Numerical Analysis of Lubrication in an Artificial Hip Joint

Shatish Ramjee

Numerical Analysis of Lubrication in an Artificial Hip Joint

by

Shatish Ramjee

A dissertation submitted in partial fulfilment
of the requirements for the degree

Master of Engineering (Chemical Engineering)

in the

Department of Chemical Engineering

University of Pretoria
Pretoria

11th February 2008

Synopsis

The ageing population has become more active and live longer, these patients require hip replacement surgery at a younger age. Artificial hip implants, consisting of the acetabular cup and femoral head, affect the lives of many people, and the longevity of these implants pose significant concerns (rarely longer than 17 years). To help understand the lubricating performance of such a system, a hip joint model was built based on the Reynolds equation; the model developed simulated hydrodynamic lubrication.

A steady-state angular rotation model was built whereby it was concluded that such motion would not support any load due to the anti-symmetric nature of the resultant pressure distribution (anti-symmetric about the axis of rotation). The pressure distribution from the steady-state rotation simulation contained a pressure source and sink which converged to the centre of the cup and whose pressure value increased in magnitude, as the eccentricity ratio increased. Infeasible results were obtained when the intermediary pressure constraint, allowing only positive pressure values, was implemented. The results obtained were not representative of the problem and it is recommended that this constraint not be implemented.

The transient walking cycle model showed that a fluid with viscosity of $0.0015 Pa.s$ is not sufficient to support a load in the walking cycle under conditions representative of hydrodynamic lubrication. Increasing the fluid viscosity promoted better results in the hydrodynamic model. Increasing the femoral head radius and decreasing the radial clearance between the components also improves the possibility of hydrodynamic lubrication.

It is recommended that the model should be extended to investigate elasto-hydrodynamic lubrication. If possible, the effects of a boundary lubrication model should be investigated, as it is believed to be a major contribution to the lubrication of hip joints.

Keywords: biotribology, hip joint, Reynolds equation, hydrodynamic lubrication, transient model, steady-state model

Acknowledgements

I would like to thank Professor Philip de Vaal for the opportunity to partake in this project, as well as his supervision and guidance to help me through the unforeseen difficulties encountered. I would also like to acknowledge the input of Mr Carl Sandrock who contributed in the problem solving aspects of this project, of both a holistic and programming nature.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

η	Coefficient of friction	-
λ	Lambda ratio	-
μ	Viscosity of fluid	Pa.s
ω	Angular velocity	rad/s
ϕ	Spherical coordinate	
θ	Spherical coordinate	
ξ	Eccentricity Ratio $(\frac{e}{c})$	-
e	Eccentricity	m
F	Resultant force	N
F_f	Frictional force	N
h	Film thickness	m
h_λ	Representative film thickness	mm
IAP	Integral of the absolute pressure	N
ISP	Integral of the square pressure	N^2/m^2
K	Wear coefficient	mm^3/N.mm
L	Overall Length of profile under examination	mm
p	Pressure	Pa

R	Radius of the femoral head component	m
R_a	Composite surface roughness	mm
R_{a-cup}	Surface roughness of the acetabular cup component	mm
R_{a-head}	Surface roughness of the femoral head component	mm
R_{head}	Radius of the femoral head component	mm
R_h	Radius of the femoral head component	m
T	Frictional torque	N.mm
t	Time	s
U	Entraining velocity	m/s
V_w	Volume of material removed	mm^3
W	Normal Load	N
X	Sliding distance	mm
x	Distance from the origin	mm
x_1	Cartesian coordinate (x)	
x_2	Cartesian coordinate (y)	
x_3	Cartesian coordinate (z)	
y	Height of the surface above the mean line	mm
Z	Sommerfeld number	-

CHAPTER 1

Introduction

The hip joint is a complex biological system. It is a synovial joint whose main component is synovial fluid. The synovial fluid provides lubrication between the femur and acetabulum, both of which are covered with hyaline cartilage; it also provides nutrients for the other joint components.

Deterioration of the hip joint has lead to the development of artificial hip joints. The most popular system in use today is the metal-on-UHMWPE system. In recent years, the focus of research has switched to trying to improve the longevity of the implant; because younger patients are requiring total hip replacements. This has lead to increased use of ceramic-on-ceramic, as well as second generation metal-on-metal bearing systems and the use of hip-resurfacing techniques. These bearing systems are known to have problems due to the large amount of revision surgeries required. A better knowledge of the biotribological aspects of the artificial hip implant can help in the understanding and ultimate improvement of the prostheses.

The aim of this study is to:

- understand the basic biological aspects of the hip joint

  This will be done by compiling a thorough literature study whereby the function and composition of the different components of the hip joint will be discussed.

- develop a basic understanding of tribology

  A basic understanding will be obtained by investigating the main topics of tribology: lubrication, wear and friction.

- review literature on the biotribology of artificial hip joints

  This will contain information about how the aspects of tribology have been applied to the artificial hip joint system and the results of such investigations.

- review literature on the modelling of an artificial hip joint

  The review will contain aspects of the type of hip joint models developed, their method of calculation and information obtained from such analysis.

- investigate the lubrication of an artificial hip joint by:
  - developing a model
  - performing the necessary numerical simulations
  - analysing the obtained results

CHAPTER 2

Literature

The hip is a complex joint system which carries much of the weight of the body. It has sophisticated interactions between cartilage, bone, synovial fluid and other connective tissue. Due to degeneration of the joint, the use of artificial hip joints has become necessary. With an increasing number of the ageing population requiring total hip replacements at a younger age (caused by both unhealthy and a more active lifestyles), it is necessary to try and improve the design and longevity of the implant.

Due to the sophistication of the hip joint, and the ambition to improve the longevity, it is necessary to understand the principles of operation of the biological hip joint as well as its tribological behaviour.

The purpose of this literature study was to:

- develop an understanding of synovial joints, in particular the hip joint and its components
- develop an understanding of the fundamentals of tribology and learn how it is applied to artificial hip joints
- to review current literature on the biotribology of hip joints, with focus on hip implants
- to review current literature on the modelling of hip joints, bearing attention to artificial hip joints

2.1 Biological Background

2.1.1 Joints

A joint can be described as the point at which two bones make contact. They are designed to allow movement of bones and provide support.

There are 3 basic types of joints:

- Fibrous joints
  These joints are connected by tissue which is mainly composed of collagen fibres, for example, between the tibia and fibula in the leg.

- Cartilaginous joints
  In these joints the bones are connected mainly by cartilage such as between the sternum and manubrium or intervertabral discs

- Synovial joints
  In synovial joints, the articulating surfaces are separated by a cavity and the bearing surfaces are covered by articular cartilage (Sokoloff, 1978: 2). Synovial fluid is present in this cavity and is confined to this space by the synovial membrane (Furey & Burkhardt, 1997). The hip joint is an example of a synovial joint.

The hip joint is the joint between the femur and the acetabulum. The main purpose of the joint is to support the weight of the body. Synovial joints will be studied in further detail in this chapter due to its relevance to hip joints.

Figure 2.1 is a simple schematic of a synovial joint. The two bones that make up the synovial joint, are covered by cartilage, referred to as articular cartilage. The synovial membrane covers the non-cartilaginous surfaces of the joint capsule or cavity. The synovial membrane secretes synovial fluid into the joint capsule. The synovial fluid provides nutrients for the articular cartilage; it also acts as a lubricant. Synovial joints allow for articulation of the surfaces with a minimal amount of friction or wear (Furey & Burkhardt, 1997). The joint is made up of several complex components; some of these will be discussed in the subsequent sections.

The function of the synovial fluid is affected by the various components of the surrounding extracellular matrices; these molecules being collagen, proteoglycans and glycoproteins (Sokoloff, 1978: 407). The synovial joint is a complex system, and to function properly, all the sub-systems need to integrate efficiently.

There are 6 types of synovial joints, which can be distinguished by their function and range of movement:

- Ball and socket joint
  In these types of joints, the head of one bone fits into the cup-like socket of another

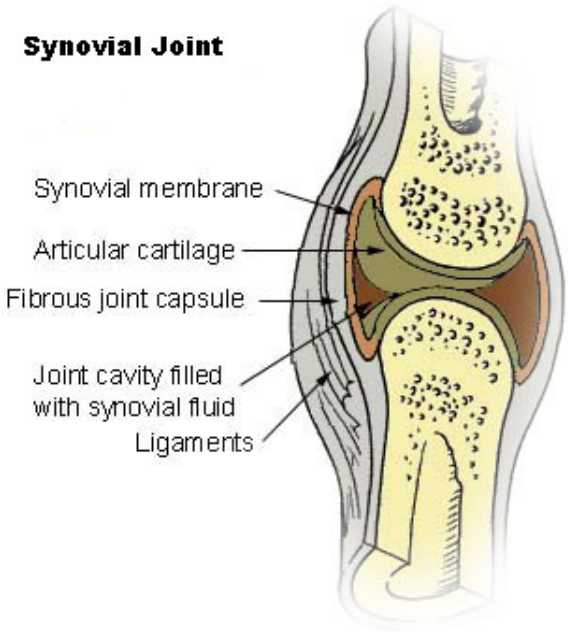

Figure 2.1: A representation of a synovial joint (Wikipedia, 2006b)

bone. This allows for freedom of movement in all directions, from the single common centre between the bones. Examples of ball and socket joints are the hip and shoulder. Figure 2.2 is a representation of a ball and socket joint.

