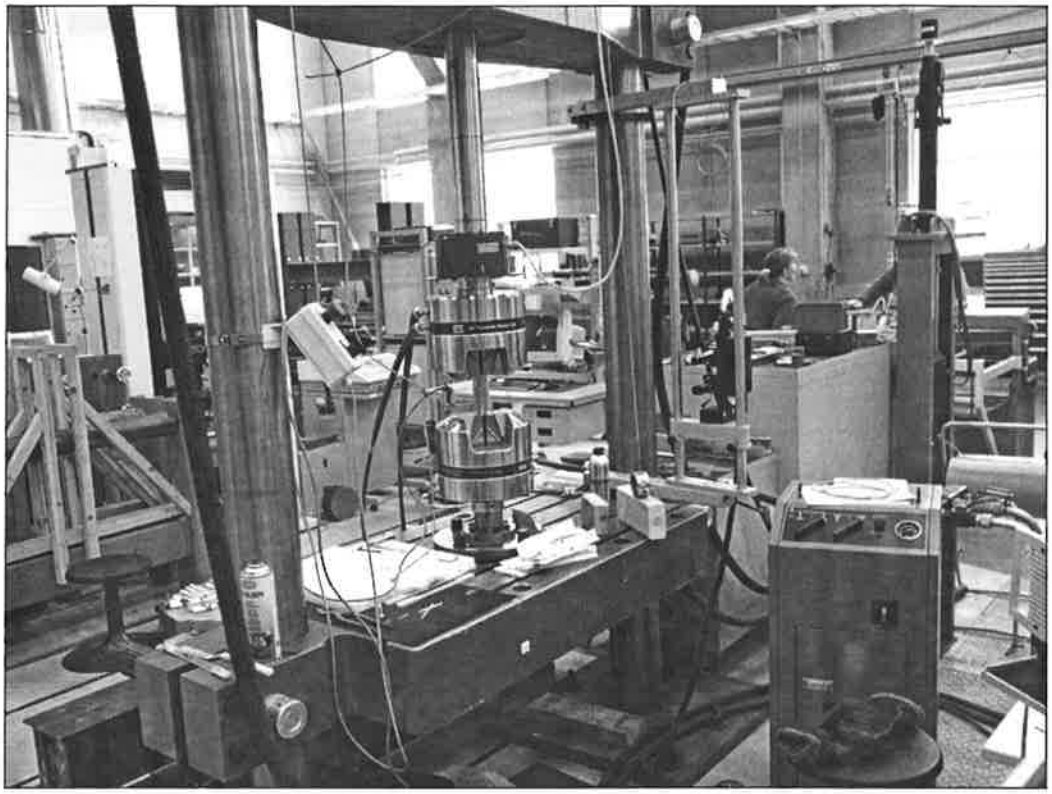
YTI -tutkimuskeskuksessa suoritetut UD –laminaattien väsytystestit tehtiin MAMK:n materiaalitekniikan laboratorion Matertest –aineenkoetuslaiteella (Kuva 1). Testilaitteessa oli hydrauliset väsytysleuat, joilla testikappaleisiin saatiin tarvittava vakiopuristusvoima. Koesauvojen puristuspaineena käytettiin 5...6MPa, mikä vastaa 38...46kN puristusvoimaa. Puristuspaine oli leukojen valmistajan suositusten mukainen dynaamiselle, maksimissaan 20...24kN, vetokuormitukselle. Kaikki väsytyskokeet olivat veto-veto kokeita, joiden minimi/maksimi –kuorman suhde oli 0.1. Testit ajettiin voimaohjattuna ja koe keskeytyi, kun siirtymärajaksi määritelty 5...6mm -arvo ylittyi. Tämä arvo ylittyi aina, kun laminaatti voitiin katsoa rikkoontuneeksi.

Kappaleen venymää seurattiin omavalmisteisella extensiometrillä, joka kalibroitiin VTT:n kalibroinnin piiriin kuuluvalla toisella extensiometrillä. Koekappaleen lämpötilaa seurattiin testin alussa ja huippulukemat kirjattiin ylös. Lämpötilat mitattiin koesauvan keskeltä ja grippien juuresta noin 15...20 minuutin päästä testin aloituksesta. Grippien juurien lämpötilat mitattiin koekappaleen sivusta, koska leuat estivät mittaamasta niitä pinnan tasosta.

M72 –sarjan väsytystaajuuksia ei määritelty tarkasti, mutta suuremmilla kuormilla taajuus oli alempi. M82 –sarjan väsytystaajuudet valittiin seuraavan taulukon 1 mukaisesti.

Taulukko 1 Väsytystaajuudet kuormituksen mukaan.

Maksimikuormitus	24	22	20	18	16	14	13	12
Taajuus	2	2.2	2.4	2.7	3	3.4	3.7	4



Kuva 1 Väsytyskoe menossa Matertest –aineenkoetuslaitteella.

3.1.2 TESTILAITTEISTON ARVIOINTI VENYMÄLIUSKAMITTAUKSILLA

Koesauvoihin syntyviä taipumia tarkasteltiin standardin ISO 527-4 liitteen mukaisesti liuskoitetulla koesauvalla. Liite antaa ohjeet koejärjestelyn suuntausvirheiden arviointiin venymäliuskamittauksilla. Sauvan keskelle oli liimattu kolme liuskaa, toiselle puolelle vastakkaisiin reunoihin kaksi liuskaa ja vastakkaiselle puolelle keskelle yksi liuska (Kuva 2). Näin voitiin tutkia sauvan kiinnityksestä ja suoruudesta aiheutuvia taipumia. Testit tehtiin sekä suoraan että vinosti suunnatulla koesauvalla.

Taivutuksen osuus suora- ja vinovetomittauksissa on laskettu taulukkoon 2. Tuloksista nähdään, että suoravedossa taivutuksen osuus on suurempi paksuussuunnassa kuin leveyssuunnassa. Paksuussuunnan virheet syntyvät suurilta osin leukojen kulmavirheestä ja koesauvan geometriavirheistä. Vinovedoissa taivutuksen osuus venymistä kasvaa myös leveyssuunnassa, mutta ei lineaarisesti kulmavirheen suhteen. Suoravedossa taivutuksen osuus paksuussuunnassa 2kN:n kohdalla on taulukon mukaan suuri, koska absoluuttiset virheet ovat syntyneet pääasiassa jo yläleukaa suljettaessa. Tulokset viittaavat siihen, että koesauvan muotovirheet (mm. sauvan suoruus ja grippien tasotoleranssi) ja aineenkoetuslaitteen geometriset virheet ovat olemassa, mutta itse vedosta ei aiheudu suurta lisätaivutusta koesauvan keskiosaan. Sinällään tämä tulos ei poissulje taivutuksen syntymistä grippien läheisyydessä. Tätä pitäisi kuitenkin tarkastella laajemmalla mittauksella.

Virhearviointien testit suoritettiin alhaisemmilla kuormilla kuin väsytystestit. Tähän oli syynä se, että sauvan matriisi haluttiin pitää ehjänä. Taivutuksen osuuksien summa on alempi väsytystestien maksimivoimilla (14…24kN), koska paksuussuunnan osuus laskee kuormituksen kasvaessa. Väsytystestien maksimivoimilla $B_b + B_h$ asettuu noin 3% pintaan ja jopa sen alle.

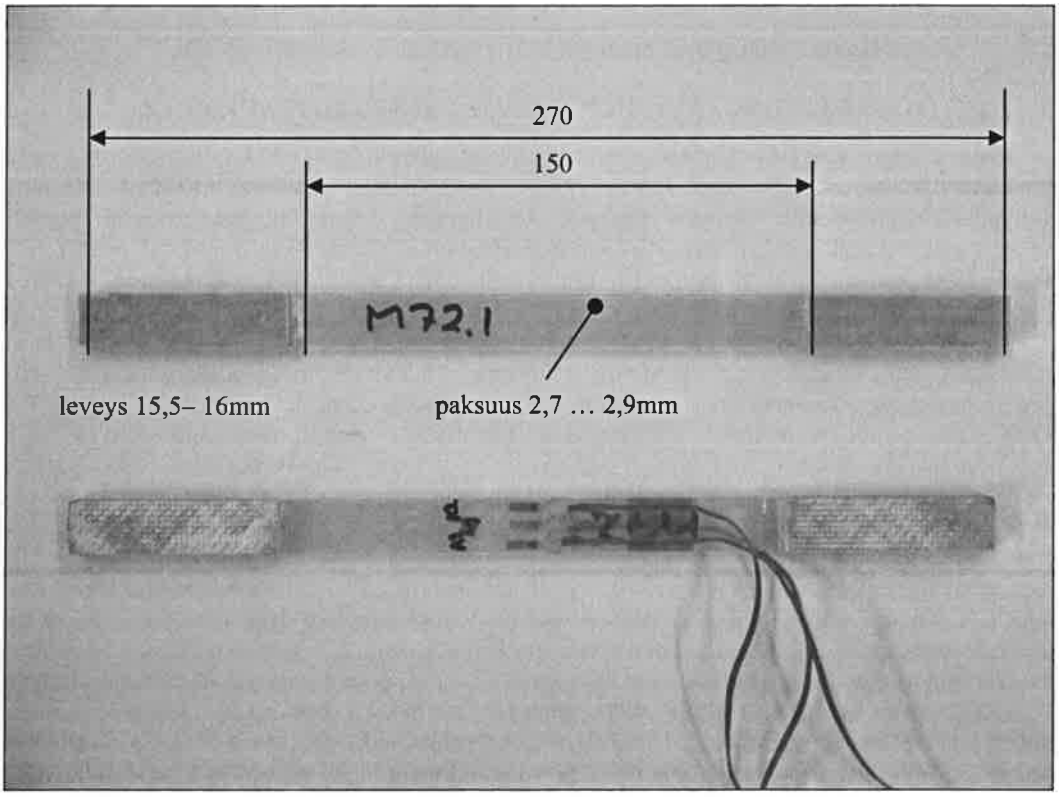
Taulukko 2 Taivutuksen osuus suora- ja vinovedoissa sekä dynaamisessa kuormituksessa.

Suoraveto	ε_{av} [με]	B_b [%]	B_h [%]	$B_b + B_h$ [%]
2kN	1139	0,6	8,4	8,9
10kN	5817	1,1	4,1	5,2
Vinoveto 1mm				
2kN	1147	18,5	7,9	26,4
10kN	5824	4,0	4,1	8,1
Vinoveto 2,5mm				
2kN	1167	5,1	7,9	13,0
10kN	5834	1,9	4,2	6,1
Dyn. kuormitus *)				
2…9kN	4107	2,4	3,2	5,6

*) Dynaamisen kuormituksen virhearvot on laskettu amplitudeista. Näitä virhearvoja ei voi verrata staattisiin virhearvoihin.

3.1.3 KOESAUVAT

Koesauvojen keskimääräiset päämitat on esitetty kuvassa 2. Testisarjoja oli kaksi, M72 ja M82. Niiden laminaatit olivat samanlaiset (UD), mutta ero oli tikkauslankojen käytössä. Molempien sarjojen sauvat oli valmistettu kolmesta valmistuserästä. Näitä merkittiin M72.1, M72.2 ja M72.3 sekä vastaavasti M82.1, M82.2 ja M82.3. Kaikki koesauvat oli valmistettu Ahlstrom Oyj.:n toimesta.



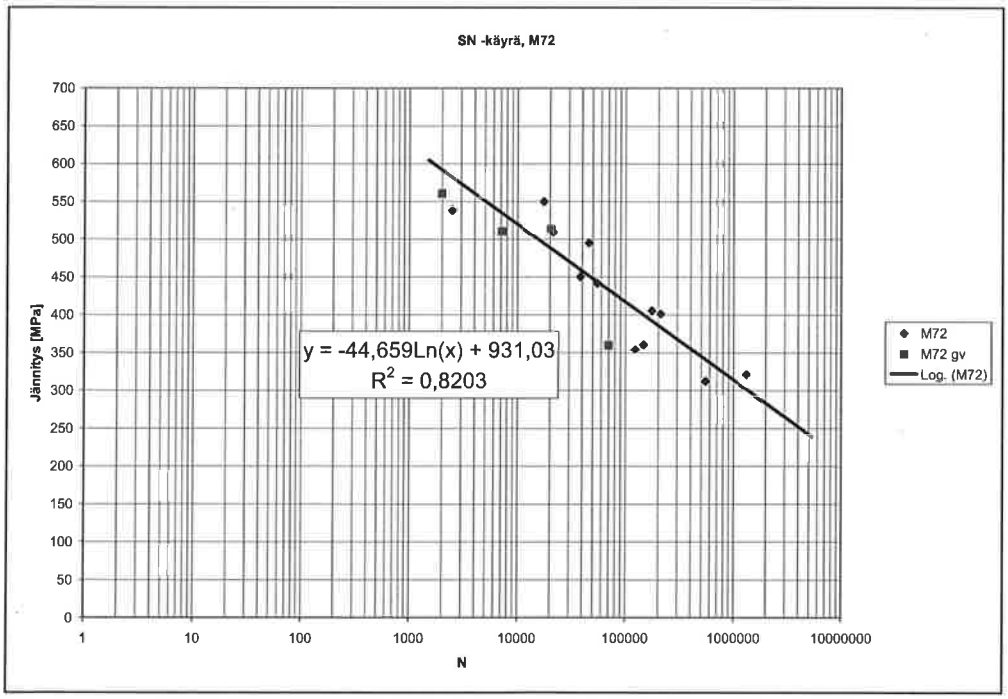
Kuva 2 Koesauvan päämitat (venymäliuskat olivat vain mittakoesauvassa).

3.1.4 TULOKSET

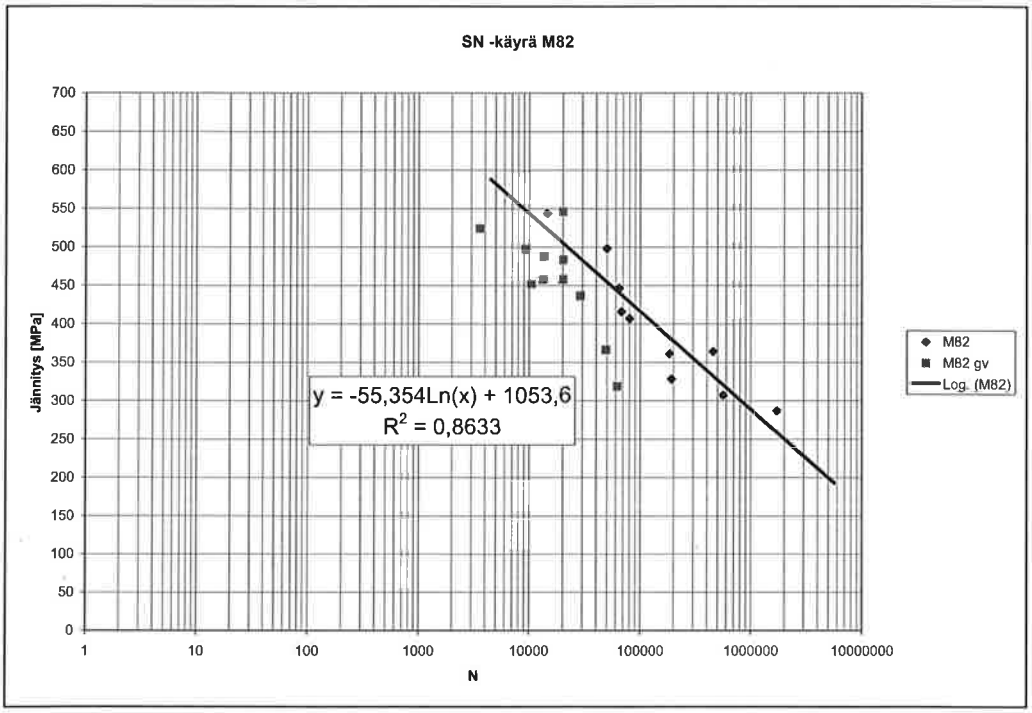
Ennen varsinaisia väsytystestejä tehtiin testisarjoille staattiset testit 2mm/min vetonopeudella. M72 –sarjasta vedettiin yksi koesauva, jonka tulos oli 903MPa. M82 –sarjasta vedettiin kolme koesauvaa: 1023MPa, 1048MPa ja 994MPa. Näiden perusteella arvioitiin väsytystestien suurimmaksi kuormitusamplitudiksi 50...60% staattisten vetotestien arvoista. Staattisista vetokokeista lasketut kimmomoduulien arvot olivat M72 –sarjassa 39.9GPa ja M82 –sarjassa 38.9GPa, 41.4GPa ja 39.0GPa.