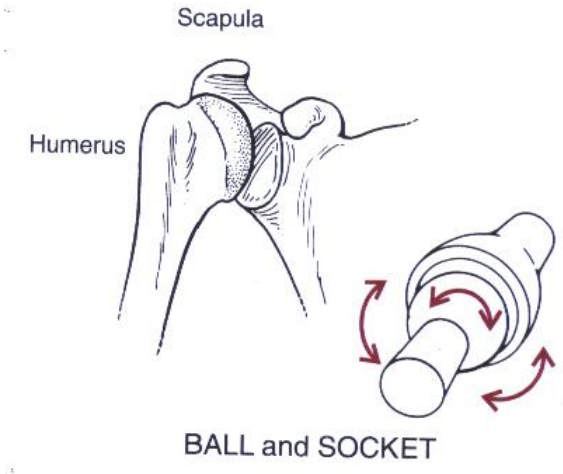

Figure 2.2: Ball and socket joint (Humanoid Animation Working Group, 2006)

- Condyloid joint
  This joint is similar to a ball and socket joint; the head of one bone fits into an oddly shaped depression of the other bone. This joint allows movement in different planes, but not rotational movement in those planes. An ellipsoidal joint is a form of a condyloid joint (can be seen in figure 2.6). Examples of condyloid joints are the wrist (ellipsoidal) and the femoro-tibial joint. Figure 2.3 is a representation of

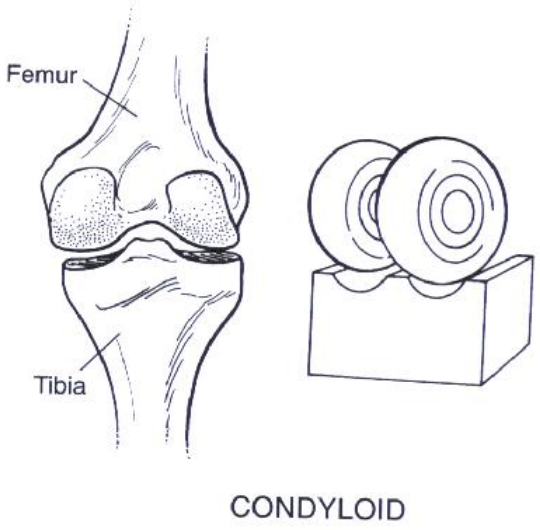

Figure 2.3: Condyloid joint (Humanoid Animation Working Group, 2006)

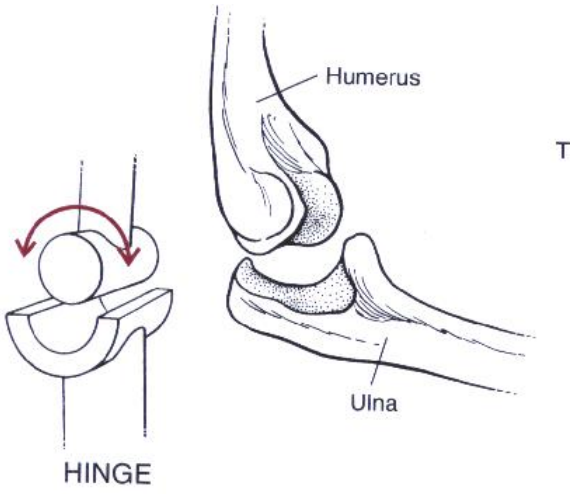

Figure 2.4: Hinge joint (Humanoid Animation Working Group, 2006)

a condyloid joint.

- Hinge joint
  These joints occur when the convex surface of one bone fits into the concave surface of the other. It allows for a flexion and extension movement in one plane only. Examples of these types of joints are the elbow (between the humerus and ulna) and phalanges. Figure 2.4 is a representation of a hinge joint.

- Pivot joint
  In this joint, the end of one bone rotates around the axis of another bone. Generally, the cylindrical surface of one bone rotates within the ring of the other. Examples of these types of joints are between the atlas and axis in the neck (the 'no' head motion) and between the radius and the ulna in the forearm. Figure 2.5 is a representation

Figure 2.5: Pivot joint (Humanoid Animation Working Group, 2006)

of a pivot joint.

- Saddle joint
  In this joint, both bones have concave and convex regions. The opposing surfaces complement each other. The movements of a saddle joint are similar to that of a condyloid joint; it allows movement in two planes, but not rotation. An example of this type of joint is the thumb. Figure 2.6 has a representation of a saddle joint.

- Plane or Gliding joint
  This joint allows for sliding or gliding movement of flat or nearly curved bones as one bone moves over the other. It allows for both back and forth and twisting movements. The movement of such joints is limited, in many cases by the the surrounding ligaments. Examples of these types of joints are those between the carpal bones; it also occurs between vertebrae and between the tarsal bones. Figure 2.6 is a representation of a plane joint.

2.1.2 Cartilage

Cartilage is a form of connective tissue. It provides a foundation for the deposition of bone; it also functions as a smooth articulating surface for the bone. Cartilage is made of chondrocytes, which are part of a gel-like substance called the matrix. Chondrocytes are the only cells found in cartilage and are responsible for the maintenance of the matrix;

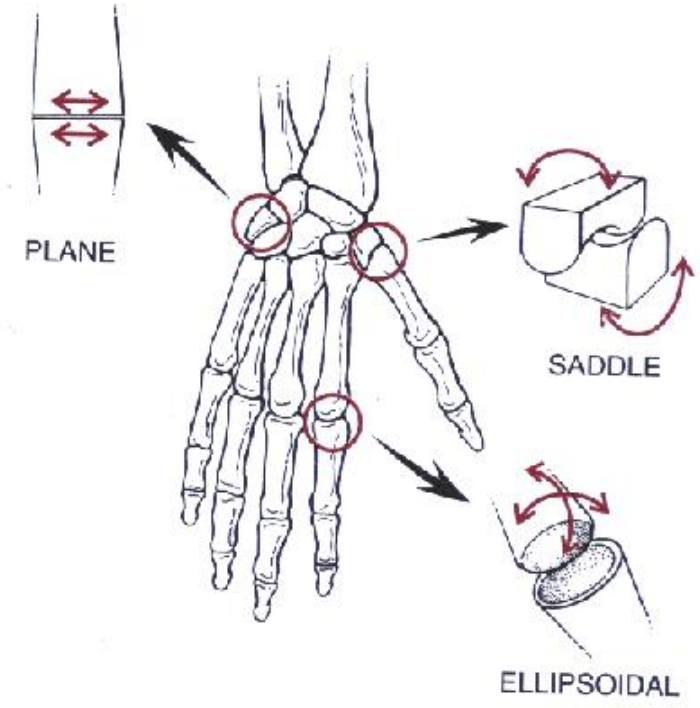

Figure 2.6: Saddle, Plane and Ellipsoidal joint (Humanoid Animation Working Group, 2006)

in the case of articular cartilage, they are responsible for the production of collagen and proteoglycans. Collagen is a long, fibrous, structural protein that provides support for most tissues; proteoglycans are proteins which exist between cells which form large complexes. When cartilage is exposed to stress or strains, fluid is forced in and out of the body; the fluid takes both nutrient and wastes with it (Nigg & Herzog, 1995: 86). Cartilage contains no blood vessels, and as a result, when transfer of nutrients needs to occur (to the chondrocytes), they diffuse through the matrix (Furey & Burkhardt, 1997). In synovial joints this would be synovial fluid.

There are 3 main types of cartilage:

- Hyaline Cartilage
  Is a translucent matrix that is found lining the bones of joints. Hyaline cartilage is also a point from which ossification, bone growth, can occur.

- Elastic Cartilage
  Is the type of cartilage that keeps tubes open. It is similar to hyaline cartilage but contains more elastin which allows it to be stiff yet elastic.

- Fibrocartilage
  It is found in areas requiring high tensile strength. The strength is as a result of the high collagen levels in the cartilage. It can be found between intervertebral discs and in places connecting ligaments or tendons to bone.

Hyaline cartilage is found on the articular surfaces of bones in synovial joints. It is an inhomogenous matrix that contains many fibres, fibrils and filaments (Sokoloff, 1978: 106); but consists mainly of water (Furey & Burkhardt, 1997). It is in essence a thin connective tissue that (Nigg & Herzog, 1995: 83) :

- transfers forces between articulating bones
- distributes forces in joints
- allows joint movement with minimal friction

Depending on the distribution and morphology of the fibres and chondrocytes, the articular cartilage can be divided into four zones and can be seen in figure 2.7 (Sokoloff, 1978: 106). These four zones are (Nigg & Herzog, 1995: 84) (Sokoloff, 1978: 106):

- Zone 1 - the superficial (or tangential) zone
  This is the thinnest zone and is the outermost zone of the articular cartilage. It can be further subdivided into a surface layer and a deeper layer. The surface layer consists of flat bundles of collagen fibrils (Nigg & Herzog, 1995: 84). The deeper layer consists of dense collagen fibres that lie parallel to the surface of the joint. As a result of the orientation of the high strength collagen fibres, both the strength and stiffness of the superficial zone are in the direction of joint movement (tangential to the surface). Therefore this zone is built to resist the force of shear stresses. This zone also contains lots of water (Nigg & Herzog, 1995: 85).

- Zone 2 - the transition (or intermediate) zone
  This zone consists of collagen fibrils that have a larger diameter than those in the superficial zone. The fibrils lie in a plane somewhat parallel to the surface, but not completely. The chondrocytes in this zone play a more prominent part than those in the superficial zone, in the synthesis of the matrix (Nigg & Herzog, 1995: 85).

- Zone 3 - the deep (or radial) zone
  This zone consists of collagen bundles, as well as chondrocytes, that lie normal to the surface of the cartilage. The chondrocytes are stacked on top of each other to form a pillar-like structure This zone has a high proteoglycan content and is responsible for much of the protein synthesis. It also has a low water content (Nigg & Herzog, 1995: 86).

- Zone 4 - the calcified zone
  This zone is the transition from articular cartilage to subchondral bone. This zone consists of hydroxyapatite, which is a component of the bone matrix. The articular cartilage is anchored into the subchondral bone by the collagen fibres found in the deep zone. The separation point between the calcified zone and the deep zone is called the tidemark (Nigg & Herzog, 1995: 86).