Väsytystesteissä sauvat vaurioituivat joko laminaatista tai grippipalojen juuresta. Gripin juuressa olevat liimaukset antoivat eriasteisesti periksi jokaisessa testissä, mutta erityisesti suurilla voimilla (≥20kN) koesauvat myös vaurioituivat ko. kohdasta. Grippien juuresta vaurioituneet koesauvat hylättiin eikä niitä sisällytetty SN -käyriin. M82 –sarjassa oli grippivikoja 10/21, mutta M72 –sarjassa vain 4/16. M82 –sarjassa grippiviat jakaantuivat seuraavasti: M82.1: 7/8, M82.2: 1/7 ja M82.3: 3/6. Tällainen jakauma grippivikojen suhteen voi viitata myös keskinäisiin eroihin koesauvoissa (tai grippien liimauksissa), koska koesauvat testattiin sekalaisessa järjestyksessä ja samoja kuormitusarvoja käytettiin kaikissa erissä tasaisesti.

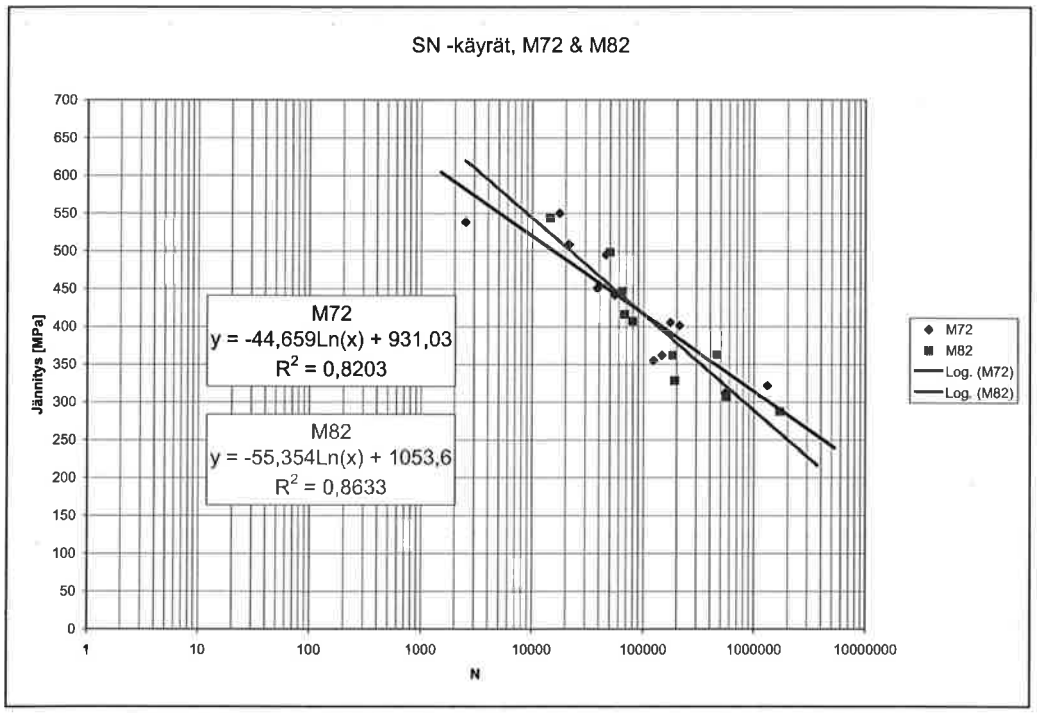
Väsytystestien tuloksena saatiin kummallekin testisarjalle SN –käyrät. Testisarjojen SN –käyrät on esitetty alla olevissa kuvissa (Kuva 3 ... Kuva 5). Kuviin on myös lisätty pistejoukkoon sovitettu logaritminen tai eksponenttiaalinen käyrä. Gv –pisteet tarkoittavat testejä, jotka ovat päättyneet grippivikaan. Kuvaajista nähdään, että grippivikaisten sauvojen elinikä oli alempi. Näin ollen yksittäisen sauvan väsytystestin hyväksymiskriteerinä oli perusteltua käyttää vaurion syntymistä muualla kuin grippien juurissa. Gv –pisteet eivät ole mukana sovitetuissa käyrissä. SN –käyrien hajonnat olivat sen verran suuret, että M72- ja M82 –testisarjojen käyrien välistä jyrkkyyseroa ei voida pitää vertailukelpoisena (Kuva 5). Näin ollen tikkaustavan vaikutuksesta elinikään ei voida päätellä tämän väsytystestin perusteella.



Kuva 3 M72 –sarjan SN –käyrä logaritmisella sovituksella. Gv –pisteet ovat grippiviallisia (ei mukana sovituskäyrässä).



Kuva 4 M82 –sarjan SN –käyrä logaritmisella sovituksella. Gv –pisteet ovat grippiviallisia (ei mukana sovituskäyrässä).



Kuva 5 M72- ja M82 –sarjan SN –käyrät logaritmisilla sovituksilla.

3.1.5 JOHTOPÄÄTÖKSET YTI:N TESTEISTÄ

Väsytystestien tuloksena saatiin SN –käyrät M72- ja M82 –sarjoille. M82 –sarjan sovitettu SN –käyrä on jyrkempi, mutta pistejoukon hajontaan suhteutettuna jyrkkyyseroa ei voida pitää luotettavana. Näin ollen tikkaustavan vaikutuksesta elinikään ei voida päätellä tämän väsytystestin perusteella. M82 –sarjan testien hajonta oli pienempi kuin M72 –sarjan hajonta. Tähän voi olla syynä M82 –sarjan testeissä käytetty koesauvan suuntaus luotilangan avulla ja väsytystaajuuden tarkka määrittely kuormituksen suhteen.

Grippivikaisten sauvojen kestoikä oli selvästi alempi kuin itse laminaatista vaurioituneiden sauvojen kestoikä. Näin ollen grippivikaisten koesauvojen hylkääminen testituloksista oli perusteltua.

3.2 VÄSYTYSTESTIT DLR:LLÄ SAKSASSA

DLR suoritti viiden laminaatin väsytystestauksen. Testeissä käytettiin standardiin ISO 527 pohjautuvia suoria testikappaleita kuten YTI:n testeissäkin. Kaikki suoritetut väsytyskokeet olivat veto-veto kokeita ja R arvona oli 0,1. Kussakin sarjassa oli 2 + 12 koekappaletta. Kahdelle kappaleelle tehtiin staattinen vetokoe murtolujuuden määrittämiseksi. Varsinaiset väsytyskokeet suoritettiin 12 kappaleella siten, että kaikkien koekappaleiden väsymiseen tarvittava syklimäärä oli alueella 10^3-10^6.

Saksassa suoritettiin kokeet seuraaville laminaateille:

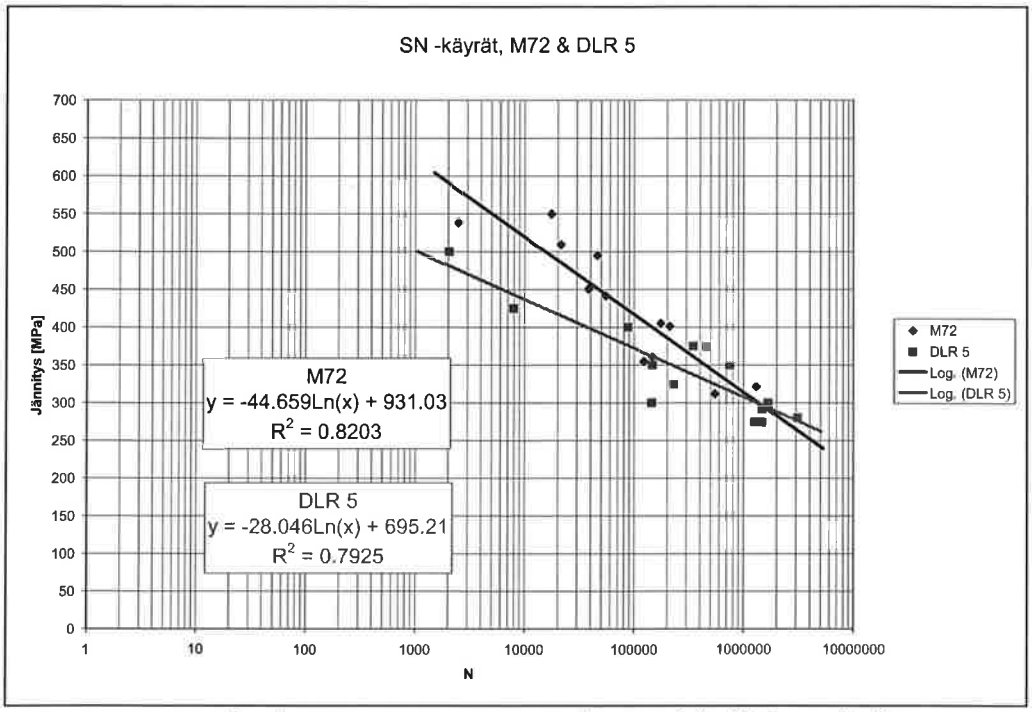
1. UD laminaatti scrim
2. UD + 90° sidontana, scrim
3. UD + 45° sidontana, scrim
4. tuotannollinen tikattu laminaatti
5. tuotannollinen tikattu laminaatti

Tässä yhteydessä käsitellään vain laminaatin nro. 5 väsytystestejä, koska samaa laminaattia testattiin YTI:n toimesta. DLR 5 -sarja vastasi YTI:n väsytystestisarjaa M72.

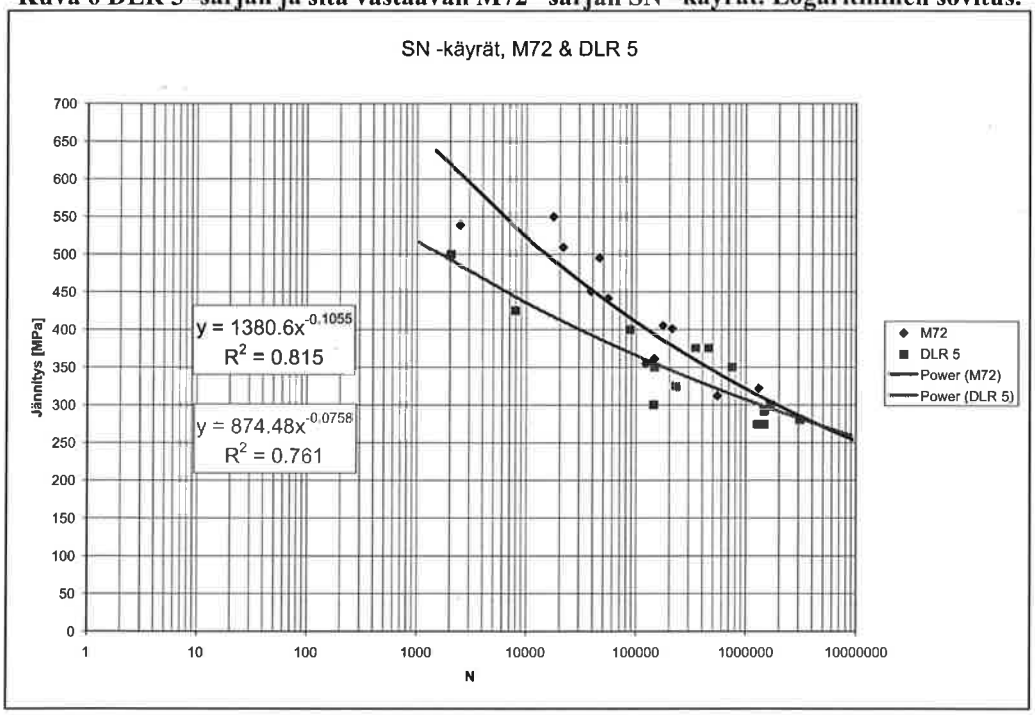
3.2.1 TULOKSET

DLR 5- ja M72 –sarjojen SN –käyrät on esitetty seuraavissa kuvissa (Kuva 6 ja Kuva 7). Kuva 7 on DLR:n esitystapa SN .käyrästä (eksponentiaalinen sovitus). Kuvista nähdään, että DLR 5 –sarjan sovitettu SN –käyrä on loivempi. DLR:n 250...400MPa:n pisteet ovat suunnilleen samoilla linjoilla M72 –sarjan pisteiden kanssa, mutta kaksi ylempää pistettä loiventaa käyrää huomattavasti. Tämä voi antaa paremman kuvan väsytyskestävyydestä alhaisilla jännityksillä ja huonomman suuremmilla jännityksillä. DLR:n tuloksissa on esitetty kaikki tulokset, myös grippivikaiset, samassa SN –käyrässä. YTI:n tekemissä väsytystesteissä tämän todettiin vaikuttavan SN –käyrään, kuten kappaleessa 3.1 todettiin.

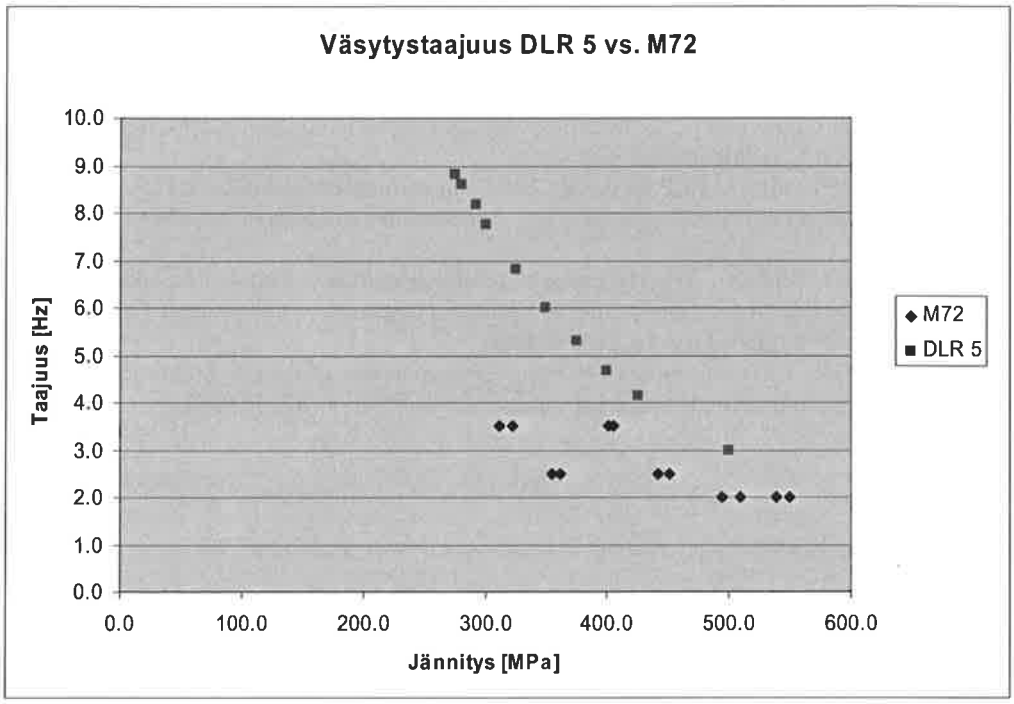
Kuva 8 esittää käytettyjä väsytystaajuuksia DLR 5 ja M72 –sarjoissa. DLR:n testeissä on käytetty korkeampia väsytystaajuuksia kuin YTI:n M72 –sarjan testeissä, mutta korkeilla jännityksillä ero oli vähäinen. M72 –sarjan väsytystaajuudet eivät kulje täysin kuormituksen funktiona, koska sarjaan sisältyi mm. väsytystaajuuksien kokeiluja.



Kuva 6 DLR 5 -sarjan ja sitä vastaavan M72 –sarjan SN –käyrät. Logaritminen sovitus.



Kuva 7 DLR 5 -sarjan ja sitä vastaavan M72 –sarjan SN –käyrät. Eksponentiaalinen sovitus.



Kuva 8 Väsytystaajuudet DLR 5- ja M72 –testeissä.

3.3 JOHTOPÄÄTÖKSET

DLR:n ja YTI:n väsytystestien erot olivat testituloksien esittämisessä sekä väsytystaajuudessa. DLR:n väsytystesteissä osa tuloksissa esitetyistä SN –käyrän pisteistä oli grippivikaisia. Väsytystaajuudet olivat DLR:n testeissä myös korkeammat. Lopullista tulosvertailua ei voida vielä tehdä, koska lopullista testiraporttia ei ole vielä saatu DLR:ltä (24.4.2006), mutta testauspaikkojen käytännöillä on testituloksiin vaikuttavia eroja. Testauskäytäntöjen ei kuitenkaan pitäisi antaa vaikuttaa SN –käyrään, koska testausolosuhteiden vaikutukset pitäisi minimoida. Testauspaikkojen tuloksia verrattaessa on siis oltava perillä testien suoritustavoista.

LÄHTEET

1. ASTM D 3479/D 3479M – 96 (Reapproved 2002). Standard test method for tension-tension fatigue of polymer matrix composite materials. ASTM International. 10. 6. 1996, Reapproved October 2002, West Conshohocken, USA.
2. Coupon Fatigue Testing of FRP laminates. Mikkelin ammattikorkeakoulu. YTI – tutkimuskeskus, Materiaaliteknologian toimiala. Materiaaliteknologian raportti MAT05-1603-007. 2005.
3. ISO 527-4. 1997. Plastics -- Determination of tensile properties -- Part 4: Test conditions for isotropic and orthotropic fibre-reinforced plastic composites. International Organization for Standardization. 15. 4. 1997, Geneve, Switzerland.
4. ISO 13003. 2003. Fibre-reinforced plastics – Determination of fatigue properties under cyclic loading conditions. International Organization for Standardization. 15. 12. 2003.

FRACTURE MECHANICS ASSESSMENT OF MIXED MODE FATIGUE CRACKS IN CFRHS BEAMS

S. HEINILÄ, G. MARQUIS and T. BJÖRK,
Laboratory of Fatigue and Strength
Lappeenranta University of Technology
P.O. Box 20
FIN-53851 Lappeenranta, FINLAND

ABSTRACT: Mixed mode crack propagation in cold-formed corners of high-strength steel plate structures has been investigated using linear elastic fracture mechanics and laboratory testing. Large- and small-scale test specimens having complex residual stress states have been fatigue tested and examined using linear and nonlinear finite element analysis. Straight, zig-zag and "S" shaped cracks were observed depending on the material strength, range of cyclic loading, residual stress field and multi-axiality of the applied loading. The multi-axial stress state was produced by both the corner geometry and residual stresses. Numerical simulations of residual stresses and linear elastic fracture mechanics were used to help understand the alternate crack paths.

1 INTRODUCTION

Cold-forming is a widely used fabrication process for plate welded structures. Cold-formed rectangular hollow section (CFRHS) tubes are frequently fabricated in a two-phase process in which a flat plate is initially cold-formed into a circular mother tube and is then further cold-rolled into its final rectangular shape. The cold-forming process introduces high levels of residual stress which may be either beneficial of detrimental with respect to fatigue strength of the structure. During plate bending or fabrication of a CFRHS, the concave inside corner surfaces experience significant compressive plastic strains. The resulting tensile residual stress along the inside of the bend region enhances the crack propagation during cyclic loading. The bend process also coarsens the initially smooth surface and enhances both micro-cracking and crack development on the surface [1].

If the residual stresses in a fabricated component are sufficiently high, a crack can propagate even if the applied local stresses are cyclically compressive. Greasly et al. [2] showed that a mode I fatigue crack within a tensile welding residual stress field can grow even when the cyclic external stresses are compressive. It was reported that the cracks arrested after a period of propagation. Hermann [3] tested compact tension aluminium alloy specimens that were pre-compressed in order to create a tensile residual stress field ahead of the notch tip. It was shown that the crack length at which a crack arrested increased with increasing levels of the pre-compression, i.e. an increasing large tensile residual stress field.

In this current study, multi-axial compressive loading was applied to the cold-formed corners of CFRHS beams, CFRHS sections and simple bent plates. Cracks always initiated along the inside of the bend radius where high tensile residual stresses exist. Cracks did not arrest as they propagated into the compressive residual stress region, but the crack mode was altered in some

cases. Several distinct crack paths were observed depending on the type and range of cyclic loading, material and residual stress conditions of the specimens.

In the case of applied mixed-mode loading, it has been previously reported that the crack growth rate alone is not sufficient for assessing the fatigue strength of a technical structure, but that the strength also depends on the growth direction of the fatigue crack [4,5]. In this study of the residual stress field induced a complex multiaxial stress field that had significant influence on the crack growth direction and thus on the lifetime of the structure.

2 BACKGROUND

The current study was partially motivated by an observed in-service failure of a CFRHS high-strength steel beam structure[1]. The observed crack was long (in the longitudinal direction of the tube) and had a small aspect ratio. It propagated undetected from the inside corner through the tube wall. The resulting fatigue crack had an "S" shaped path as seen in Fig. 1.

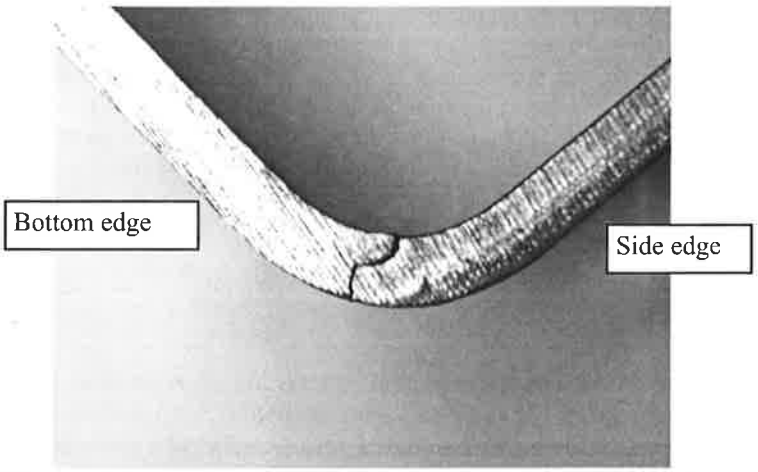


Figure 1. Observed crack path from a CFRHS that failed in service.

3 LABORATORY TESTS

Analytical and experimental studies were planned to both reproduce the observed cracking behaviour and to better understand the fatigue crack propagation behaviour. Both large and small scale specimens were tested.

3.1 Large Scale Specimens

The large scale laboratory specimen was a CFRHS manufactured by cold rolling and was identical to that used in the failed in-service structure. The test configuration included two beam-type specimens of which four corners, two of each specimen, were monitored for cracks using back side strain gauges, i.e. strain gages were glued to the outside corner of the tube. The CFRHS profile is generally shown in Fig. 2 and a schematic of the test arrangement is shown in Fig. 3. In total, eight specimens were tested. The CFRHS was fabricated from high strength steel with nominal yield stress f_y = 650 MPa. The middle support shown in Fig. 3 consisted of a series of rollers that travelled back and forth in-phase with the applied bending loading. The lateral roller translation was about 10% of the length of the beam-type specimen. Fatigue cracking was observed within

this region of travel and also in the regions just beyond travel of the rollers. Multiple cracks initiated near the centre of the inner corner of the tube as can be seen in Fig. 4.

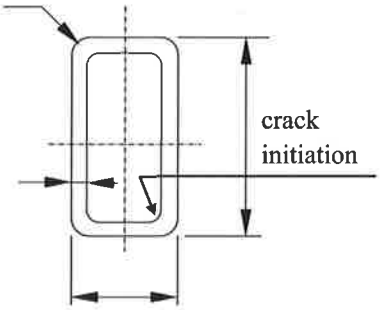


Figure 2. CFRHS profile.

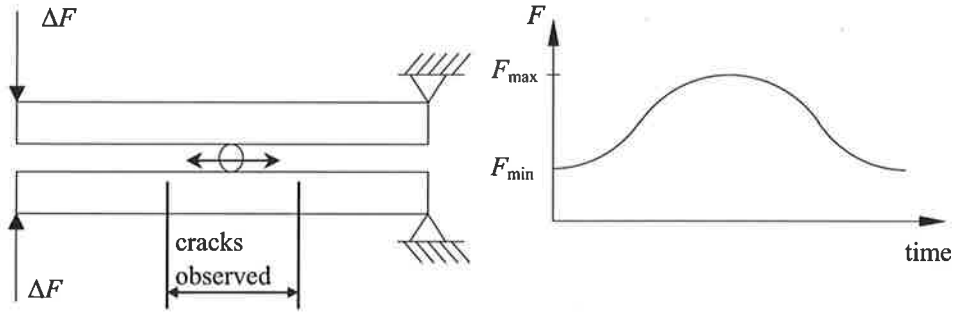


Figure 3. Testing configuration and schematic diagram showing the applied loading ΔF.

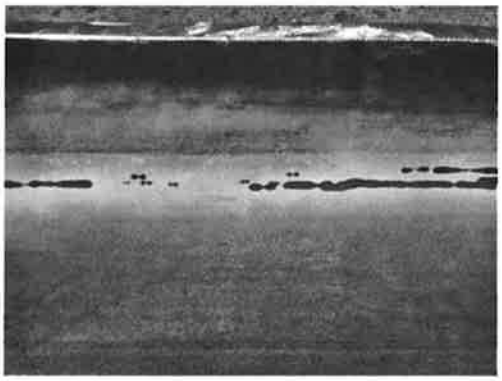


Figure 4. Cracks initiated along the inner corner of the tube. Dark regions are cracks highlighted using a dye-penetrant solution.

The residual stresses at the centre of the corner were measured using the X-ray diffraction method. Residual stresses on the inner and outer surfaces and at small depths from the inner surface were measured. Measured tensile residual stresses on the inner surface were in the range 30…50% of

the nominal yield strength of the material. Results for one tube corner are shown in Table 1. Depth measurements in Table 1 are presented as a portion of the wall thickness, t, of the tube. The outer surfaces of the tubes were found to have compressive residual stresses.

Table 1. The tangential residual stresses σ_{res} at the centre of the CFRHS corner

depth	σ_{res}	
	MPa	**± MPa**
0 (inner surface)	213	25
0.003 * t	313	32
0.006 * t	292	17
0.012 * t	252	33
0.022 * t	188	33
0.035 * t	174	33
0.051 * t	173	44
0.066 * t	124	55
1.000 * t (outer surface)	-73	44

Residual stresses along the inner corner at positions other than the centre were also measured. These stresses were also tensile, but the magnitudes were lower and are not here reported. Crack initiation was observed to occur in the centre of the corner where the tensile residual stress was highest. Numerous small cracks, longitudinal with respect to the beam-type specimen, propagated along the inner surface and eventually joined to form a single long small aspect ratio crack. This crack then propagated transverse through the wall of the tube. Crack growth was initially orthogonal to the inner surface of the tube. However, near the neutral plane of the wall the, crack appeared to turn 90° with respect to the original direction forming a zigzag crack as seen in Fig. 5. Closer inspection revealed, however, that the crack did not turn but rather joined with other cracks that had initiated tangential to the tube corner radius near the centre of the wall. This is seen in Fig. 6.

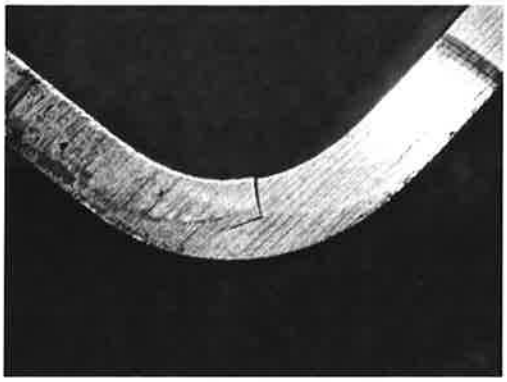


Figure 5. Crack branching observed in a large scale laboratory specimen.

Testing also included two beam-type specimens that were thermal treated so as to relieve, or at least reduce, the residual stresses in the beam corners. These specimens showed completely different cracking behaviour. No cracks were observed to initiate on the inner surfaces of these tubes. Cracks in these specimens initiated on the outer surface of the tube due to contact fatigue with the moving middle support rollers. Fatigue strength was significantly improved with respect

to the non-heat treated beams. This observation clearly demonstrated that the residual stresses have a major effect on fatigue crack path and fatigue strength of the structure.