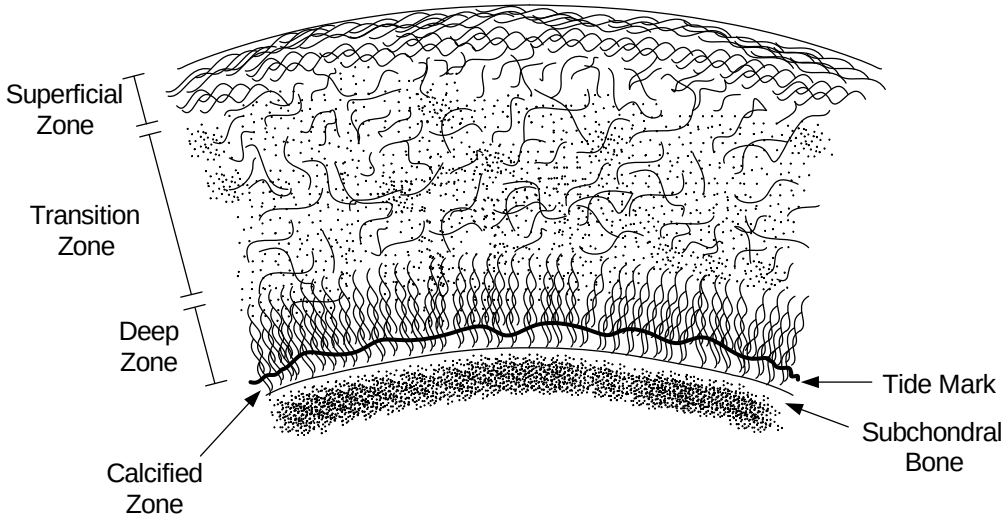

Figure 2.7: The four zones of articular cartilage

The intracellular matrix in cartilage consists of macromolecules and tissue fluid. The macromolecules make up 20-40% of the cartilage, with the rest being made of the tissue fluid (Nigg & Herzog, 1995: 87). These macromolecules are produced by the chondrocytes and are responsible for the mechanical behaviour of the articular cartilage.

The macromolecules consist of collagen, proteoglycan and other proteins (Nigg & Herzog, 1995: 86).

Collagen has a high tensile strength and is responsible for much of the structural framework of the articular cartilage; but it is weak under compression and shear forces. Collagen fibrils do also cross-link to form a macromolecule with higher tensile strength and stiffness. Collagen fibres and fibrils also interact with other macromolecules to give a more rigid and immobile structure (Nigg & Herzog, 1995: 86).

Proteoglycans are found in large quantities in articular cartilage. It is a group of disaccharides linked together to join a protein core which gives it a highly branched structure. The proteoglycans are intertwined with collagen fibres and fibrils to form the structural framework of the articular cartilage. The branched structure adds to the resistance of the framework to compressive forces (Nigg & Herzog, 1995: 87).

The tissue fluid is the main constituent of the articular cartilage; its main component is water. It also contains macromolecules which adds to the viscosity of the fluid. The tissue fluid moves freely through the open spaces in the cartilage structure (Nigg & Herzog, 1995: 89). As the cartilage framework is sponge-like in its nature, the tissue fluid flows in and out of the cartilage. The fluid permeates through the porous cartilage; the permeability is proportional to the pressure applied to the cartilage. The cartilage has a low permeability due to the micro-porous structure and high viscosity of the fluid; the resistance to compression and stiffness of the cartilage also contribute to the low permeability. The low permeability is important as it contributes to the formation of a lubricating layer on the articular surface.

2.1.3 Bone

Bone is a form of hard connective tissue found in the human body. It is a lightweight composite that consists mainly of organic material. The organic material is predominantly collagen, which gives bone its flexibility. The inorganic material is calcium and phosphate salts in the form of hydroxyapatite; this gives the bone its hardness and strength. The main functions of bone are (Nigg & Herzog, 1995: 53):

- to support the body from external forces
- to transfer forces in conjunction with muscles to allow for movement
- to protect vital internal organs
- to form blood cells (hematopoiesis)
- to store calcium and other minerals

Bone structure can be divided into two parts: the compact and cancellous (spongy) parts. The compact bone is referred to as the cortical bone and accounts for most of the bone mass of an adult; it is dense, strong and rigid. Spongy bone is trabecular and has a relatively low density due to its open structure. Trabecular bones are resistive to compressive loads. The trabeculae orient and reorient themselves depending on the direction of the stresses applied to the bone (Martini, 2001: 169).

Bone can be either woven or lamellar. Woven bone is deposited *de novo*, without any hard tissue or cartilage substrate. This gives it the ability to grow during rapid bone growth or repair. It is called woven bone because of the random orientation of the collagen fibres; the disorganised fibres give woven bone poor strength. Lamellar bone grows slowly; it cannot grow spontaneously or *de novo*. Lamellar bone only grows on hyaline cartilage or bone. Collagen fibres of lamellar bone are highly organised and grow in parallel; this gives lamellar bone its strength. Woven bone grows initially, during the growth or repairing process, after which it is covered by lamellar bone (Martini, 2001: 174)(Saladin, 2007: 220).

There are several different types of bones: long, short, flat, irregular and sesamoid (Nigg & Herzog, 1995: 55). Most of these bones have similar structures, but that of the long bone (like the femur) will be discussed.

The ends of the long bone are called epiphyses; these are the articulating ends of the bone over which the hyaline cartilage grows. The shaft between the epiphyses is hollow and is called the diaphysis. The diaphysis is bone that surrounds the medullary cavity (there is no medullary cavity in flat bones). The medullary cavity is lined by a thin membrane called the endosteum; it functions as a barrier for certain ions. The bone is covered by a fibrous tissue called periosteum; this layer penetrates and is anchored into

the bone. It provides nourishment by supplying blood to the bone (Nigg & Herzog, 1995: 55). Cortical bone forms the outer shell of the diaphysis around the cancellous bone at the end of the joints. Figure 2.8 is a representation of the different components of a long bone.

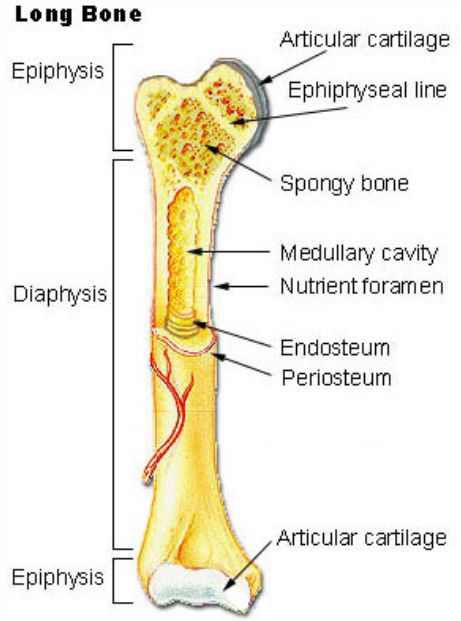

Figure 2.8: Long bone composition (Wikipedia, 2006a)

Bone marrow is found in the medullary cavity of the long bone. Most of the new blood cells are produced here. There is red and yellow marrow. Red marrow produces red and white blood cells and platelets (Saladin, 2007: 220). Yellow marrow contains a higher concentration of fat cells which gives the marrow its yellow appearance. Red marrow is found in flat bones such as the hip bone; and yellow marrow is found in the trabeculae at the proximal end (end pointing towards the skull) of the femur (Martini, 2001: 169)(Saladin, 2007: 220).

2.1.4 Synovial Membrane

The synovial membrane, also referred to as the synovium, is a soft tissue that covers the surfaces, except the cartilage, of synovial joints. The function of the synovial membrane is to:

- produce synovial fluid
- remove synovial fluid and waste from the joint capsule

The synovial membrane can be divided into two parts. The layer adjacent to the joint space, the inner layer, which is called the synovial lining or intima. The outer layer which

is used for support, is referred to as the subsynovial tissue or subintima (Dijkgraaf et al., 1996).

The membrane, and cells of the intima, provide a barrier-like duty during the loading of joints; it prevents the loss of the synovial fluid from the joint, to keep the surfaces lubricated; by doing so, it prevents the joint from becoming dry. There are two types of intima cells, fibroblasts and macrophages. Fibroblasts produce hyaluron; this gluco-protein works in tandem with lubricin, within synovial fluid, to lubricate the joint. Macrophages assist in the maintenance of the joint by removing waste products from the synovial fluid (Sokoloff, 1978: 135).

2.1.5 Synovial Fluid

The main constituent of synovial fluid is water; it can be viewed as a dialysate of blood plasma with hyaluronic acid (Furey & Burkhardt, 1997). The synovial fluid has two main functions in a joint: to provide lubrication, nutrition and as a carrier for waste products. These functions are strongly coupled with the synovial membrane, as it controls both the flow and composition of the synovial fluid (Sokoloff, 1978: 407).

Composition of Synovial Fluid

Electrolytes, proteins and hyaluronic acid are the main constituents of synovial fluid.

Electrolytes themselves are distributed between the blood and synovial fluid. Plasma is believed to be the source of most of the proteins. The concentration of the proteins are dependent on the activity and the location of the joint; the concentration also varies according to the changes, disturbances, the joint experiences. Other factors that contribute to the variation of the concentration is the structure of the synovial membrane and the surrounding matrices (Sokoloff, 1978: 409).

Synovial fluid also contains catabolic enzymes (enzymes that break down molecules into smaller units and release energy). It happens that it is possible for articular cartilage to lose proteoglycans; this reduces the mechanical properties and increases wear of the articular cartilage. The synovial fluid acts as a nutrient source. It also provides a service by removing certain catabolic enzymes (such as collagenase), proteinase inhibitors and waste products (Sokoloff, 1978: 424–245).

1. Hyaluronic Acid

   Hyaluronic acid is a polysaccharide that forms part of the extracellular matrix of connective tissue. It is usually found in small amounts, but in specific tissues, like that of synovial joints, it is a major component (Sokoloff, 1978: 412).

   Hyaluronic acid is a linear polysaccharide of $2-aceamido-2-deoxy-3-O-\beta-D-glucopyranosyluronic\ acid-D-glucose$ linked by $1,4\beta-glucosidic$ linkages.