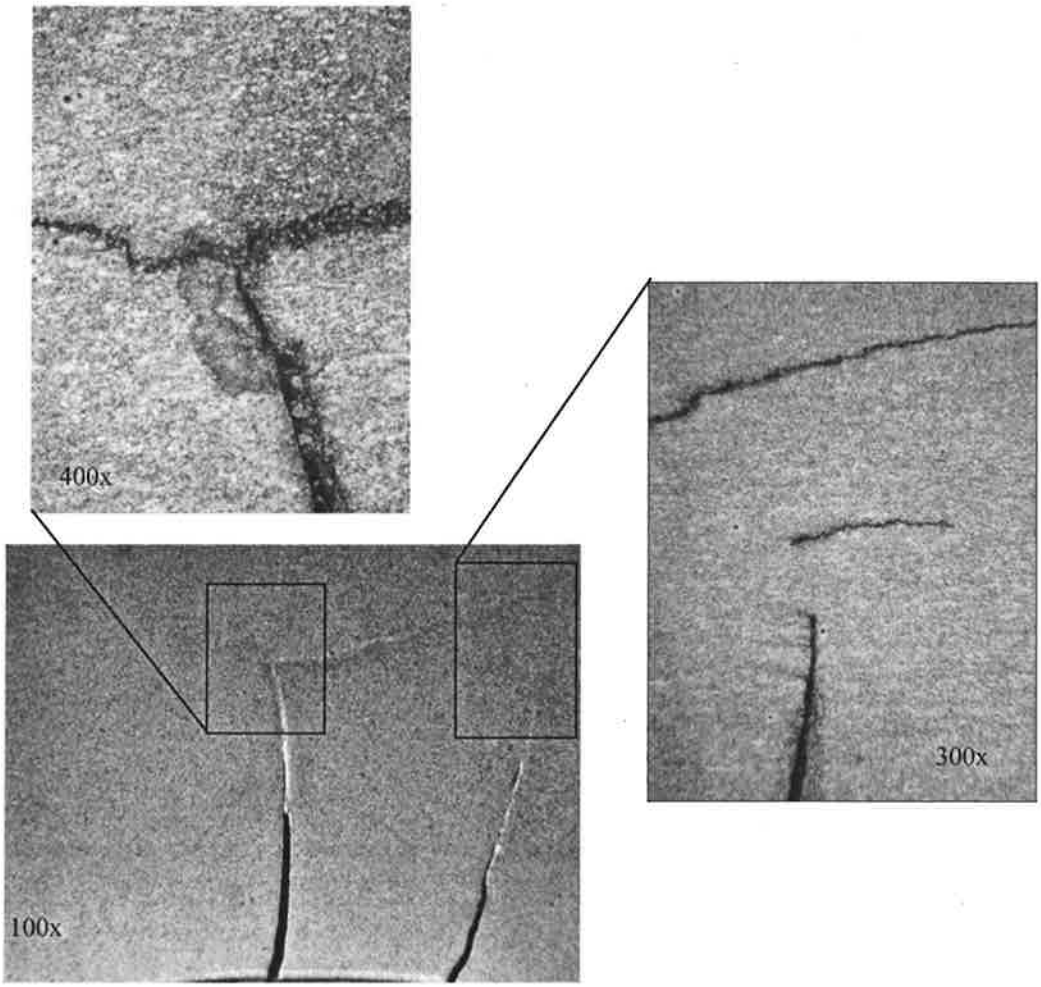


Figure 6. Crack kinking found in large scale laboratory specimen turned out to be a new crack initiation.

3.2 *Small scale CFRHS sections*

Sections of CFRHS tubes were fatigue tested using external compressive loading as shown in Fig. 7. These small scale CFRHS specimens were made of structural steel with yield strength f_y = 650 MPa. Loading was applied to opposite corners so as to create pure bending condition in the corners. Some specimens were found to fail from the corners under applied cyclic compressive stresses while other specimens fractured in the corners subject to applied cyclic tension. In all these cases the cracks propagated relatively straight through the wall thickness as seen in Fig. 7. Similar crack path have previously been reported by Bäckström et al. [6].

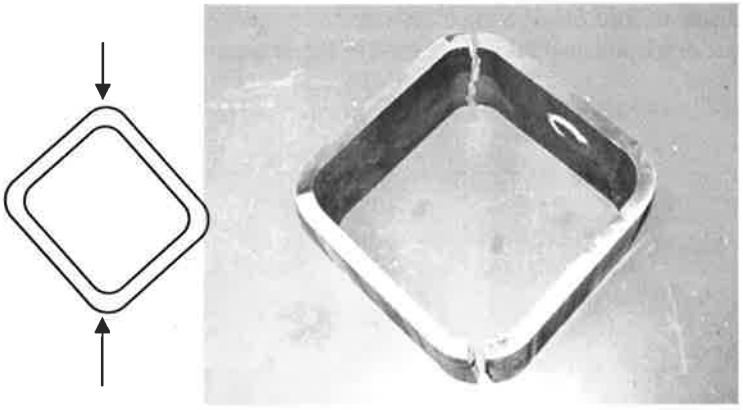


Figure 7. Straight crack paths were found in a small scale specimen cut from CFHS tubes subject to compressive external loading.

3.3 *Small scale "L" specimens*

Small scale "L" shaped plate specimens were manufactured both by cold forming small plate sections and by cutting "L" shaped corner sections from existing CFRHS tubes. These were fatigue tested using compressive loading as shown in Fig. 8. In the case of specimens produced by plate bending, the initially straight plates were bent in a single operation with special tools to achieve a final angle of 90° angle with a defined inner radius, r. The radius and the material were chosen so as to be similar to the large beam-type specimens.

During testing, the specimens were clamped along one end and an external cyclic force, ΔF was applied at the start position of the corner, see Fig. 8. This loading mode induced a bending stress in the corner region similar to that produced by the compression loading shown in Fig.7. However, the test configuration in Fig. 8 also produced a significant alternating shear stress component. The local stress state in the corner region was, therefore, similar to the stress state in the large beam-type specimens.

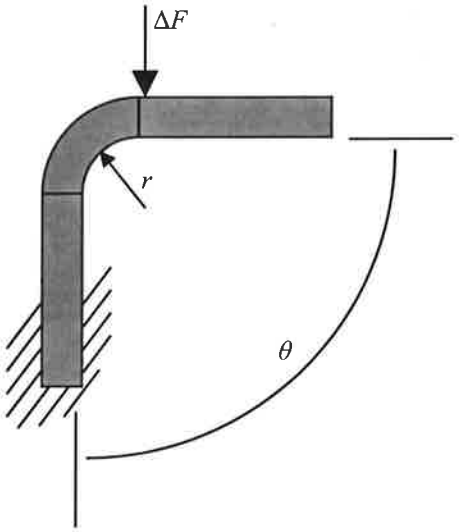


Figure 8. Loading configuration for small scale "L" plate specimens.

Fatigue cracking was always observed to initiate on the inside surface of the corner. Curvilinear crack paths similar to the service failure crack (see Fig. 1) were observed in specimens cut from CFRHS tubes. Specimens formed by simple bending of plates with nominal yield strengths greater than 650 MPa had straight crack paths. Typical paths are seen in Fig. 9.

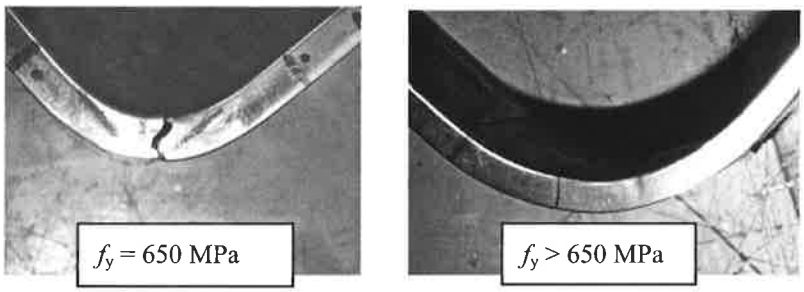


Figure 9. Samples of curvilinear and straight crack paths observed in small scale "L" plate specimens.

4 NUMERICAL MODELLING

In order to better understand the initial residual stress state, the finite element method was used to simulate the cold forming operation for a structural tube [7]. The computed through-thickness distribution of tangential residual stresses is shown in Fig. 10. Numerous simplifying assumptions were made in this simulation and, as a result, the computed residual stresses were not in full agreement with the stresses measured by X-ray diffraction. The errors were mainly due to the overly simple material model implemented and the use of a 2D simulation. The true cold forming process for a CFRHS tube is 3D and it difficult to simplify.

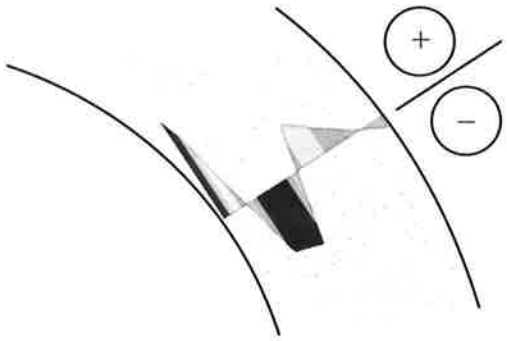


Figure 10. Tangential residual stresses of CFRHS according to FE-simulation.

Residual stresses along two crack paths were assessed. Paths were chosen based on the observed crack paths in the laboratory test specimens and the in-service failure. The first path was "S" shaped (see Fig. 1) while the second was straight through the thickness in the corner region (see Fig. 7). The stress intensity factors (SIF's), K_I and K_{II}, as a function of crack length due to the residual stress alone were calculated. These are shown in Fig. 11. For calculations residual stress distribution with crack growth was assumed, however, contact between the crack faces was not included so SIF values contain some errors, especially in the cases where $K_I < 0$. Some mesh simplifications were also used. K_I and K_{II} could be calculated based on the J-integral or its domain integral conversion [8]. For Fig. 11 it is clearly seen that K_I for the straight crack approaches zero

at a crack depth of about 3 mm. On the other hand, K_I remains greater than zero in the case of a curvilinear crack path. If the external cyclic loading also produces compressive stresses, it is clear that a straight crack path through the thickness is not physically possible. On the other hand, the curvilinear crack will remain open even at greater crack depths. The fact that the mode II SIF becomes negative is not significant.

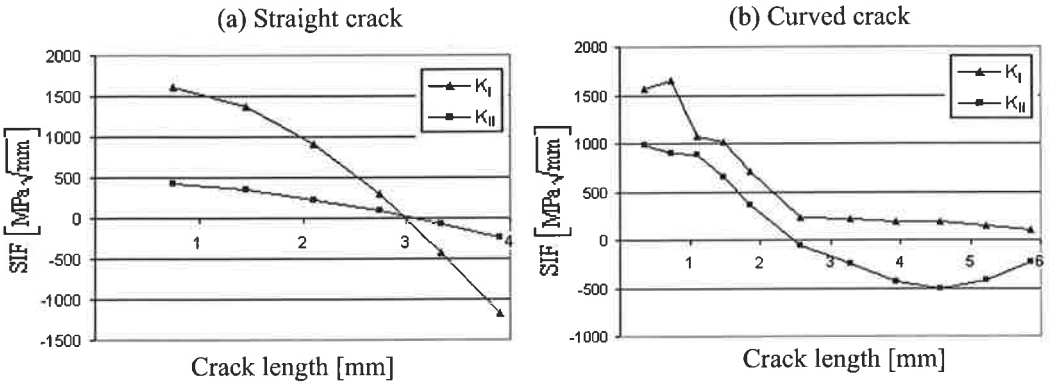


Figure 11. Stress intensity factors for straight (a) and curved crack (b) in case of no external loading.

Estimates of the residual tangential stress redistribution due to crack advance were also obtained for the two crack paths. The residual stress field for the "S" shaped crack case evolves into a simple tension-compression field as the crack approaches the neutral plane, see Fig. 12.

In the case of a straight crack, the tensile residual stresses ahead of crack tip slowly diminish and compressive stresses are developed, see Fig. 13.

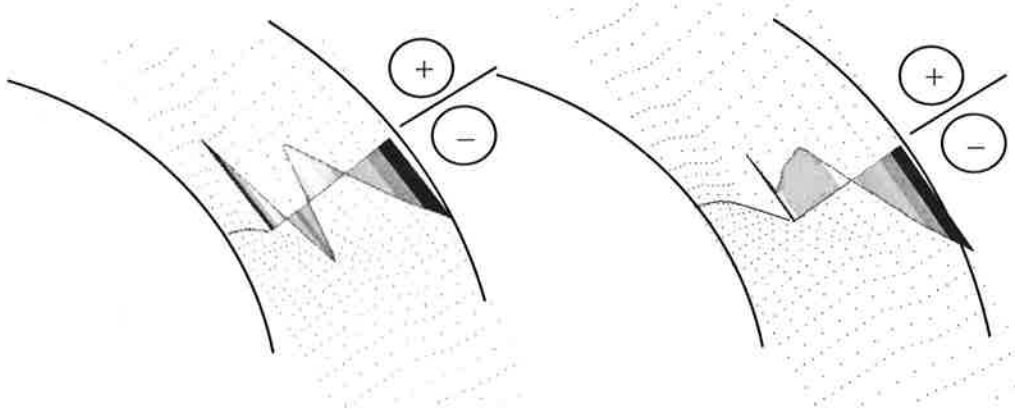


Figure 12. Estimated redistributed residual stresses tangent to the corner in the case of a curved crack.

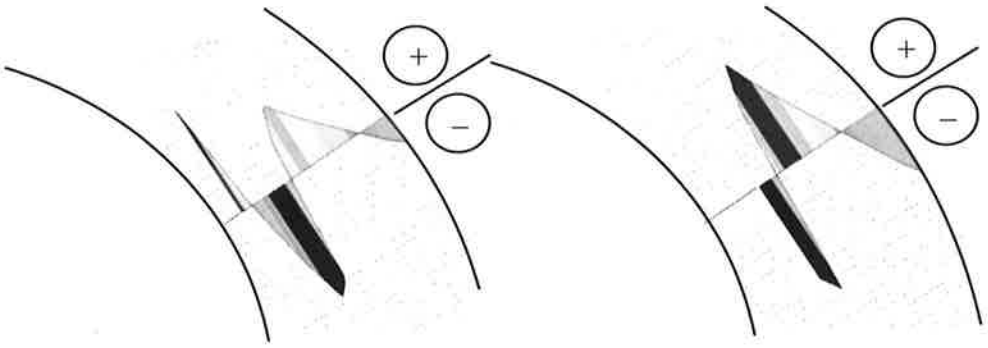


Figure 13. Estimated redistributed residual stresses tangent to the corner in the case of a straight crack.

5 DISCUSSION

The local tangential stress on inside surface at the corner is the sum of the stress induced by the external loading and tensile residual stress. In general the local tangential stress under alternating external loading can be stated as

$$\sigma^{\theta}_{min} = \sigma^{\theta res} + \sigma^{\theta ext}_{min} \quad (1\ a)$$

$$\sigma^{\theta}_{max} = \sigma^{\theta res} + \sigma^{\theta ext}_{max} \quad (1\ b)$$

$$\Delta\sigma^{\theta} = \sigma^{\theta}_{max} - \sigma^{\theta}_{min} \quad (2)$$

where $\sigma^{\theta res}$ is local tangential residual stress, $\sigma^{\theta ext}$ is local tangential stress caused by remote external loading and $\Delta\sigma^{\theta}$ is the range of the local tangential stress caused by the external loading cycle. The range of local stress is clearly independent of the residual stress while the mean stress depends on both the residual stress and the applied external stress.