Thoughts have also led to the possibility of different hyaluronic acid structures in different tissues. In different tissues, the molecular weight of hyaluronic acid is known to vary (Sokoloff, 1978: 414).

2. Lubricating Glycoproteins (LGP)

   Swann and colleagues isolated fractions of synovial fluid using sequential sedimentation techniques and gel permeation chromatography (Furey & Burkhardt, 1997). In trying to isolate the glycoproteins, three main fractions were identified (Sokoloff, 1978: 417):

   - a high molecular weight glycoprotein called LGP-1
   - a reduced form of γ globulin
   - another glycoprotein constituent, LGP-2

   These compounds don't separate easily, they aggregate, and suggest the formation of high molecular weight complexes (Sokoloff, 1978: 420).

   Studies have shown that the boundary lubricating properties of synovial fluid are provided by the high molecular weight glycoprotein components. These are active lubricants for articular cartilage (Sokoloff, 1978: 416).

Lubricating Function of Soft Tissue

1. Lubrication of Soft Tissue

   Lubrication of soft tissue occurs by a boundary lubrication mechanism; the most active lubricant constituent is hyaluronic acid. The lubricating ability of synovial fluid on soft tissue is not necessarily a function of kinematic viscosity, but rather, it is dependent on the concentration of hyaluronic acid and intrinsic viscosity. The intrinsic viscosity is affected by the molecular structure of hyaluronic acid, thus affecting the lubricating ability; therefore when the organisation, structure and length of the hyaluronic acid polysaccharide chains alters, the lubricating ability changes (Sokoloff, 1978: 425–427).

2. Lubrication of Articular Cartilage

   Hyaluronic acid does not lubricate the articular cartilage as it does soft tissues; this is the work done by the protein constituents. These proteins are boundary lubricants that interact with the surface of the articular cartilage. LGP-1 is believed to be the active lubricant of the synovial fluid; however, the effectiveness is dependent on the presence of other components in the fluid (Furey & Burkhardt, 1997) (Sokoloff, 1978: 428–430). LGP-1 is effective in reducing friction but not effective in reducing cartilage wear (Furey & Burkhardt, 1997).

2.2 Tribological Background

Tribology can be described as the science of interacting surfaces in relative motion (Hutchings, 1992) (Bhushan, 2002: XV). Tribology involves the study of friction, wear and lubrication as well as the design of bearings (Jin et al., 2006) (Hall et al., 2001).

Tribology is an important aspect in almost all engineering disciplines; this is evident from most projects involving moving parts. This is prominent in solid and fluid mechanics, but to almost all engineering subjects (Jin et al., 2006).

2.2.1 Surfaces

When surfaces are analysed on a significantly fine scale, all surfaces are found to have some sort of roughness (Hutchings, 1992: 4). The solid surface properties are especially important with respect to solid-liquid and solid-gas interactions. The surface properties affect the area of contact, friction, lubrication and wear (Bhushan, 2002: 9).

One of the most important surface properties is roughness. The average roughness R_a can be calculated using equation 2.1 (Hutchings, 1992: 9).

$$R_a = \frac{1}{L}\int_0^L |y(x)|\,dx \tag{2.1}$$

There are many methods of measuring surface roughness, but for most engineering applications mechanical or optical methods are generally used. As the need arises to measure ever finer surfaces for engineering purposes, low-energy-electron diffraction, molecular beam methods, field-emission and field-ion microscopy, and atomic force microscopy, are gaining more popularity (Bhushan, 2002: 55).

2.2.2 Surfaces of Artificial Joints

Surface roughness is an important characteristic when determining the wear on very smooth surfaces in the presence of lubricants containing proteins. Minor flaws on the surface of the implant can lead to much higher rates of wear in the case of a polymer. To reduce wear rates ultra-smooth surfaces are used; hard and scratch resistant materials, like ceramics, are also used (Dowson, 1995).

The smoothest surfaces for artificial joints are found on ceramics, with average surface roughness of approximately $0.005\mu m$, while metallic surfaces are in the region of $0.01\mu m$. The arithmetic mean deviation (R_a) is the parameter most often used to quantify surface roughness. Different processing techniques produce components with different surface roughness, table 2.1 shows the surface roughness for engineering processes for relevant orthopaedic implants (Jin et al., 2006).

Production process	$R_a(\mu m)$
Sand casting	12.5–25
Sawing	3.2–25
Forging	3.2–12.5
Drilling	1.6–6.3
Turning	0.4–6.3
Die casting	0.8–1.6
Grinding (coarse)	0.4–1.6
Grinding (fine)	0.1–0.4
Polishing	0.05–0.4
Super-polishing	0.025–0.2
Superfinishing	0.005–0.01

Table 2.1: Surface roughness values for different engineering processes

For artificial hip joints, the composite surface roughness is calculate by using equation 2.2.

$$R_a = \sqrt{(R_{a-head})^2 + (R_{a-cup})^2} \tag{2.2}$$

R_a values for different bearing surfaces used in hip replacements can be seen in table 2.2 (Jin et al., 2006).

Bearings	Femoral	$R_a(\mu m)$	Acetabular	$R_a(\mu m)$	Composite $R_a(\mu m)$
UHMWPE-on-metal	Cobalt chrome	0.01–0.025	UHMWPE	0.1–2.5	0.1–2.5
Metal-on-metal	Cobalt chrome	0.005–0.025	Cobalt chrome	0.005–0.25	0.0071–0.035
Ceramic-on-ceramic	Alumina	0.005–0.01	Alumina	0.005–0.01	0.0071–0.014

Table 2.2: Surface roughness values for different engineering processes

2.2.3 Friction

Friction is the resistive force that a body experiences when one body moves over another (Hutchings, 1992: 22).

The basic laws of friction are as follows (Hutchings, 1992: 23) (Jin et al., 2006):

1. The frictional force (F_f) is directly proportional to the normal load (W).
2. The friction force is independent of the contact area.
3. The kinetic frictional force is independent of the sliding velocity,

The first two laws are attributed to Amontons, who rediscovered the work of Da Vinci (Hutchings, 1992: 23). This can be stated simply by equation 2.3.

$$F_f = \eta W \tag{2.3}$$

However Amontons laws of friction are not directly applicable in the case between polymers, and polymers and metals, because the contact in these cases is largely elastic. The visco-elastic behaviour of polymers, mostly shear-thinning, attributes to the inapplicability (Hutchings, 1992: 24).

The third law, attributed to Coulomb, merely states that the coefficient of dynamic friction is less than the coefficient of static friction, and is independent of sliding velocity in most cases (Hutchings, 1992: 24).

2.2.4 Friction of Artificial Joints

Friction between surfaces plays an important role in artificial hip joints. The first artificial hip joints were designed to reduce friction, but failed due to excessive wear. Failure between metal-on-metal implants were caused by high friction and frictional torque generated (Jin et al., 2006).

Not much research has been done on friction of artificial joints, as it is believed that frictional resistance is not a major contributing factor to the loosening process of modern prostheses; this is due to the increasing effort to overcome loosening and thus improve the longevity of the prostheses (Hall et al., 2001).

The hip joint in the human body is a rotational system, therefore it is more meaningful to describe frictional resistance by torque, T, rather than force. This is shown in equation 2.4 (Hall et al., 2001).

$$T = F_f R_{head} = \eta W R_{head} \tag{2.4}$$

Frictional torque is the affect the frictional force, F_f, has about the point of rotation. From equation 2.4, it can be deduced that femoral components with smaller diameter heads, will cause less frictional torque than a head with larger diameter in an acetabular cup of equal outer radius (Hall et al., 2001).

Friction testing has been used to compare the designs of different implants. The friction factor is used to compare the effect of design variables on operating conditions. Calculation of the friction factor (f) is shown in equation 2.5 (Jin et al., 2006).

$$\eta = \frac{T}{R_{head} W} \tag{2.5}$$

Some of the parameters under these conditions can be used to form the dimensionless Sommerfeld number, Z, as shown in equation 2.6 (Jin et al., 2006).

$$Z = \frac{\mu U R_h}{W} \tag{2.6}$$

The change in friction factor with Sommerfeld number can be used to identify different lubrication regimes, in what is called the Stribeck Plot (Jin et al., 2006); a qualitative example of a Stribeck Plot can be seen in figure 2.9.

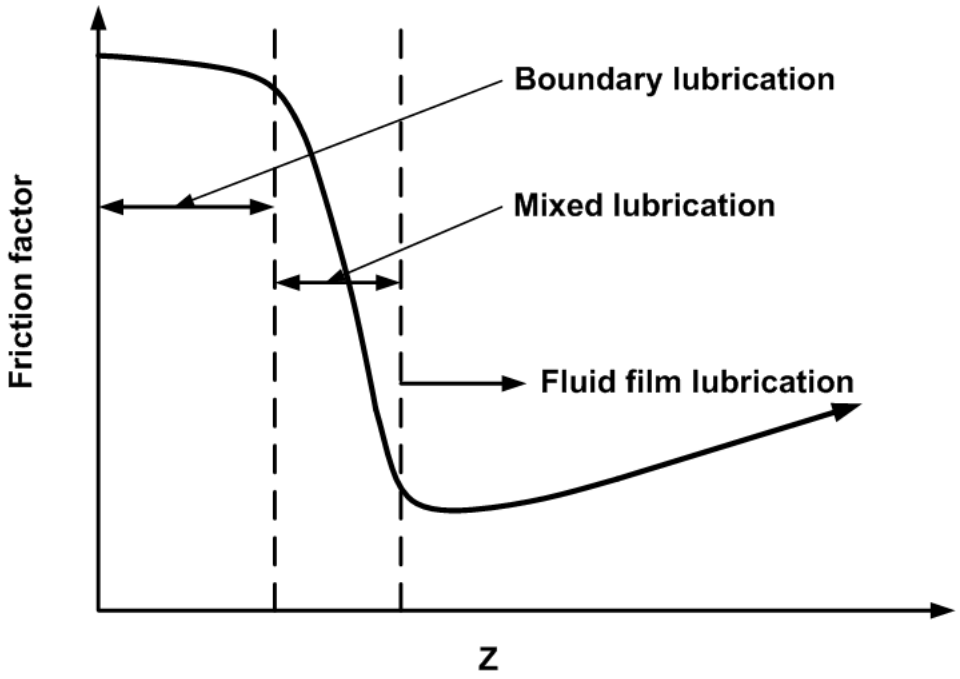

Figure 2.9: Stribeck plot

2.2.5 Lubrication

The performance of a lubricant in a bearing system, depends on, amongst other things (Hall et al., 2001):

- the viscosity of the lubricant
- the geometry of the bearing surfaces
- the entraining velocity at which the bearing operates
- the material properties of the surface
- the surface roughness

There are generally three accepted lubrication regimes:

- boundary lubrication
- mixed lubrication
- fluid-film lubrication

Fluid-Film Lubrication

Fluid-film lubrication can be divided into hydrodynamic and elasto-hydrodynamic lubrication.