Tensile or compressive mean stresses are known to have a strong influence on whether a crack grows predominantly by mode I or mode II [9,10]. Once a crack begins to branch, the mean stress does not influence the orientation of the brack crack. Branch cracks initially propagate in the direction perpendicular to the local maximum tangential stress (σ^{θ}_{max}) and gradually changes to the direction perpendicular to the remote principal stresses [11,12]. Residual stresses would therefore be expected to have an influence on the dominant mode of crack growth but do not promote growth in arbitrary direction. The issue of residual stresses is of course further complicated by the fact that they are secondary stresses that continuously redistribute with crack advance so as to be self-equilibrating. Residual stresses also frequently have large grandients that change from tensile to compressive over very short distances.

As mentioned previously, the loading configuration illustrated by Fig. 8 produces both tangential stresses and shear stresses in the corner region. A fatigue crack that initiates on the inner surface of the corner will initially propagate in the direction perpendicular to the local maximum tangential stress range. Even though the external load produces a local stress that alters between zero and a compressive minimum, the crack will be open during the cycle due to the high tensile residual stresses. Eventually the mode I crack growth rate will reduce or go to zero as the crack proceeds into the region of reduced tensile residual tangential stresses or even into the region of compressive residual stress. The driving force for a straight crack is very small due to the significant compressive residual tangential stresses associated with this path, see Fig. 13. If the

alternating shear stress was sufficiently small, the crack would arrest as was observed in the previously referenced cases [2, 3].

With relatively small external loads, the crack advance is small and the crack path turns. The turning is aided by the redistribution of the residual stresses in which the stresses tangential to the crack tip plane remain tensile. The alternating shear stresses near the neutral plane of the plate combined with a small tensile K_I are sufficient to continue crack advance on the curved path. The propagation progresses in this fashion until the crack reaches the neutral plane and the crack has turned to direction of the maximum shear stress. After a short period of crack extension along the centre of the plate, mode I crack growth caused by the external loading is again preferable and a branch crack develops that rapidly advances through the plat thickness. This process produces the "S" shaped crack path as seen in Fig. 1.

The zigzag crack path seen in Fig 5 for the large laboratory specimen was subject to loading similar to the in service beam shown in Fig 1. The in-service beam had an "S" shaped crack. The only significant difference is the amplitude of applied loading which is greater in the case of the zigzag crack. The large stress amplitude leads to the network of tensile and shear cracks observed in Fig. 6. In this case the main crack does not turn as in the case of a single crack, but the straight mode I crack links to a mode II crack initiated near the neutral axis of the plate. The final branch crack the leads to failure is the same for both the "S" crack and the zigzag crack. The increased stress amplitude probably also has some influence on the residual stress redistribution, e.g., Lee et al. [13] found that residual stress redistribution was affected also by the cyclic loading range. This issue, however, is not fully understood for the cold-formed corners studied here and remains an area for further investigation.

An interesting contrast exists for the small "L" shaped specimens cut from CFRHS tubes and those formed by bending of simple plates. An "S" shaped crack, similar to that found in the failed-in-service beam, was observed for specimens cut from tubes sections while straight cracks occurred for plate bending specimens. The external loading conditions were identical and the major difference was the residual stress state. The plate bend specimens had somewhat higher yield strength and therefore potentially higher residual stresses. However, the corner radii and plate thickness were slightly different. Residual stress measurements for the simple bending case have also not yet been done and this difference in behaviour is not yet fully explainable.

The loading configuration shown in Fig. 7 produces predominantly tangential normal stresses and only negligible shear stresses. The resulting crack paths observed for the small scale CFRHS specimens were nearly straight (Fig. 7). In this case mode I crack initiation is assisted by the tensile residual stresses. The combination of external loading and redistributing residual stresses does not lead to a crack arrest condition.

6 CONCLUSIONS

Fatigue crack propagation in cold-formed corners of high-strength structural steel plates has been investigated using large- and small-scale laboratory specimens, finite element analysis and linear elastic fracture mechanics. The cold-formed corners have complex residual stress states and the stress state in the corner region was multi-axial due to the geometry/loading combination. Straight, zig-zag and "S" shaped crack paths were observed. The compressive tangential residual stresses due to cold forming and the alternating shear stresses due to the external loading were found to have a dominating influence of the observed fatigue crack path.

7 REFERENCES

1. Heinilä, S., Björk, T., Marquis, G., Bäckström, M., Ilvonen, R. (2004) In: *Proceedings of ECF15*, Stockholm.
2. Greasley, A., Naylor, S. G. W. (1986) *Engng Fract. Mech.*, **24**, 717-726.
3. Hermann, R. (1994) *Fatigue Fract. Engng Mater. Struct.* **17**, 93-103.
4. Sumi, Y. (2003) In: *Proceedings of FCP 2003*, Parma.
5. Filippini, M., Giglio, M., Gorla, C., Rosa, F. (2003) In: *Proceedings of FCP 2003*, Parma.
6. Bäckström, M., Savolainen, M., Ilvonen, R., Laitinen R. (2002) In: *Proceedings of FATIGUE 2002*, Stockholm, 277 – 302 .
7. Heinilä, S., Björk, T., Marquis, G., Ilvonen, R. (2004) In: *Proceedings of ECCOMAS 2004*, Jyväskylä.
8. Lei, Y., O'Dowd, N. P., Webster, G. A. (2000), *Int. J. Fract.*, **106**, 195-216.
9. Marquis, G. and Socie, D., (2000) *Fatigue Fract. Engng Mater. Struct.* **23**, 293-300
10. Pook, L. P., (2002) *Crack Paths*, WIT Press, Southhampton.
11. Murakami, Y., Takahashi, K., Toyama, K. (2003) In: *Proceedings of FCP 2003*, Parma.
12. Tanaka, K., Kato, T., Akiniwa, Y. (2003) In: *Proceedings of FCP 2003*, Parma.
13. Lee, Y., Chung, C., Park, Y., Kim, H. (1998) *Int. J. Fat.* **20**, 565-573.

HITSIN LAADUN MÄÄRITYKSESTÄ

T. BJÖRK

Konetekniikan osasto
Rakenne ja lujuustekniikan laboratorio
PL 20, 53851 LPR 85
e-mail: timo.bjork@lut.fi

TIIVISTELMÄ

Hitsin laatu voidaan määrittää tarkoituksenmukaisuusperiaatteen mukaisesti monella eri tavalla. Yleisimmin tarvitaan luokittelua kestävyyden perusteella. Vaikka hitsin laatuluokitus standardi on olemassa, ei se palvele kustannustehokkaasti nykyisiä kasvaneita laatuvaatimuksia. Energian kallistuminen, keveysvaatimus tai hyötykuorman kasvattaminen edellyttävät teräksen käyttöasteen kohottamista etenkin liikkuvissa rakenteissa. Tämä johtaa korkealujuuksisten teräksen hyödyntämistarpeisiin myös väsyttävästi kuormitetuissa rakenteissa, mikä puolestaan edellyttää hitsien suunnittelun ja valmistuksen laatutason kohottamista. Keskeiseksi muodostuu tällöin laadun määräytymisen ymmärtäminen. Laatu olisi pystyttävä määrittämään kvantitatiivisesti siten, että se korreloi väsymislujuuden kanssa. Tässä on esitetty menetelmän perusteita laadun määrittämisestä silloin, kun lujuus on oleellinen laatuparametri.

1. JOHDANTO

Hitsin laadulla tarkoitetaan hitsin kykyä suoriutua tehtävästään elinkaarensa aikana. Alilaatua tulee välttää, koska se voi aiheuttaa vakavia turvallisuus- tai kustannusseuraamuksia. Ylilaatu puolestaan on varmuusmarginaalin ylittävä osuus, jota asiakas ei suostu maksamaan. Ylilaatu voi olla sikäli harmiton, että sen tuottaminen ei välttämättä aiheuta ylimääräisiä kustannuksia. Huonompaa laatua ei kannata tuottaa, mitä prosessissa luontaisesti syntyy.

Tarkoituksenmukaisen laadun tekeminen edellyttää suunnittelun ja valmistuksen yhteistyötä. Suunnittelija asettaa tavoitetason ja antaa valmistukselle mahdollisuudet sen toteuttamiselle. Valmistuksessa on hyvä ymmärtää vaatimusten taustat sillä tasolla, että suunnittelusta tuleva vaatimukset ovat motivoivia. Vain varmuuden vuoksi laadun teettäminen tulee kalliiksi ja vie pohjan vaatimuksilta silloin, kun laatua todella tarvittaisiin. Laatu on kohdennettava oikeaan paikkaan. Korkeatasoisen täsmälaadun tekeminen maksaa ja saakin tehdä sitä, koska se korvautuu esim. suurempana hyötykuormatehona, pienempänä energiankulutuksena tai pitempänä käyttöikänä.

Suomessa ja muissa Pohjoismassa on samantapainen hitsattujen rakenteiden tuotantokulttuuri. Sarjakoot ovat pienehköjä, jolloin kilpailu käydään innovatiivisella tuotesuunnittelulla ja valmistuksella. Tuotteet pitää saada asiakkaalle houkuttelevaksi esim. alhaisemmilla käyttökustannuksilla. Hitsatuissa teräsrakenteissa tämä johtaa hitsauslaadun merkityksen ja myös yhä lujempien teräslaatujen käytön lisääntymiseen erityisesti väsyttävästi kuormitetuissa rakenteissa. Lujasta teräk-

sestä tehtävät rakenteet on osattava suunnitella ja valmistaa niin laadukkaasti, että suuri lujuus näkyy myös käyttöasteen kasvuna. Tulevaisuudessa energian hinta tullee säilymään korkeana korostaen liikkuvien laitteiden painon merkitystä käyttökustannuksissa. Tämän kehityksen myötä laatuun panostaminen on palkitsevaa.

Hitsin laatu kuten laatu yleensäkin määritellään usein vain kvalitatiivisesti parempi- huonompiasteikolla. Tällainen laadun määrittely toimii vain kahta vaihtoehtoa keskenään verrattaessa, jos siinäkään. Vertailussa voi nimittäin esiintyä tekijöitä, jotka vaikuttavat vastakkaisiin suuntiin. Siten todellinen ja yksiselitteinen laadun määritys edellyttää kvantitatiivisuutta. Ensiksi täytyy vain sopia, mikä on se suure, joka kuvaa hitsin laatua yksiselitteisesti. Valitettavasti tällaista yksiselitteistä suuretta ei ole olemassakaan, mutta hyviä kompromisseja kuitenkin.

Hitsin laatuluokitusta käsittelevä standardi ISO 5817 [1] jakaa hitsit laatuluokkiin B, C ja D. Hitsin laatuluokka on vain apusuure, jonka avulla yritetään lajitella joukko laatua kuvaavia tekijöitä yhteen karsinaan. Menetelmän heikkous piilee siinä, että kokonaisuutta käsitellään huonoimman yksittäisen tekijän mukaisesti, vaikka todellisuudessa se muodostuu eri osatekijöiden yhteisvaikutusta. Lisäksi lajitteluun on otettu mukaan parametreja, joiden vaikutukset laaduntuottotekijöinä ovat kyseenalaisia.

Tässä esityksessä käsitellään edellä mainittuja laatuluokkia, mutta niiden sisältöä pyritään selkeyttämään ja määrittämään laadulle numeerinen arvo vaativassa sovellusympäristössä, eli väsyttävästi kuormitetuissa rakenteissa.

2. HITSIN LAADUN MÄÄRITYSPERUSTEITA

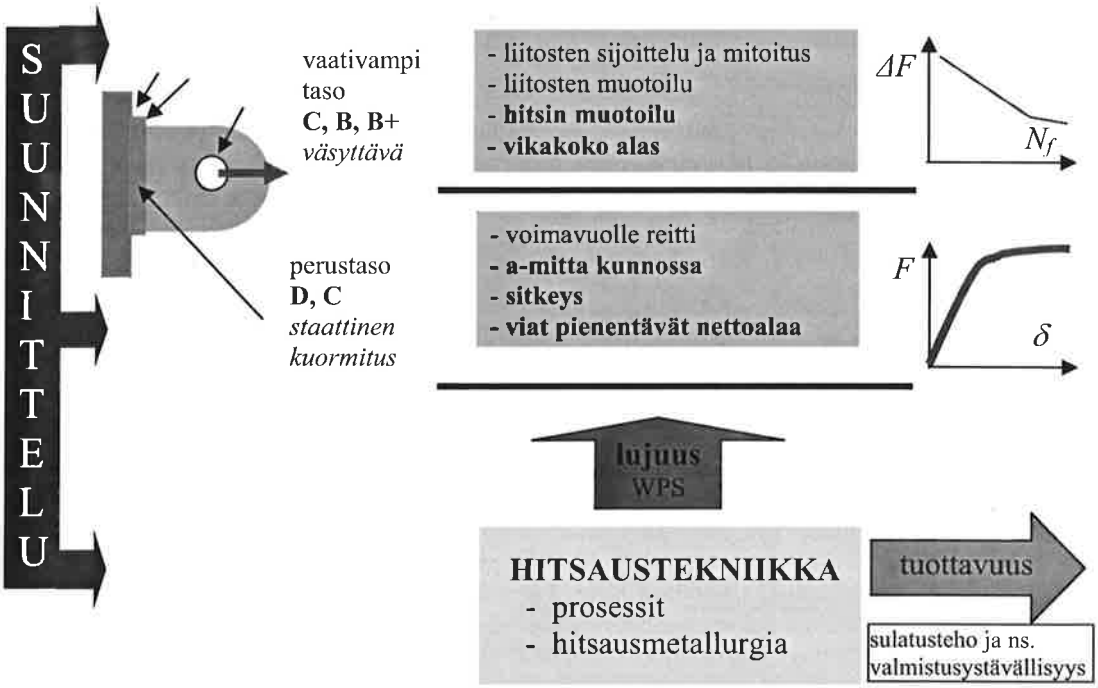
2.1 Tarkoituksenmukaisuusperiaate

Johdannossa esitetystä laadun määritelmästä käy ilmi, että elinkaariaikainen toiminto määrittää laadulle tarkennetut kriteerit. Sellaisia kriteerejä voivat olla mitkä tahansa hitsausliitoksen pää- tai sivutoimintoon liittyvä ominaisuus, kuten kovuus, kulumiskestävyys, korroosionkestävyys, tiiviys, ulkonäkö, pintakäsiteltävyys, jne. Vanha ruotsalaisen hitsin laatuluokitus otti tämän tärkeän tarkoituksenmukaisuusperiaatteen huomioon ja hitseille pystyttiin vaatimaan käyttötarkoituksen mukaisesti oikeita ominaisuuksia painottava laatu. Tämä tarkoituksenmukaisuus viestitettiin valmistukseen lisäämällä laatuluokituksen perään vaadittavaa ominaisuutta kuvaava lisäkirjain, kuten Y, T, U, jne. Laadun lisämääre mahdollistaisi sellaisten hitsien käytön, joilla ei olisi mitään muita vaatimuksia kuin lisämääreen mukainen ominaisuus. Jos tällaisen luokaton hitsin merkittäisiin E-kirjaimella, esim. ET tarkoittaisi liitosta, jolle asetetaan vain tiiveysvaatimus. Tällainen tarkoituksenmukainen laadun määritys on kustannustehokas tapa täyttää asiakkaan tarpeet, mutta jostain syystä sitä ei ole kelpuutettu uuteen kansainväliseen hitsin laatuluokkastandardiin. Tässä esityksessä laatu kohdennetaan tarkoittamaan lujuutta ja erityisesti väsymislujuutta.

2.2 Kuormituksen luonteen vaikutus

Kuormituksen luonne on yksi tärkeä laadun tekemistä ohjaava kriteeri, jota ei ole otettu huomioon riittävän selkeästi nykyisissä laatuohjeissa. Laatutasot voidaan jakaa kuvan 1 mukaisesti perustasoon ja vaativampaan tasoon kuormituksen väsyttävyyden mukaisesti. Staattisesti kuormitetuille liitoksille, joissa ei ole haurasmurtumavaaraa, laatu voidaan määrittää perustason mukaisesti. Perustason keskeisin vaatimus on hitsin ja lämpövyöhykkeen (HAZ) sitkeys, joka taataan tarvittaessa esim. esituotannollisella menetelmäkokeella varmennetun hitsausohjeen (WPS) avulla. Saman

ohjeen pitäisi taata se, ettei särömäisiä vikoja synny hitsauksessa. Materiaalin lujuus on tärkeä ominaisuus, mutta sen menetystä hitsauksessa voidaan korvata materiaalin lisäyksellä. Sen sijaan sitkeyden menetystä ei voida korvata, joten se on kaikista tärkein hitsausliitoksen perusvaatimus. Kestävyyden määritys lepää tämän oletuksen varassa. WPS on siten pohjimmiltaan ohjeistus, jolla taataan hitsin muodonmuutoskyky ja lujuus. Siten hitsausparametrien määrityksessä on tärkeää priorisoida aina ensin riittävän laadun tuottavat parametrit ja vasta sitten tuottavuutta parantavat asiat, ei toisin päin. Tietysti voi olla rakenteita, joiden liitosten laatuvaatimukset ovat vähäiset. Näissä valmistuskustannusten optimointi voi tuoda merkittäviä säästöjä konepajalle. Tyypillisiä tällaisia kohteita ovat staattisesti kuormitettujen rakenteiden pitkät kiinnityshitsit. Sitä paitsi hitsin laatu ja tuottavuus eivät välttämättä ole vastakkaissuuntaisia vaatimuksia, kuten kuvassa 1 symbolisesti osoitetaan. Perustason kolmas laatukriteeri on hitsin *a*-mitta. Hitsityypillä (päittäis-, piena-, V-, K-hitsi) ei ole perustason vaatimuksiin oleellista merkitystä, jos kaikissa on sama *a*-mitta. Edellä esitettyjen ehtojen täyttäminen riittää staattisesti kuormitetuissa rakenteissa. Siten hitsin laatuluokkaohjeessa ISO 5817 [1] on staattisesti kuormitetun liitoksen laatuluokkia liikaa, kaksi luokkaa riittäisi.



Kuva 1. Laatu voidaan jakaa kuormituksen mukaan perus- ja vaativampaan tasoon

Vaativampi taso tuo mukaan lisävaatimuksia hitsin geometrialle ja on tarkoitettu lähinnä väsyttävästi kuormitetuille rakenteille. Perustason laatuvaatimukset ovat tietenkin edelleen täysmääräisesti voimassa. C-luokkavaatimusten mukaiset vikaisuuskriteerit voidaan vaatia myös staattisesti kuormitetuille liitoksille, jos mahdollisen vaurion seuraukset ovat erityisen ankarat, kuten esim. painelaitteissa. Toinen perustelu C-tason vaatimuksille staattisessa kuormituksessa voisi olla liitoksen huono käytönaikainen tarkastettavuus. Tarkastuslaajuuden kytkeminen laatuluokkavaatimukseen estää saman vaikutuksen huomioon ottamisen suoraan varmuuskertoimella. Myös haurasmurtumavaaran ollessa mitoittava kriteeri vaativampia laatuluokkia käytetään staattisesti kuormitetuille rakenteille, koska tällöin perustason sitkeyskriteeri ei ole voimassa. Väsyttävästi kuormitetun liitoksen kapasiteetti määräytyy karakteristisesti heikoimman detaljin mukaisesti, kun staattisesti kuormitetun sitkeästi käyttäytyvän liitoksen kapasiteetti määräytyy koko nettoleikkauksen lujuuden perusteella. Tästä ratkaisevasta eroavaisuudesta aiheutuu se, että jos staattinen kuor-

mitus muuttuu väsyttäväksi, laajenee kriittisten detaljien lukumäärää koko rakenteessa, yksittäisessä liitoksessa ja myös yksittäisessä hitsissä. Koska väsymiskestävyyteen vaikuttaa monta parametria ja niiden yhteisvaikutuksesta lujuus voi vaihdella merkittävästi, on tarkoituksenmukaista ottaa käyttöön ainakin kolme eri lujuustasoa, joita vastaavia laatutasoja merkitään tässä symboleilla C, B ja B+. Jos oletetaan suunnittelun ja mitoituksen olevan kunnossa ja juurenpuolen väsymismahdollisuuden olevan siten eliminoitu, hitsin muoto ja rajaviivan geometria määräävät liitoksen väsymislaadun poikittain rajaviivan nähden kuormitetuissa liitoksissa. Juuren puolen väsyminen voi esiintyä huonon valmistuslaadun seurauksena, mutta sille ei voida määrittää omaa laatuluokkaa, koska väsymislujuus voi vaihdella lähes nolla-tasolta rajaviivan asettamaan laatuluokkaan saakka.

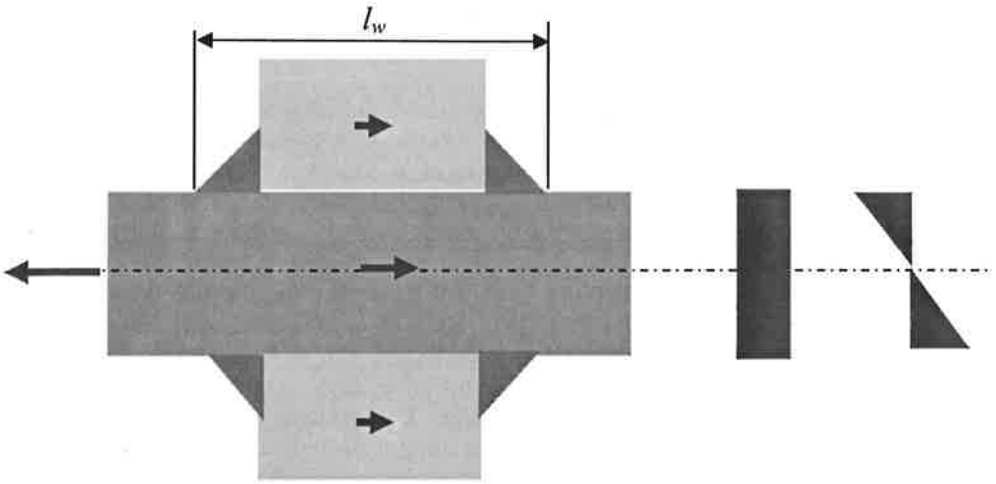
2.3 Hitsaus- ja suunnittelulaatu

Kun lähtökohdaksi otetaan juuren puolelta väsymätön hitsi, lujuustekninen laatu voidaan määrittää väsymisominaisuuksien perusteella rajaviivan geometriasta määräytyvänä. Kriittisiä rajaviivoja voi olla 1-4 hitsityypin ja kuormitussuunnan mukaisesti, monipalkohitsauksessa myös palkojen välit voivat olla kriittisiä. Valmistusteknisesti ko. rajaus typistää hitsin laadun pelkästään prosessin tuloksena syntyväksi ja siten myös sen kautta ohjattavaksi suureeksi. Muut valmistusvaiheessa syntyvät ja mahdollisesti liitoksen kestävyyteen vaikuttavat tekijät suunnittelija ottaa huomioon laskennallisesti. Tämän periaatteen mukaisesti kuvassa 2 esitetyt kulmavetäymät ja sovitusvirheet eivät ole hitsin laatuluokkiin kuuluvia laatutekijöitä, vaikka joku voisi kytkeäkin ne samalla tavalla rajaviivageometrian kanssa hitsauksessa syntyviksi tekijöiksi. Ko. poikkeamien vaikutus on mukana rakenteellisen jännityksen suuruudessa ja vaikutus on analoginen eripaksuisten levyjen aiheuttamaan epäsymmetriseen porrastukseen. Suunnittelun laatua kuvaa se, miten suuria tällaisia rakenteellisia epäjatkuvuuksia suunnittelija sallii eri kuormitustilanteissa ja miten hyvin valmistus on informoitu näistä. Ko. poikkeamat ovat syytä määritellä rakenteellisen jännityksen tekijöiksi siksikin, että niillä on aivan erilainen merkitys kalvo- ja taivutusjännityksen alaisuudessa. Tämä luokitusrajaus on ristiriidassa nykyisen ohjeistuksen [1] kanssa, mutta uusi menetelmä selkeyttäisi vastuukysymyksiä, laadun määritystä ja ohjeistusta.

Kuva 2. Sovitus- ja kulmavetäymä eivät ole laatuparametreja vaan osa rakenteellista jännitystä

Hitsin laatu ei siis määräydy koko liitoksen geometriasta, vaan se on rajaviivan geometriaan sidottu paikallinen ilmiö. Näin ollen hitsin laatu ei määräydy rakennedetaljin muotoilusta, joka on puhtaasti suunnittelijan vastuulla oleva asia. Siten laadun määritys ja vertailu täytyy tehdä jännitystasolla, joka sisältää liitoksen muodon ja siitä johtuvan vaikutuksen jännitykseen, mutta ei sisällä hitsin laadukkuudesta määräytyvää lovijännitystä. Ruotsalaisen BSK 99 [2] mukaisesti laatu on otettu mukaan nimellisen jännityksen tasolla, mikä on periaatteessa käypä ratkaisu. Ongelmana on kuitenkin nimellisen jännityksen menetelmään liittyvä suuri hajonta, kun sekä rakenteelliset että paikalliset geometriaparametrit sisältyvät samaan S-N -käyrään. Toisaalta menetelmässä esitetty laadun määritys rajaviivan liittymiskulmien mukaan ei yksin riitä luokitukseen ja laadun sisällyttäminen nimelliseen väsymisluokkaan edustaa siten lähinnä suuntaa-antavaa arvausta. Rakenteellinen jännitys on oikeampi lähtökohta laadun määrittämiselle. Tosin ongelmaksi jää edelleen se, että lovijännitys, jonka pitäisi kuvata hitsin laatua, sisältää laatua kuvaavan liittymägeometrian lisäksi myös rakennedetaljin mittojen, kuormituksen ja reunaehtojen vaikutuksen. Jotta näiden sekundaaristen tekijöiden vaikutus voitaisiin eliminoida ja laatuluokka määrittää pelkästään raja-

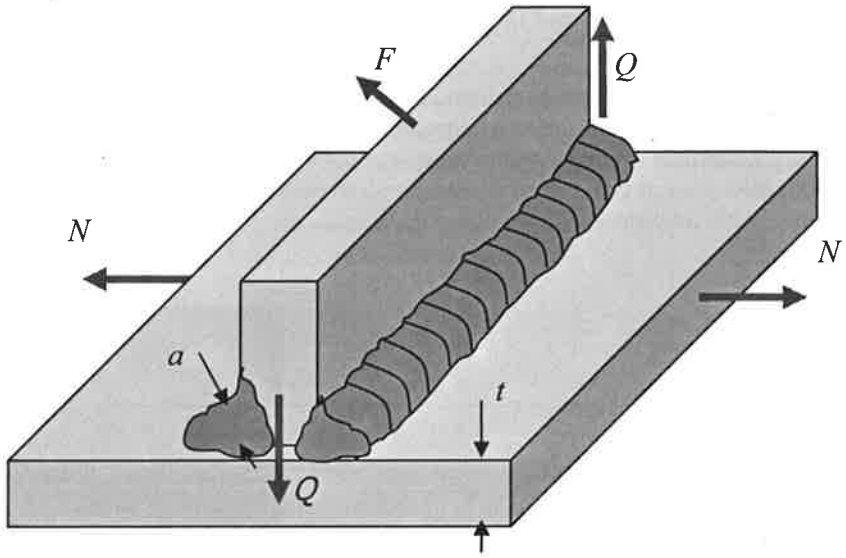
viivageometriasta, tarvitaan laadulle muutama lisäluokitus esim. Niemen ja Hobbacherin [3] esittämällä tavalla. Kahden eri rakennedetaljin hitsien rajaviivoilla voi olla sama rakenteellinen jännitys ja sama liittymisgeometria, eli laadun ja kestoikien pitäisi olla identtiset. Kuitenkin muutokset kuvassa 3 esitetyssä symmetrisyydessä tai alusrakenteen jäykkyydessä vaikuttavat lovijännitysjakauman erilaisuuden vuoksi väsymiskestävyyteen, samoin kuin rajaviivojen välinen etäisyys l_w kuormitussuunnassa. Myös hitsin kuormituksenkantoaste (0-100 %) vaikuttaa Poutiaisen [4] mukaan lovijännitysjakaumaan. Myös kalvo- ja taivutusjännityksen suhde vaikuttaa kestoikään, samoin kuin seinämäpaksuus, mutta nämä seikat voidaan ottaa yleisesti käytetyillä korjaustermeillä huomioon..



Kuva 3. Rajaviivan geometria ja reunaehdot yhdessä määräävät lovijännityksen suuruuden

2.4 Liitostyypin vaikutus

Liitostyyppi kuormituksineen vaikuttaa myös laadun määräytymiseen. Kuvassa 4 on esitetty T-liitos, jota kuormittavat voimat F, Q ja N. Mikäli mikä tahansa primaarinen voimakomponentti (F) kulkee liitoksen läpi, on kyseessä voimaliitos. Kiinnitysliitoksen tunnistaa siitä, että sitä kuormittaa kahden rinnakkain toimivan komponentin yhteen sitomiseksi tarvittava leikkausvoima (Q). Varusteluliitoksen läpi ei kulje lainkaan primaarista voimavuota (N), ainoastaan siirtymätilan jatkuvuuden yhteensovittamiseksi tarvittava hajavuo. Jos juuren puolen väsymismahdollisuus on eliminoitu, hitsin a-mitat määräytyvät staattisen kestävyyden ja lämmön tuonnin perusteella liitostyypeittäin lausekkeen 1 mukaisesti. Lausekkeessa β -kerroin ottaa huomioon perusmateriaalin ja hitsin lujuuden suhteen, γ ja γ_{M2} ovat kuormituksen ja materiaalin osavarmuuskertoimia, l on hitsin pituus, f_u on perusmateriaalin murtolujuus, n on rinnakkaisten hitsien lukumäärä, S ja I ovat poikkileikkauksen staattisen- ja jäyhyysmomentin arvot ja t liitettävien osien suurin paksuus. Voimaliitoksessa alimitoitus on kohtalokasta, kiinnitysliitoksen ongelmana on usein liian isot a-mitat aiheuttaen ylimääräisiä valmistuskustannuksia ja muodonmuutoksia. Kaikkien hitsien täytyy täyttää lämmöntuontirajoitukset, mutta varusteluliitoksessa se on ainoa kriteeri hitsin a-mitalle.