1. Hydrodynamic Lubrication

   In hydrodynamic lubrication, there is always a thick film of fluid lubricant between the sliding surfaces. The lubrication occurs because the load is supported by the pressure within the fluid, in a hydrodynamic manner. The pressure that supports the load is generated from the viscous forces of the fluid; and therefore depends on the shear rate (forces) the fluid experiences (Hutchings, 1992: 62–63).

2. Elasto-Hydrodynamic Lubrication

   Elasto-hydrodynamic lubrication results when elastic deformation occurs between solids involved in the lubrication process. This usually happens with solids that have low geometrical conformity. In this case, the manner in which the viscosity of the lubricant changes with pressure is important. This results from the film produced, thinner than hydrodynamic lubrication, because of the heavy loads on the system (Bhushan, 2002: 427). The heavy loads lead to the elastic deformation of the solids, thereby changing the geometry and affecting viscosity (Hutchings, 1992: 67–68).

Boundary Lubrication

Boundary lubrication occurs at very high loading or slow speeds. In this case, the solid surfaces are so close that solid contact between the asperities occur. There is interaction on a molecular level between the lubricant and the solid (Hutchings, 1992: 70–71). The lubricant molecules are adsorbed onto the surface; they produce forces to try and repel the surfaces. These forces try to withstand the pressure. These films are mono- or multi-molecular films that lubricate when solid asperities dominate the contact (Bhushan, 2002: 428).

Mixed Lubrication

Mixed lubrication is the transition between the hydrodynamic\elasto-hydrodynamic lubrication and boundary lubrication regimes. In this regime, there is some solid contact, but some of the load is still being carried by the hydrodynamic film (Bhushan, 2002: 427) (Hall et al., 2001).

2.2.6 Lubrication of Artificial Joints

In the case of artificial joints, the lubrication is dependent on lubricant viscosity, sliding speed, radius of the femoral head, clearance and surface roughness.

Theoretical means can be used to analyse the possible lubrication regime; this can be done by obtaining a ratio of the film thickness (h_λ) to the composite surface roughness of the head and cup, this is known as the lambda ratio (see equation 2.7).

$$\lambda = \frac{h_\lambda}{R_a} \tag{2.7}$$

The different lambda values are associated with different lubrication regimes; these are shown below:

$$\begin{aligned} \lambda \leq 1 &\quad \text{Boundary Lubrication} \\ 1 < \lambda < 3 &\quad \text{Mixed Lubrication} \\ \lambda > 3 &\quad \text{Fluid-Film Lubrication} \end{aligned}$$

Generally, as λ increases, wear decreases. In most femoral head designs, the larger the head diameter, the larger the value of λ, which then leads to a better lubrication mechanism. Better manufacturing processes for reducing surface roughness will improve the possibility of fluid-film lubrication, and therefore less wear. λ is optimised by also taking into account the diametral clearance.

2.2.7 Wear

Wear can be described as the removal of solid material from rubbing surfaces. This can be caused by chemical or mechanical action. The presence of friction or heat, in most cases, accelerates the wear process.

There are three laws of wear (Jin et al., 2006):

1. Wear volume increases as the normal load increases.
2. Wear increases as the sliding distance increases.
3. Wear of the softer sliding component decreases as its hardness increases.

These laws can be expressed mathematically as follows (Jin et al., 2006):

$$K = \frac{V_w}{WX} \tag{2.8}$$

where V_w is the volume of material removed, K is the dimensionless wear coefficient and X is the sliding distance.

The wear factor is dependent on the condition of the bearing surface, the material used and roughness (Hall et al., 2001).

There are several forms of wear:

- Adhesive Wear
  This occurs when two solid bodies are in relative motion to each other (Jin et al., 2006). The asperity contacts are sheared by sliding, causing the particle to break off one surface and attach to the other surface. This transference is referred to as adhesion (Bhushan, 2002: 332).

- Abrasive Wear
  Abrasive wear causes plastic deformation and removal of material when hard particles or surfaces slide on a softer surface (Jin et al., 2006). In most situations of abrasive wear, scratches are observed in the direction of sliding. Abrasive wear can also be caused by fracture; a material can be pushed beyond its threshold load and brittle fracture can occur, this cracking can cause wear (Bhushan, 2002: 346).

- Fatigue Wear
  Fatigue wear occurs under conditions of repeated cyclic loading and unloading (Jin et al., 2006). This causes the surface to develop surface or subsurface cracks which eventually lead to the removal of particles from the surface; this generally occurs after a critical number of cycles. The amount of material removed by fatigue is not really important; what is more important or relevant is the time, number of cycles, before fatigue failure occurs (Bhushan, 2002: 360–361).

- Erosion
  Erosion, like abrasion, can be caused by plastic deformation or brittle fracture (Bhushan, 2002: 369). The stress is a result of the kinetic energy of the solid particles or liquid droplets that strike the surface of the body (Bhushan, 2002: 370 & 374). The material can therefore be removed from the surface by plastic deformation or by the formation and intersection of cracks. This can also be caused by a fluid which contains solid particulates (Jin et al., 2006).

- Chemical or Corrosive Wear
  Corrosive wear occurs by means of chemical or electrochemical reaction and rubbing (Jin et al., 2006). The chemical or electrochemical reaction produces a layer of chemical product on the surface. A rubbing action wears the layer away, thereby exposing the material to further chemical attack (Bhushan, 2002: 379).

- Fretting
  Fretting is caused by a low amplitude oscillatory motion which is tangential to the contact motion between the surfaces, like vibration. In fretting, the normal load causes adhesion and the oscillatory motion causes abrasion which result in wear particles (Bhushan, 2002: 383).

2.2.8 Wear of Artificial Joints

Wear can be defined as the continuous loss of material from the operating surface caused by relative motion on the surface. Wear particles from artificial joints can cause adverse tissue reactions, osteolysis and loosening (Jin et al., 2006). The probability of wear can also be reduced in the full-fluid film regime; this is the region within which engineers try to design artificial joints (Hall et al., 2001).

Factors that influence the wear rates of implants are: type of materials, contact stresses, surface hardness and roughness, type of articulation caused by motion, number of cycles, solution particle count and distribution, oxidation of materials and surface abrasion particulates (Buford & Goswami, 2004).

Different types of contact stresses produce different wear rates. There are three predominant kinematic contact stresses in a hip joint: sliding, gliding and rolling. *Sliding* is a motion caused by the relative contact position of the cup that remains stationary. *Gliding* is where the contact position of the cup reciprocates. *Rolling* occurs when the contact position of the cup varies and relative velocities of both components are the same. Studies showed that rolling contact stresses result in higher wear rates than gliding or sliding (Buford & Goswami, 2004).

There are three basic methods of measuring wear of implants:

- pin on disc machines
- pin on plate machines
- joint simulators

Pin on disc machines are disadvantaged since they cannot replicate the pendulum-like movement of joints; joint simulators are now an important tool when testing new materials and designs for replacement hip joints (Dowson, 1995).

There are various types of hip simulators that are used to assess the performance of difference designs and materials. The main output used to judge performance is the linear or volumetric wear done according to ISO standard 14242-1, however in some cases friction can be measured as well. This incorporates a load profile which is a combination of the Paul and Bergmann cycles (Burger, 2005). Most hip joint simulators are capable of applying by both flexion-extension and abduction-adduction motion whilst applying a vertical load.

A disadvantage of the use of the hip joint simulators are the high capital and maintenance costs. The simulators themselves are also largely inaccessible for research, as simulators are mainly situated at research institutions near major manufacturing companies (Burger, 2005).

2.3 Biotribology: Friction, Lubrication and Wear of Artificial Hip Joints

Biotribology refers to all aspects of tribology related to biological systems (Jin et al., 2006) (Hall et al., 2001).

Examples of tribology applied to biological systems are (Jin et al., 2006):

- Wear of dentures
- Friction of skin and garments
- Tribology of contact lenses and ocular tribology
- Wear of replacement heart valves
- Lubrication of the pump in total artificial hearts
- Wear of screws and plates in bone fracture repair
- Tribology of natural synovial joints and artificial replacements

Total hip replacements were initially designed to be low friction and low wear prostheses; this occurring from low penetration of the femoral head into the cup, thus improving longevity. A few decades later these objectives were met; the penetration and wear suggested that the prosthesis would last for about thirty years (Dowson, 1995).

It was believed that the tribological problems had been overcome and that focus of improvement could be switched to loosening. This proved to be a premature thought, as the main cause of loosening was osteolysis; this happens with cellular reactions caused by wear debris. As a result of these outcomes, the focus of study has switched towards the better understanding of wear and the method by which it is produced (Dowson, 1995).

The design of new prostheses has also evolved. As the ageing population become more active and live longer, these patients require surgery at a younger age; therefore more emphasis is put on the longevity of the device (Tipper et al., 2005).

2.3.1 UHMWPE Acetabular Cup Bearing Systems

The use of UHMWPE acetabular cups was as a result of trying to produce low-friction bearing systems. Sir John Charnley originally used PTFE, but this produced too much wear debris; he then moved on to UHMWPE, which gave better wear characteristics at the expense of higher friction (Hall et al., 2001); actually UHMWPE is preferred over all other polymers for this reason. In the mid 1990's, approximately 90% of all hip replacement surgeries consisted of implants which had an UHMWPE cup and a ceramic or metallic head; this, despite the concerns over the effect of UHMWPE wear debris on the surrounding tissue (Fisher & Dowson, 1991) (Bigsby et al., 1998).

Sterilisation and Oxidation

Sterilisation affects the wear rate of UHMWPE (Burger, 2005). Before UHMWPE cups are implanted, they are sterilised. During γ or electron beam sterilisation, radiation-induced free-radicals are formed; it will undergo one of the following free-radical reactions (Hall et al., 2001):

- recombine to form the original polymer
- combine with oxygen (present due to diffusion)
- cross-link

When UHMWPE is oxidised, like during sterilisation, its crystallanity is increased, which results in reduced fatigue strength. This causes a subsequent increase in wear rate and therefore the production of wear debris. The wear debris are small, less than $10\mu m$, and are the most biologically active wear particles which could result in loosening the components (Hall et al., 2001).

γ-sterilisation also affects the hardness of the cup; there are significant differences between the surface and core hardness. Toughness is also reduced, and reduced toughness is more severe as the cup ages (Burger, 2005).

Some free-radicals do however remain after γ-sterilisation, but given that the component is implanted soon after sterilisation, the risk of oxidation is reduced due to the low oxygen concentration in synovial fluid (Hall et al., 2001).

The sterilisation also allows for cross-linking to occur; this improves the wear characteristics of UHMWPE (Hall et al., 2001) (Burger, 2005).

Molecular weight of the UHMWPE also affects the wear rate; as the molecular weight decreases, the wear rate increases. This happens as a result of the oxidative chain scission during sterilisation (Fisher & Dowson, 1991).

Ethylene-oxide sterilisation produces no free radicals, and therefore no cross-linking or oxidation occurs. Providing no method for cross-linking results in relatively poor wear performance of the UHMWPE cup. This is observed by the slightly better wear characteristics of γ-sterilised cups as compared to ethylene-oxide sterilised cups (Burger, 2005).

Like metals, UHMWPE has different manufacturing processes which affect its properties. These processes are: heat processing, polishing, γ-irradiation, carbon reinforced polymer and pressure crystallisation. Most of these processes have their shortcomings, but γ-irradiation has proved to be effective due to the cross-linking and reduced oxidation just discussed (Buford & Goswami, 2004).

Penetration and Creep

When total hip replacements have been implanted, it has been observed that the acetabular cups have been penetrated. In UHMWPE, the penetration of the femoral head into the cup is a combination of both creep and wear (Dowson, 1995).

Creep is a result of the visco-elastic properties of UHMWPE; creep is caused by the time-dependent deformation under constant load. Creep dominates the initial penetration over the first few years\millions of cycles. The deformation, and change of the geometry, is initially caused by the creep; which is associated with the clearance between the cup and the head (Fisher & Dowson, 1991).

Penetration caused by creep can be substantial, but its recovery during unloading can be equally substantial. Researchers have also shown that creep is also greater for steady loads than for time-varying loads; when the applied load is about three times the body weight (Dowson, 1995).

UHMWPE also softens above $40°C$, this leads to increased creep (Burger, 2005). Another contribution to creep is the annealing after manufacture, this indicates that this processing method should be avoided (Burger, 2005). Cups that are γ-sterilised to improve cross-linking, also improve the creep characteristics of UHMWPE; this results in resistance to plastic flow in hot spots on the bearing surfaces (Burger, 2005).

Wear occurs continuously throughout the use of the prosthesis, and is responsible for the long term penetration aspects. The direction at low penetration is superomedial (towards the center), caused by creep; and is superolateral (towards the side) at high penetration caused by wear (Hall et al., 2001). Studies also show that there are lower penetration rates for the UHMWPE acetabular cups from ceramic femoral heads, than there are from metallic heads (Fisher & Dowson, 1991).

Burger (2005) showed that creep is a function of the manufacturing process, material heat treatment, orientation and temperature. All these aspects should therefore be taken into account during the design of UHMWPE acetabular components.

Biological Activity of UHMWPE Wear Debris

Prostheses with UHMWPE cups will fail due to aseptic loosening. The wear debris formed from UHMWPE enter the peri-prosthetic tissue where it is phagocytosed by microphages which release cytokines (Tipper et al., 2005) (Germain et al., 2003). This causes bone resorption which leads to osteolysis and eventually to the loosening of the implant (Fisher & Dowson, 1991) (Germain et al., 2003). This reaction occurs because the UHMWPE wear debris produced are in the biologically active size range of $0.2 - 7\mu m$ (Germain et al., 2003).

As a result of the osteolysis caused by UHMWPE wear debris, the possibility of reverting back to ceramic-on-ceramic and metal-on-metal systems are being investigated.

There is production of less volume but more reactive metallic wear debris, as compared to larger volumes of more inert polymeric wear debris (Dowson, 1995).

Wear

For metal or ceramic heads articulating in UHMWPE sockets, the designs proved to be somewhat unsuccessful because the softer, more elastic, UHMWPE failed due to fatigue as a result of cyclic stresses (Fisher & Dowson, 1991). In hip joints, fatigue wear is associated with the size of the asperities on the polymer surface; but shows itself as subsurface strains (Dowson, 1995). There are three domains for the wear of UHMWPE (Hall et al., 2001):

- Microscopic asperity wear - takes place on the micron level and results in wear debris of similar size.
- Macroscopic polymer asperity wear - when fatigue of whole polymeric asperities occur; these particles are 10–100s of microns in size and are associated with the grain size of the powder used to make the UHMWPE.
- Structural failure - is known as delamination and occurs mainly in non-conforming bearing systems.

The UHMWPE can also suffer from wear due to abrasion. The asperities of the hard femoral head can press into the soft polymer. A shearing mechanism would cause strain to the polymer; this would lead to scratches or grooves in the surface. Some of the material abraded, would detach from the surface to form wear debris; the debris would be in the same order of magnitude as the asperities (Dowson, 1995).

The femoral heads also seem to suffer a degree of wear. Stainless-steel and cobalt-chrome heads have been damaged due to third body abrasion caused by bone cement. Titanium heads also suffer considerable wear, even in the presence of a scratch resistant oxide layer; the oxide particles could be a source of third body abrasive wear (Hall et al., 2001).

Increased wear can also be caused by third particle abrasion from wear debris from abrasion, metal or ceramic particles, corrosion products or even bone fragments (Dowson, 1995). The ceramic femoral heads are more resistant to abrasion than the metallic heads (Fisher & Dowson, 1991).

Comparison of Metal and Ceramic Femoral Heads

The femoral head components made of ceramic behave differently to there metallic counterparts. Both alumina and zirconia have better wear properties than metal, this can be attributed to (Hall et al., 2001):

- ultra smooth surface finish
- increased scratch resistance
- greater wettability

Implants have shown that penetration of ceramic femoral heads can be up to half that observed with metal heads with similar topographic characteristics (Dowson, 1995); if penetration does occur on the surface; it does not leave pile-up similar to those on the sides of scratches of metal components (Hall et al., 2001). A unique characteristic of ceramic over metal alloys, is that it does not form an oxidative layer, and therefore cannot suffer from oxidative or corrosive wear (Buford & Goswami, 2004).

Scratch damage on metal femoral heads is known to increase the wear rate of the UHMWPE. Ceramic femoral heads are much more scratch resistant and results in reduced wear of UHMWPE, and therefore also osteolysis (Tipper et al., 2001).

Ceramic heads also show lower levels of joint diseases, but its use has been limited due to the expense of the device (Hall et al., 2001).

During the early use of ceramic femoral head components, it was susceptible to fracture, despite its good wear resistance; by changing the manufacturing process, the resistance to fracture was increased and thereby reduced failure due to fracture (Buford & Goswami, 2004).

Design

The main factors that influence the wear of the UHMWPE cups are the material and roughness of the surface of the femoral head; for this reason it needs to be accounted for in the design process. The femoral head is designed from a hard material and also to have a highly polished surface to try and reduce wear of the softer polymer (Fisher & Dowson, 1991). If a smooth femoral head is used, the wear rate of the UHMWPE increases as the surface roughness of the femoral head increases. Single asperities on the surface also increase the wear rate. Not only is the surface roughness a major factor of the topography, but so is the wavelength of the asperities(Fisher & Dowson, 1991).

The different levels of stress, speed and geometry of the implants, affect the friction of the UHMWPE and metal/ceramic pairs, which in turn affect the lubrication of the joints. The high speeds cause frictional heating which also increases the wear rate. The asperities of UHMWPE are larger in comparison to its respective femoral head component; this results in a higher surface roughness. As the surface becomes more rough, the film thickness reduces; therefore the joint generally operates in a boundary or mixed lubrication regime (Fisher & Dowson, 1991) (Bigsby et al., 1998).

There is also a relationship between the contact pressure and wear of UHMWPE; the wear increases as the contact stress approaches the limiting compressive stress of

UHMWPE. This factor is important in the design of femoral heads (Fisher & Dowson, 1991).

Other important design aspects are the femoral head radius and radial clearance of the bearing system. These aspects have been addressed in papers discussing the modelling of joints as the aim is generally to help improve the design of artificial hip joints. These aspects are discussed in section 2.4.2, but it has been concluded that larger head diameters and smaller radial clearances are best suited to improve lubrication and therefore better performance of the bearing system.