Kuva 4. Liitoksen luokittelu voima, kiinnitys ja varusteluhitsiksi määräytyy kuormituksesta

$$a = \frac{\sqrt{3}\beta\gamma\gamma_{M2}F}{\sum l f_u}$$

$$a = \frac{\sqrt{3}\beta\gamma\gamma_{M2}QS}{nIf_u} \qquad (1)$$

$$a = \sqrt{t} - 0.5$$

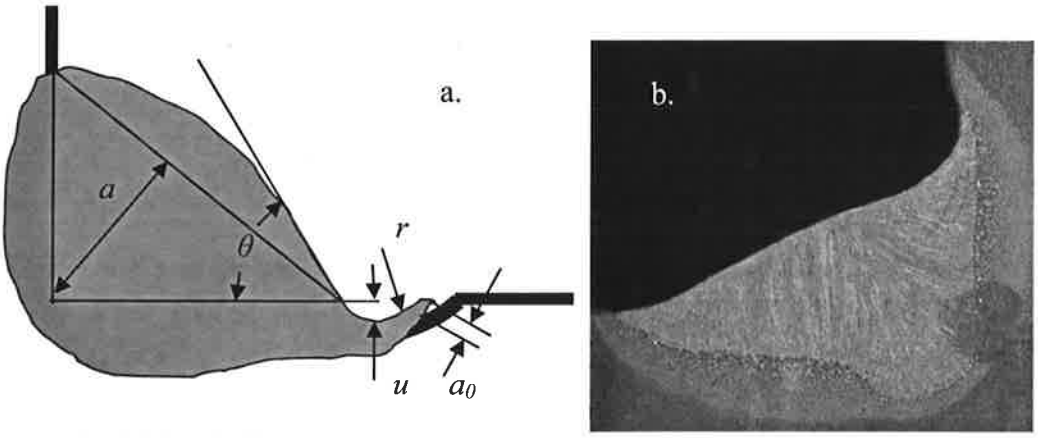
Väsymiskestävyyden mukaisessa tarkastelussa voimaliitoksen kuormitetun rajaviivaan juohevuuteen panostaminen parantaa liitoksen laatua. Vastaavasti varusteluliitoksessa kuormitetun osan rajaviiva on aivan yhtä merkittävässä osassa laadun tekemisessä kuin voimaliitoksen vastaava rajaviiva. Kiinnitysliitoksessa rajaviivan poikittaisgeometrialla ei ole oleellista merkitystä väsymisominaisuuksiin ja laatu määräytyy hitsin pitkittäissuuntainen juohevuudesta. Tämä erittely olisi syytä sisällyttää laatuluokitukseen, jossa yleensä tarkastellaan vain liitoksen poikkileikkausta pituussuunnasta ja kuormituksesta piittaamatta.

2.5 Hitsin laatuparametrit

Hyvästä suunnittelulaadusta viestittäisi sellainen tilanne, että rakenteen väsymisominaisuudet määräytyisivät kiinnityshitsien laadusta. Käytännössä usein kuitenkin voima ja varusteluhitsit ovat kriittisiä ja koko rakenteen käyttöaste määräytyy niiden laadusta. Siten kriittisiä laadun tekijöitä ovat rajaviivaan nähden poikittainen lovijännitys tai yleistetymmin jännitysintensiteettikerroin K, mikäli hitsin rajaviivalla on kasvukelpoinen alkusärö. Uuden tuotteen jännitysintensiteettikerroin sisältää rakenteellisen jännityksen σ_{hs} lisäksi valmistuksessa syntyvän alkusärön suuruuden a_0 ja särön muodon korjausfunktion $Y(a_0)$ sekä varsinainen laadun huomioonottavan lovijännityskertoimen M_k. M_k -kerroin määräytyy kuvan 5a mukaisesti hitsin rajaviivan liittymiskulmasta θ, pyöristyssäteestä r sekä alkusärön koosta a_0.

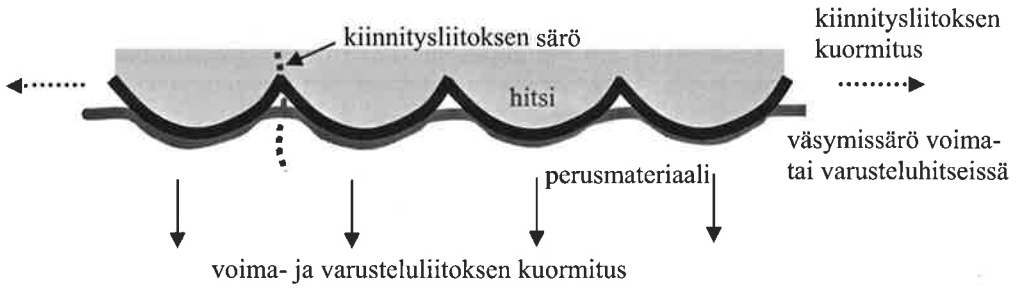
$$K = \sigma_{hs} M_k(\theta, r, a_0)\sqrt{\pi a_0} Y(a_0) \quad (2)$$

Mikäli alkuvika a_0 on terävästä reunahaavasta aiheutuvaa, on jännitysintensiteetti avaavaa moodia (mode I), kun taas kylmäjuoksun aiheuttamaa liitosvirhe synnyttää leikkaavan moodin (mode II) tai näiden yhdistelmän. Jännitysintensiteetin suuruuteen vaikuttaa oleellisesti myös se, sijaitseeko sularajalla mahdollisesti oleva alkuvika pyöristyksen pohjalla vai sen vieressä. Vastaavat parametrit ovat muissakin hitsityypeissä (päittäis-, V-, K-hitsi...).



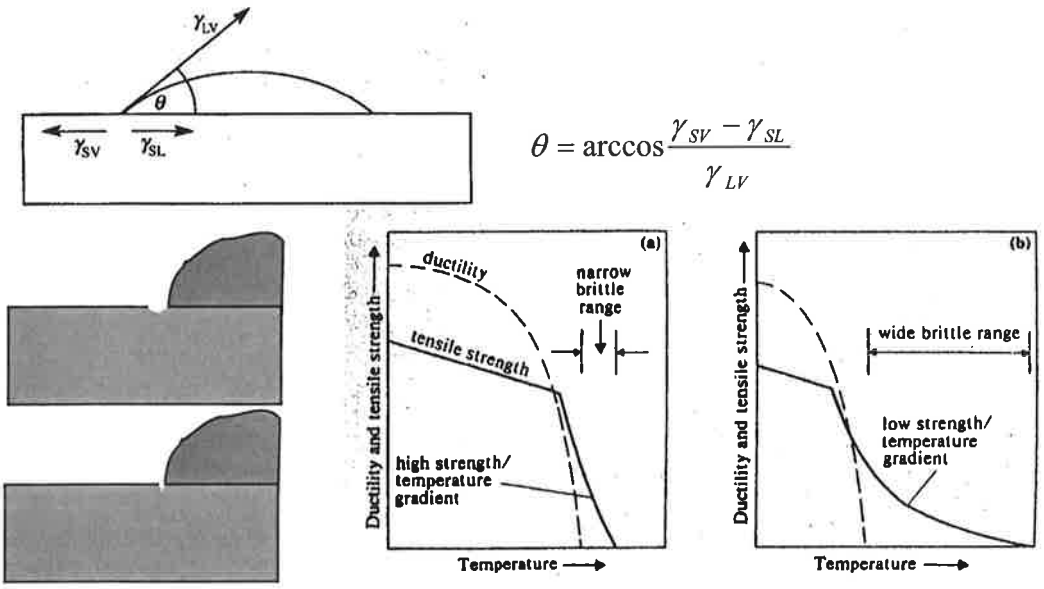
Kuva 5. Pienahitsin rajaviivan laatuparametrit a. ja hyvä rajaviivaliittymä pienahitsissä b.

Muita sekundaarisempia hitsin rajaviivan laatuparametreja ovat mahdollinen pyöreäpohjaisen reunahaavan syvyys u ja hitsin kupu, hitsin leveys, hitsin yleinen koveruus-kuperuus sekä hitsin a-mitta. Myös kuvan 6 mukainen hitsin pituussuuntainen rajaviivan muoto vaikuttaa poikittaissuunnan väsymisominaisuuksiin. Robottihitsauksessa kannattaa käyttää ns. vaaputusta, koska aaltoileva muoto on rajaviivan nähden poikittain kuormitetuissa voima- ja varusteluliitoksissa parempi kuin suoraviivainen. Kiinnitysliitoksessa sen sijaan aaltoileva muoto synnyttää lovia ja on siten laatua heikentävä.



Kuva 6. Vaaputus parantaa laatua voima- ja varusteluhitseissä ja huonontaa kiinnityshitseissä

Hitsin rajaviivageometria määräytyy hitsin jähmettymisvaiheessa. Vaikuttavina tekijöinä liittymiskulmaan ja reunahaavaan ovat kuvan 7 mukaisesti mm. pintajännitys, kaaripaine, vapaalangan pituus ja gravitaatio. Ko. kuvassa rajapinnan energiatasoja edustavat γ_{SV} (kiinteä-kaasu), γ_{SL} (kiinteä-neste) ja γ_{LV} (neste-kaasu) rajapintojen pintajännitykset. Reunahaavan pohjan terävyys voi aiheutua Lancasterin [5] mukaan liian hitaasta lujuuden kehityksestä sulan jähmettymisvaiheessa.



Kuva 7. Rajaviivageometrian syntymisen metallurgiset taustat [5]

3. LAADUN JA VÄSYMISKESTÄVYYDEN VÄLINEN YHTEYS

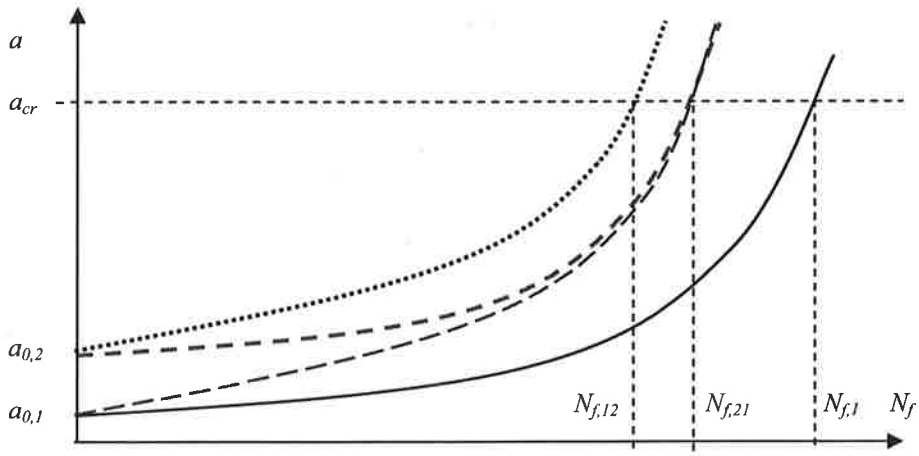
Hitsin laatu määräytyy hitsin rajaviivageometriasta, jota ohjataan hitsausparametreilla [6]. Rajaviivageometria yhdessä sekundaarisempien tekijöiden kanssa määrittää liitoksen väsymiskestoiän N_f, joka voidaan määrittää rakenteellisen jännitysvaihtelun $\Delta\sigma_{hs}$ ja murtumismekaniikan avulla.

$$N_f = \left[\frac{FAT_Q}{\Delta\sigma_{hs}}\right]^m \cdot 2\cdot 10^6 = \int_{a_0}^{a_f} \frac{da}{C\left[\Delta\sigma_{hs} M_k(\theta,r,a)Y(a)\right]^m} \tag{3}$$

Yhtälössä C ja m ovat murtumismekaanisia särönkasvuvakioita ja a_f lopulliseen vaurioon johtava särön koko a. Lausekkeista voidaan määrittää hitsin laadukkuutta kuvaava rakenteellisen jännitysvaihtelun mukainen laatuluokka FAT_Q.

$$FAT_Q = \sqrt[m]{\frac{1}{2\cdot 10^6 C}\int_{a_0}^{a_f} \frac{da}{\left[M_k(\theta,r,a)Y(a)\right]^m}} \tag{4}$$

FAT_Q=100 kuvaa normaalia konepajalaatua C ja siitä suuremmat arvot parempaa laatua. Liitoksen rajaviivalta alkavan särönkasvun ja väsymiskestoiän välinen yhteys on tyypillisesti kuvassa 8 esitettyjen käyrien muotoinen. Sama kestoikä $N_{f,2l}$ voidaan saavuttaa eri parametriyhdistelmällä. Rakenteellisen jännitysvaihtelun ollessa vakio käyrän loivuus kuvaa liitoksen juohevuutta (suuri r/θ -suhde) ja sen sijainnin määrää alkusärön suuruus (a_0). Käyrästöstä nähdään, että huonoa liittymää ei pieni alkusärö paranna ja vastaavasti alkusärön pienuudesta on eniten hyötyä juohevassa liittymässä.

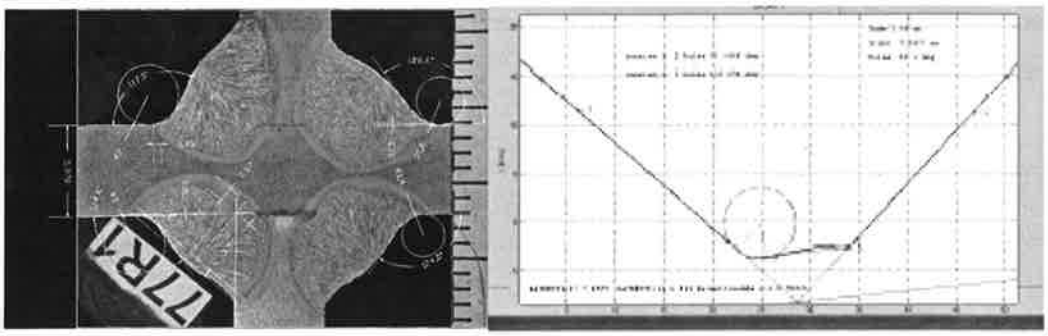


Kuva 8. rajaviivaliittymän ja kestoiän välinen yhteys skemaattisesti

Nykänen [7, 8] on esittänyt lineaariselastiseen murtumismekaniikkaan (LEFM) perustuvilla laskelmillaan laadun tasa-arvokäyriä ja voimassaolorajoja eri rajaviivan parametriyhdistelmille kuvan 8 periaatteen mukaisesti.