2.3.2 Ceramic-on-Ceramic Bearing Systems

Alumina-on-alumina systems have been used since 1970 and has a very low coefficient of friction (Fisher & Dowson, 1991). Wear is at least an order of magnitude less than metal-on-UHMWPE systems (Germain et al., 2003); wear particles are also less biologically active (Hall et al., 2001). The alumina pair bearing systems have not been widely used because of the difficulties encountered during the manufacturing of components with such smooth surfaces, and their associated cost (Fisher & Dowson, 1991). A new method of processing improved alumina is by hot isotatic pressing; it is commonly referred to as HIPed alumina. It is believed that HIPed alumina has better wear characteristics than non-HIPed alumina; Nevelos et al. (2001) showed that HIPed alumina has lower wear rates than non-HIPed alumina, but with no statistical significance.

The resistance to fracture of ceramic components is dependent on its mechanical strength; this is affected by the manufacturing process. As the quality of medical grade alumina improves, so does its mechanical strength and resistance to fracture (Hall et al., 2001).

Performance of ceramic-on-ceramic systems is highly dependent on the surgery. Impingement caused by incorrect positioning has resulted in high wear rates and osteolysis; it has also been associated with the "'stripe"' and increased wear debris formed from Mittelmeier designed prostheses. However, if properly positioned, it produces small amounts of wear debris (Germain et al., 2003).

The "'stripe"' phenomenon also occurred in first generation ceramic implants; it results in increased wear and was believed to be caused by steep acetabular cup positioning or even repeated dislocations (Nevelos et al., 2001). Studies have shown that steep acetabular cup position does not cause the wear "'stripe"' on the femoral head as was first believed (Tipper et al., 2002). It was shown that there is separation of the ball and socket during the swing phase of metal-on-UHMWPE joints; this was termed micro-separation and occurs when a negative load is applied during the swing phase. Tests were done by Nevelos et al. (2000) on a joint simulator using the micro-separation environment to simulate joint laxity that occurs after surgery. This was found to increase wear rates,

especially on ceramic-on-ceramic systems; it also showed a wear "'stripe"' on the femoral head, which is found on most in-vivo samples; this was accompanied by damage on the rim of the acetabular cup (Tipper et al., 2005).

The wear debris from ceramic-on-ceramic systems have a bimodal distribution. Particles in the nm range are caused by relief polishing; whereas particles in the μm range are obtained as a result of contact between the the acetabular cup rim and femoral head; this was believed to be the action that happened when the heel strikes the ground. This is the cause of the wear "'stripe"' and source for larger wear debris, caused by the granular fracture of the ceramic (Tipper et al., 2002). The wear "'stripe"' is believed to be caused by the heel strike on the ground; however investigations by Walter et al. (2004) suggest it is more likely that it is caused by edge-loading (of the rim of the cup) when the hip is flexed; this occurs in the movements of rising from a chair or climbing a high step.

Osteolysis is hardly observed in ceramic-on-ceramic prostheses, in the cases where it does occur, it is associated with extremely large wear volumes. The peri-prosthetic tissue shows no obvious pathology; there are areas rich in macrophages and there are parts of tissue that have necrosis (Hatton et al., 2002). The mixed pathology could be caused by the bimodal distribution of the wear debris. Particles within the μm range can activate macrophages to produce mediators of osteolysis, (Tipper et al., 2001); but at high concentrations of the nm particles, tissue necrosis is possible (Germain et al., 2003) (Hatton et al., 2002).

2.3.3 Metal-on-Metal Bearing Systems

The first metal on metal total hip replacement was performed in 1938, with poor results. The first cemented total hip replacement was performed in 1960, which was successful and widely used; but the long term results were poor with many patients requiring revision surgery due to acetabular cup loosening (Hall et al., 2001).

Early prosthesis could attribute some of the problems to poor bearing design and choice of materials. These systems had a small diametral clearance, which increased contact during loading. The increased contact area is also believed to produce more friction that could possibly contribute to failure of the material (Hall et al., 2001). By the mid 1970's, the use of metal-on-metal prostheses was being phased out, in favour of metal-on-UHMWPE systems because of the high loosening rates of the metal bearing systems and concerns over the biological activity of the metallic wear debris (Dumbleton & Manley, 2005).

In the 1980's metal-on-UHMWPE systems were predominantly used; the cause of osteolysis, and subsequent loosening, was attributed to UHMWPE wear debris. This lead to alternative implants, one being the re-introduction of metal-on-metal bearing systems (Dumbleton & Manley, 2005).

Metal particles produced are on average much smaller than UHMWPE particles. Even though metal-on-metal systems produce less wear by volume, they produce more wear particles; 1–10 million particles per step (Tipper et al., 2005). Despite the large number of particles, osteolysis is hardly observed in metal-on-metal systems (Hall et al., 2001).

Cobalt-chromium pairs have been used in hip prostheses; they have low wear (due to hard surfaces), but they have high friction. This is a concern as the friction is believed to contribute to the loosening process (Fisher & Dowson, 1991). Cobalt-chromium is used in favour of titanium alloys as it has shown better wear resistance (Buford & Goswami, 2004).

Wimmer et al. (2003) also showed that a tribochemical reaction takes place on the oxide layer of metal prostheses. This results in a predominantly organic protein layer with salts. This layer functions as a solid lubricant and covers much of the surface area, by doing so it prevents much adhesion and reduces surface fatigue (Wimmer et al., 2001).

First Generation Metal-on-Metal Implants

There were several designs of these first generation prostheses, namely: McKee-Farrar, Ring, Stanmoore, Meuller-Huggler and Sivash. The designs incorporated both cemented and uncemented components; and all of these implants also used cobalt-chromium alloys. Unfortunately all of the designs had pitfalls as they were generally of a poor quality and design; they resulted in high frictional torque and equatorial contact which contributed to the loosening of the implant (Germain et al., 2003). The prostheses worked well despite their flaws, requiring revision surgery at an average of 13-14 years after the initial hip replacement; some lasted well over 20 years (Dumbleton & Manley, 2005).

These first generation designs, showed relatively low wear rates. They produced small wear debris, in the range from $0.25 - 2\mu m$, and as a result produced many wear particles. The tissue around the prosthesis showed sign of infection; a large number of metal particles were found in the tissue surrounded by macrophages. This caused staining of the tissue and osteolysis. This could have been caused by the impingment of the loose components (Dumbleton & Manley, 2005).

Second Generation Metal-on-Metal Implants

Second generation implants were developed in the late 1980's, became more widespread in the 1990's, and was an improvement on the first generation implants due to better bearing design. These were the Metasul designs with cobalt-chromium-molybdenum alloy femoral and acetabular components (Tipper et al., 2005). The sphericity, clearance and surface finish were improved for better lubrication. Most of the designs used cobalt-chromium allows with varying carbon content and were either wrought or cast. Some of the implants were unsuccessful, with staining of the tissue around the prosthesis, caused

by impingement (Dumbleton & Manley, 2005).

The metal components used in these designs can have a low or high carbon content. The high carbon content alloys have a biphasic structure; it has small grains surrounded by scratch-resistant carbides. Hardness also improves with higher carbon content (Tipper et al., 2005). Tipper et al. (2005) also suggested for these reasons, that the femoral and acetabular components both be made of high carbon content alloys as they would produce less wear debris. The cobalt-chromium alloys can also be cast or wrought. Cast alloys produced less wear debris than alloys that have been wrought (Buford & Goswami, 2004). This happens since the wrought components have carbides that wear below the matrix surface during use. This results in micropitting and possible delamination. Many of the carbides of the cast component remain on the surface, preventing pitting; this allows for a much stronger surface, thereby possibly reducing wear (Wang et al., 1999).

Second generation hip implants did not show lower wear rates than the first generation implants. Revision surgery of most patients was a result of dislocation or loosening. The femoral head diameter for the implant was generally small, which caused impingement and dislocation and reduced the probability of full-fluid lubrication (Dumbleton & Manley, 2005).

Hip resurfacing of these newer implants has also become more prevalent, especially for younger patients, as this pushes back the date for total hip replacement surgery. Advantages of resurfacing are: femoral bone preservation, reduced risk of dislocation and increased range of movement (Tipper et al., 2005); it also leaves the femoral canal intact, which makes it easier if total hip replacement is eventually necessary (Ebied & Journeaux, 2002). The femoral head is also closer to the anatomic head size of the femur. This reduces the possibility of dislocation and increases the range of movement. For hip resurfacing, the larger diameter femoral component can allow for larger diametral clearances, which increases the probability of fluid-film lubrication. New hip resurfacing techniques also use hybrid fixation of the components; the femoral component is cemented and the acetabular cup is uncemented (Ebied & Journeaux, 2002).

Metal-on-metal articulations produce significantly less wear debris than UHMWPE. This is because metal bearing systems don't form debris as fast as UHMWPE cup systems; the metal systems also show reduced acceleration of the production of wear debris, whereas UHMWPE does not (Buford & Goswami, 2004). The wear of hard-on-hard bearing systems, like metal, have 2 distinct phases; The first million cycles or so, is the bedding in wear period; this is the self polishing action of the metal wear particles. The second period is the lower, steady state wear period (Tipper et al., 2005).

The formation of these wear debris can mainly be attributed to mechanical stresses, and not thermal stresses. These stresses result from surface fatigue and abrasion, as well as tribochemical reaction (Wimmer et al., 2001).

Dumbleton & Manley (2005), also suggested increasing the size of the femoral head

component to reduce impingement and increase range of motion as well as probability of full-fluid lubrication. The bedding-in wear volume of large diameter femoral heads are low; but in comparison to the steady-state wear volume it is significant; the bedding in wear volume is a major contributor to the total volume of wear debris, even over long periods of time (Dowson et al., 2004). Metal-on-metal systems are also lubrication sensitive; the diametral clearance directly affects the lubrication and subsequently wear. Systems with smaller diametral clearances show reduced wear. Optimal clearances are however dependent on the diameter of the femoral head (Tipper et al., 2005) (Dowson et al., 2004). This can also explain the better performance of resurfaced femoral heads, as they have a larger diameter than before. Another important design variable is sphericity; as sphericity increase, wear decreases (Ebied & Journeaux, 2002).