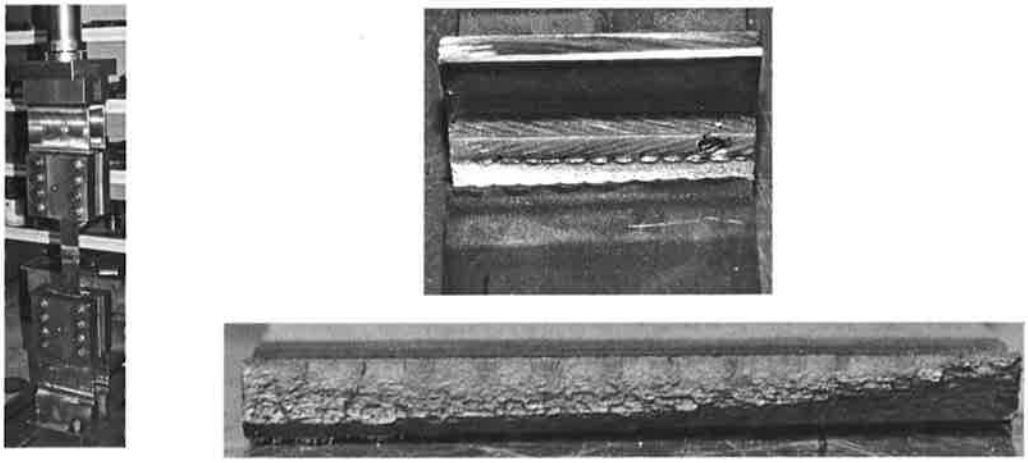
4. KOETULOKSIA

Teorian ja laskentamallin toimivuus pitää varmistaa kokeellisesti. Kokeissa testikappaleena on käytetty kuormaa kantamatonta X-liitosta, josta on määritetty globaalisten muotojen lisäksi hitsin paikallinen muoto makrokuvausta hyödyntäen [9] sekä SLM -menetelmää käyttäen [10]. Kuvassa 9 on esitetty periaatekuvat näistä menetelmistä.



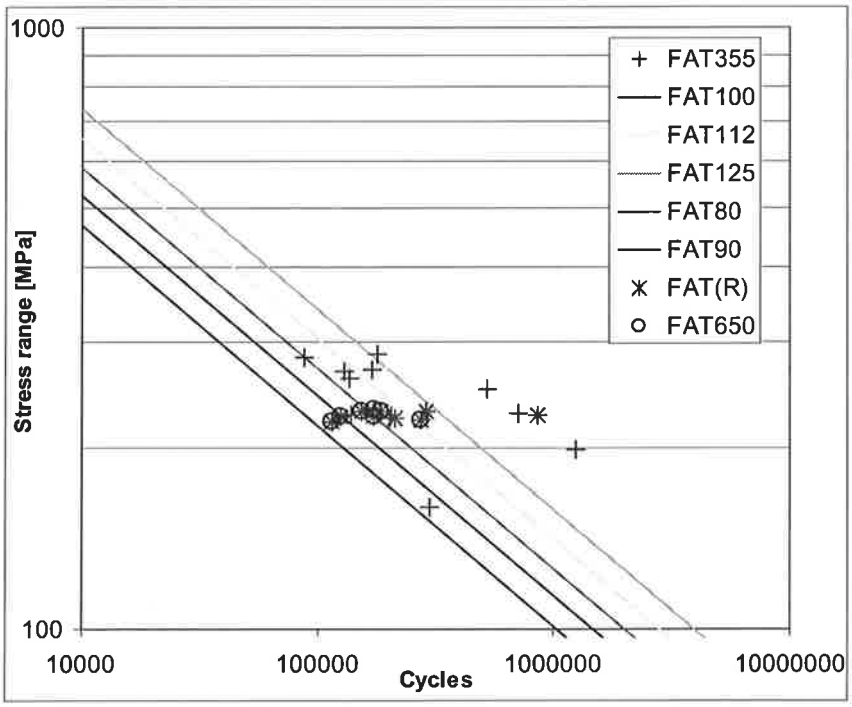
Kuva 9. Hitsin rajaviivan geometria makrokuvausta ja SLM -tekniikkaa soveltaen.

Kuvassa 10 on esitetty väsytyskoe, koekappaleen X-liitos ja sen väsymismurtunut pinta, jossa näkyy vaaputuksen vaikutus särön kasvurintamaan. Kuormitus on ollut vakioamplitudista aksiaalista vetotykytystä joko siten, että jännityssuhde $R = \sigma_{min}/\sigma_{max} = 0.1$ tai siten, että maksimi jännitysarvo on sama kuin materiaalin nimellinen myötöraja (FAT (R)-tulokset). Materiaalina on ollut joko S355 tai S650 ja liitokset on hitsattu robotisoidulla MAG -prosessilla.

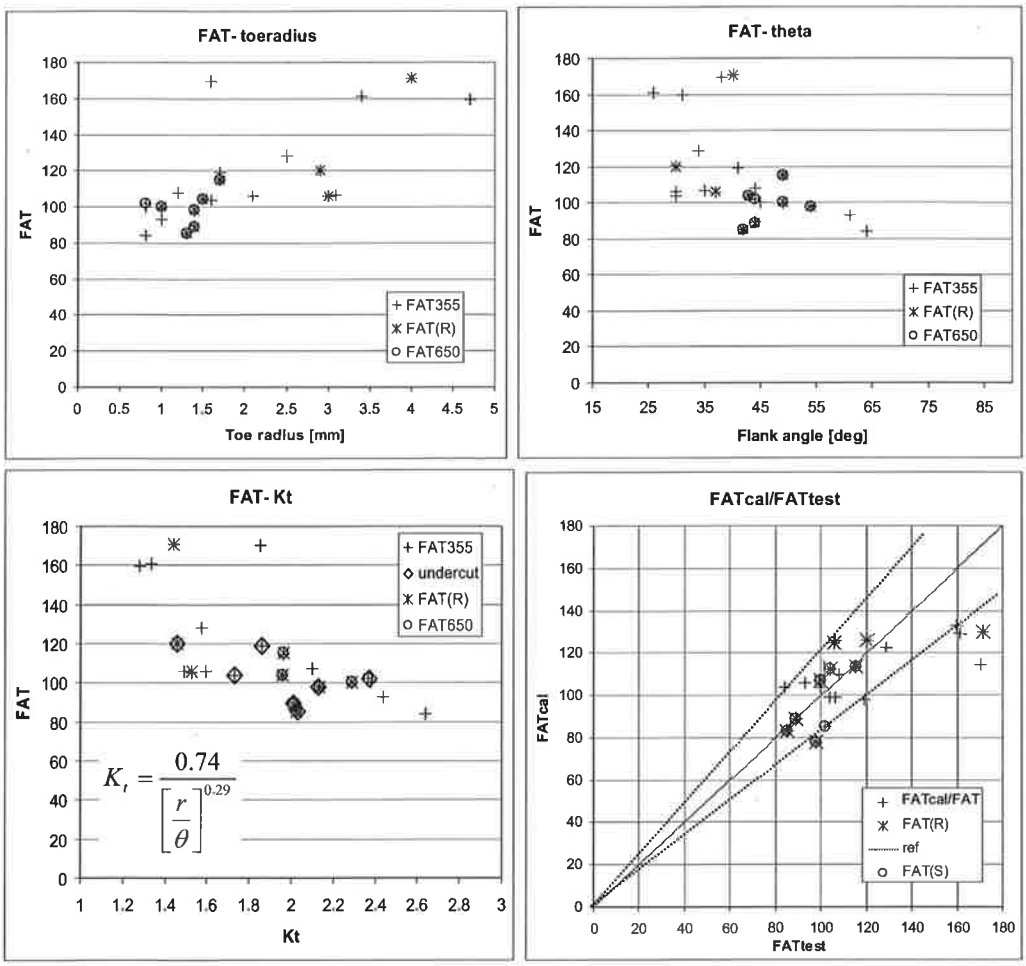


Kuva 10. Väsytyskoejärjestely sekä koekappaleen hitsin rajaviiva ehjänä ja rikkoutuneena

Koetulokset ovat esitetty kuvassa 11. Kuvassa 12 on verrattu mitattujen väsymiskestävyyksien riippuvuutta liitosparametreista (r, θ) ja niiden yhteisvaikutuksesta K_t -kertoimen avulla. K_t -kerroin on määritetty lähteessä [11] esitetyllä tavalla ja edustaa M_k-kerrointa tilanteessa, jossa a_0 = 0. Lisäksi on verrattu lähteessä [8] esitetyllä menetelmällä määritettyjä FAT -arvoja kokeellisesti määritettyihin tuloksiin.



Kuva 11. Väsytyskoetulokset



Kuva 12. Väsytyskoetulosten vertailuja liitosparametreihin ja niillä laskettuihin ennusteisiin

Tuloksista nähdään, että hitsin laatuluokka voi vaihdella FAT_Q = 80...170 MPa. Lisäksi näyttää siltä, että rajaviivapyöristys tai liittymiskulma yksin ei selitä kovinkaan luotettavasti väsymislujuutta. Tosin jonkinlainen trendi ko. yhteyksissä on havaittavissa, kuten myös K_t –kertoimen ja väsymisluokan välisessä yhteydessä. Laskentamalli ennustaa väsymisluokan melko hyvin 20 % virhemarginaalilla FAT 125 MPa asti. Tätä parempaa laatua laskentamalli ei pysty ennustamaan, koska siinä ei ole ydintymisvaihetta mukana.

5. YHTEENVETO JA JOHTOPÄÄTÖKSIÄ

Tarkoituksenmukaisen hitsauslaadun tuottamiseksi tarvitaan yhteisesti hyväksytty menetelmä [12]. Sen perusteella suunnittelu pystyy tulevaisuudessa laatimaan yksiselitteiset ohjeet, joita ammattitaitoinen valmistus noudattaa. Huonon laadun tuottamiseen riittää kumman tahansa laaduntuottotekijän epäonnistuminen. Taulukossa 1 on esitetty yhteenveto kuormituksen vaikutuksesta laatuparametreihin ja taulukossa 2 alustava esitys laatuluokista ja niihin sisältyvistä vaatimuksista rajaviivan ollessa kriittinen.

Taulukko 1. Parametrien vaikutus laatuun

parametrit		kuormitus	
		staattinen	**väsyttävä**
mekaaniset ominaisuudet	sitkeys	ehdottoman tärkeä	tärkeä turvallisuusmarginaali
	lujuus	korvattavissa lisämateriaalilla	ei kovin tärkeä
geometria	hitsi (piena-,V-,K-)	ei oleellista merkitystä	vaikuttaa lovijännitykseen
	hitsin a-mitta	tärkeä	ei vaikuta merkittävästi
	hitsin muoto	kuperuus ei haitta	kovera muoto hyvä
	rajaviivageometria	huono liittymä ei haittaa	tärkeä
viat	ei särömäiset	vähäisinä harmittomia	pintaa avautuvat huonontavat
	särömäiset	nettopintavähennys	vain hyvin pieniä sallitaan
liitostyyppi (juuren puoli ei kriittinen)	voimaliitos	alimitoitus kohtalokasta	hitsin rajaviiva tärkeä
	kiinnitysliitos	ylimitoitusvaara	pituussuunta oleellinen
	varusteluliitos	lämmöntuonti varmistettava	hitsin rajaviiva tärkeä
rajaviivan tilanne	kuormitusaste	a-mitta, varustelu - voimahitsi	laadun merkitys kasvaa
	kiertymä estynyt	ei vaikutusta	laadun merkitys kasvaa
	kolmiaksiaalisuus	parantaa (ei haurasmurtumaa)	ei oleellista vaikutusta

Taulukko 2 Esitys laatuluokkien peruskriteereiksi

kriteerit	**laatuluokat**				
	E	D	C	B	B+
sitkeys	ei vaati-muksia	vaatimus voimassa			
a-mitta (=staat. lujuus)		liitostyypeittäin			
vikatyyppi kriteeri		nettopinta-ala	ei särömäisiä		
menetelmä	manuaalinen tai automatisoitu			automatisoitu	
rajaviivakontrolli		ei tarvita	r, θ, a_0 väsyttävästi kuormitetut voima- ja varusteluhitseille		
tarkastus		esituotannollinen WPS ja silmämääräinen			
väsymislujuus		(FAT 80)	FAT100	FAT112	FAT 125
sovellus		staattinen	väsyttävä vaativa staat.	vaativa väsyttävä	erittäin vaativa väsyttävä
erityisvaatimukset U, T, Y…	x	x	x	x	jälkikäsittely x

Seuraavat johtopäätökset voidaan esittää:

- staattiselle kuormitukselle riittäisi yksinkertaistetut vaatimusohjeet ja väsyttävään kuormitukseen tarvittaisiin nykyistä yksityiskohtaisemmat
- liitostyyppi (voima, kiinnitys, varustelu) vaikuttaa oleellisesti laatuvaatimuksiin
- hitsausparametreilla ja niistä aiheutuvalla rajaviivageometrian vaihtelulla on suuri merkitys väsymisominaisuuksiin väsymisluokan vaihdellessa FAT 80…170 MPa
- rajaviivan laatua ei voi määrittää yksittäisellä parametrilla, vaan tarvitaan niiden yhteisvaikutuksen huomioon ottamista
- LEFM -pohjainen laskentamalli ennustaa melko hyvin väsymisluokan FAT 125 saakka, kun ydintyminen ei vaikuta väsymiskestävyyteen kovin merkittävästi.

LÄHTEET

1. SFS-EN ISO 5817, Hitsaus. Teräksen, nikkelin, titaanin ja niiden seosten sulahitsaus Hitsiluokat, 2003

2. Boverkets handbook om stålkonstruktioner BSK 99, Boverket 2001

3. Hobbacher, A., Fatigue design of welded joints and components. Recommendation of IIW Joint Working Group XIII 1965-03/XV 1127-03, IIW 2005

4. Poutiainen, I., Marquis, G., A fatigue assessment method based on weld stress, International Journal of Fatigue 28 (2006), 1037-1046

5. Lancaster J. F., Metallurgy of Welding. Sixth edition. Abington Publishing , Cambridge 1999

6. Kallio T., Hitsausparametrien vaikutus rakenneteräksen MAG-hitsin latuun. Diplomityö, Lappeenrannan teknillinen yliopisto, Konetekniikan osasto. Lappeenranta 2006

7. Nykänen T., Marquis G., Björk T., Effect of local geometrical variations on the fatigue strength of fillet welded joints. IX Suomen mekaniikkapäivät. Lappeenranta 2006

8. Nykänen T., Li X., Björk T., Marquis G., A parametric fracture mechanics study of welded joints with toe cracks and lack of penetration. Engineering Fracture Mechanics 72 (2005), 1580-1609

9. Seshardi A., Statistical variation of a weld profiles and their influence on fatigue strength. Diplomityö, Lappeenrannan teknillinen yliopisto, Konetekniikan osasto. Lappeenranta 2006

10. Nieminen T., Kosketukseton hitsin pintageometrian määritys. Diplomityö, Lappeenrannan teknillinen yliopisto, Konetekniikan osasto. Lappeenranta 2006

11. Colchen D., Prise en compte de la qualite et des defaults de realisation dans les assemblages bout-a-bout en alliage d'aluminium sollicites en fatigue. Etudes et Recherche, 2004, 27-34

12. Marquis G., Toward a weld quality system for fatigue loaded complex structures. Lecture note, 2006

Lappeenrannan teknillinen yliopisto
Digipaino 2006

ISBN 952-214-227-1
ISSN 1459-2924