There is no evidence thus far to suggest that second generation metal-on-metal implants are better than metal-on-polyethylene implants, with regards to length of survival; but they have showed better survival rates than the first generation metal implants. In fact, Naudie et al. (2004) concluded that metal-on-metal bearing systems showed lower risk of stem and/or cup loosening than metal-on-polyethylene systems; but with no statistical significance. There is also thus far, no evidence to suggest that metal-on-metal systems cause osteolysis (Tipper et al., 2001).

Biological Effect of Metal Wear Debris

The metal levels of patients with metal-on-metal bearing systems were reported to be high; cobalt and chromium levels were higher in both urine and blood serum samples from implant patients (Tipper et al., 2005). The levels decrease when the implants last longer; increased levels have been reported for surface replacements (Dumbleton & Manley, 2005).

High cobalt and/or chromium levels can lead to hypersensitive reactions. Reports have also emanated of cobalt hypersensitivity leading to early loosening of the implant (Tipper et al., 2005).

Metallic wear debris are generally very small, some in the nanometre range, with a large surface area,which leads to improved dissolution of the metal into the blood stream. The particles are small and are below the size range needed for macrophage activation. Metal particles are known to travel to the lymph nodes via the lymphatic system (Tipper et al., 2005). Granulomas have been found in the liver and spleen, which have been caused by these travelling particles (Dumbleton & Manley, 2005). Chromium particles have been observed in tissue around the prosthesis; it showed only mild infection as most of the metal particles move away via the lymphatic system. Corrosion products are also believed to be the cause of bone resorption and may cause osteolysis (Dumbleton & Manley, 2005). Cobalt-chromium wear particles have also been shown to cause necrosis

of the peri-prosthetic tissue; these particles may also be carcinogenic (Germain et al., 2003)

2.3.4 Cushion form Bearings

A method of reducing wear debris is to improve the possibility of the joint working in the full-fluid lubrication regime (Bigsby et al., 1998). Cushion bearings with low elastic modulus, similar in form to articular cartilage, are being investigated as a possible covering for the rigid acetabular cup and femoral head components, to try and obtain this mode of lubrication. Possible materials to be used are polyurethane, silicone rubbers and hydrogels, which would enhance the probability of full-fluid lubrication by one or many proposed lubrication mechanisms (Fisher & Dowson, 1991).

It is believed that the bearing layer would result in lower shearing surfaces which would cause less debris due to reduced fatigue. This configuration will also help to transfer the load from the cushion to the rigid component (Fisher & Dowson, 1991).

Bigsby et al. (1998) had developed a composite cushion bearing made of both high and low modulus polyurethane. The composite was adhered to a metal shell by injection moulding the high modulus polyurethane followed by moulding of the low modulus polyurethane. The results on this composite cushion metal backed acetabular cup showed that volume change after the initial stage of creep is negligible. It also produced almost no wear debris after being able to operate in the full-fluid lubrication regime. Long term tribological effects were promising but problems arose from the degradation at the internal interfaces of the composite.

2.4 Modelling of Synovial Joints

The modelling of synovial joints has gone through stages, and different focuses. It begins with the modelling of the natural synovial joint and cartilage; each of these a tough task to model in their own right. Simple models were developed to describe these systems; and naturally, as time progressed these models became more complex in both the physics and geometry with the progression of both computational power and applied mathematics. With the introduction of arthroplasty, surgical reconstruction or replacement of a malformed or degenerated joint, the modelling of artificial joints has also been thrust into the spotlight. With the need to improve the quality of artificial joints currently implanted, the lubrication of the joints and the effect of implant design variables on the lubrication has been investigated.

2.4.1 Modelling of Articular Cartilage and Natural Joints

The article by Sinha et al. (1982), summarises the history of human joint lubrication until 1982. It begins with MacConaill, who suggested that synovial joints work as hydrodynamically lubricated bearings. Charnely proposed boundary lubrication. McCutchen proposed "'weeping"' lubrication, based on the characteristics of articular cartilage. Dintenfass introduced the concept of elasto-hydrodynamic lubrication. Dowson suggested that more than one form of lubrication mechanism was necessary to describe joint lubrication. He suggested that the main contributing factor was elasto-hydrodynamic lubrication with squeeze film lubrication provided by the cartilage; in the context of severe loading he suggested boundary lubrication. Walker then also proposed a "'boosted"' lubrication mechanism whereby the squeeze film leads to a diffusion of hyaluronic acid and other lubricant constituents between the cartilage surfaces, thereby increasing the viscosity of the fluid.

Torzilli and Mow made an attempt to model articular cartilage in a two part paper. In part I, Torzilli & Mow (1976a) derived equations, based on mixture theory, and applied it to describe fluid motion through articular cartilage during motion.

Cartilage was viewed as both a solid and liquid phase; which is a mixture of two interacting continua. This mixture theory is used to try and develop concepts which take into account fluid flow based on pressure gradients and elastic deformation based on the porous, solid matrix.

By applying Hamilton's principle (requiring variational calculus) and mixture theory based on entropy, an alternative formulation of the differential equations of motion as an equivalent integral equation was obtained. The equations of motion of interacting continua of the porous, elastic solid and incompressible fluid were used to derive the Navier equations of motion of each of the two phases.

The mixture was considered as a linear elastic solid filled with an incompressible fluid which encounters frictional resistance to relative motion. Both the solid and liquid were considered inert and therefore no chemical reaction took place; it is known that reactions do take place because of the metabolic activity of the chondrocytes.

The equation developed was for the flow of the interstitial fluid through the elastic solid matrix. During the derivation, body forces were assumed to be zero and stresses were assumed isotropic in the fluid. The fluid equation of motion was reduced to a form of Darcy's equation of fluid flow through a porous medium; this equation takes into account the solid matrix and pressure terms representing hydraulic pressure.

This paper investigates the results that can be obtained when applying a time varying load as opposed to spatially varying area of loading on the articular cartilage; Spatially varying area of loading represents flexion-extension motion, which is similar to mechanical pumping.

In part II of the paper, Torzilli & Mow (1976b) reduced the equations derived in paper I by applying an order of magnitude analysis; these equations were solved by a double Fourier-Laplace transform procedure to obtain a solution. The model developed was derived for a cyclically loading process.

The authors were able to deduce from their model and results that the flow process is highly dependent on the tissue thickness and surface porosity; the frequency of the loading as well as the permeability, stiffness and Poisson ratio of the matrix also influenced the solution.

Collins (1982) then developed an articular cartilage model to address the fluid interchanging across the deforming articular layers, and the change in hyaluronic acid concentrations in the articular cavity.

Transport of the fluid within the cartilage is described by a modified form of the equations of Torzilli & Mow (1976a). The concentration of hyaluronic acid is modelled by a simple, transient, convection-diffusion equation describing quasi-one dimensional channel flow. The equation takes into account the kinematics when the cartilage is deformed, thereby affecting porosity. With large pores, very little resistance to outflow is modelled. With smaller pore sizes, the gelling effect of the fluid is modelled.

A fluid-mechanical model for the flow of fluid through an elastic, porous matrix deforming under time varying load was analysed. It was concluded that at small gaps between articular surfaces, the maximum concentration of mucin, a glycoprotein of synovial fluid, at the articular surface was obtained, thereby also determining the rate of lubricant gel buildup on the cartilage during contact.

Sinha et al. (1982) then attempted to mathematically model and explain the observed effective increase in viscosity at the solid boundary. As synovial fluid has a complex chemistry and rheology, the authors used a micro-polar fluid model to describe the fluid behaviour while the synovial joint was modelled as a rigid, porous spherical bearing. Slip was considered at the boundary surfaces. Cartilage elasticity was ignored. Equations were derived neglecting body forces, couples and inertial effects as well as other lubrication theory assumptions. It was assumed that flow through the cartilage was Newtonian; therefore the flow could be described by Darcy's law. Sliding kinematics were not considered. The authors concluded that a microcontinuum approach utilising micropolar theory was effective to model hyaluronic acid in synovial fluid. This was verified when an increase in effective viscosity was observed, leading to higher loading capacity.

2.4.2 Modelling of Artificial Hip Joints

The modelling of artificial hip joints has in many cases centred around the extensive contributions of Jin and Dowson, as well as their co-authors. Many other authors have also contributed to this field, some of these authors will be discussed in this study. The

models developed have in most cases tried to show the effect of the different material combinations on the lubrication of the artificial hip joint; in most cases, the material combinations chosen are based on the different designs of current implants.

Most early work on this topic suggested that the lubrication mechanism within the hip joint is in the mixed regime with contributions from both boundary and fluid film lubrication, depending on the operating conditions. Squeeze film lubrication is also believed to be a large contributor to the film thickness and as such Mabuchi & Sasada (1990) decided to attempt the numerical modelling of squeeze film lubrication.

Mabuchi & Sasada (1990) state that the elasto-hydrodynamic lubrication film is generated by both entraining and squeeze actions, and that the squeeze action is dominant during the stance phase as the entraining velocity is too low to produce a wedge film. The model developed, did not contain the contributions of the wedge film. The elasto-hydrodynamic equations were developed by including the deformation of the surface and solving the Poisuelle equation for fluid flow to obtain the pressure needed to balance the force of the load applied to the joint. The geometry used was a spherical ball-in-socket joint in polar co-ordinates; a sinusoidal load was used to simulate the loading cycle of the system. It was assumed that the space between the femoral head and acetabular cup was filled with synovial fluid. The main result from this paper was that the squeeze film effect was negligible under a constant load, even if the material is soft.

In the many papers published by Dowson and later by Jin, with their respective co-authors, the method of lubrication modelled was elasto-hydrodynamic lubrication. One of the main assumptions made in these models was that synovial fluid is an incompressible, isoviscous Newtonian fluid (Jin et al., 1997). It is accepted that synovial fluid is indeed a non-Newtonian, shear-thinning fluid; but that the viscosity changes only at high pressures, up to 100MPa, and at low shear rates, lower than that modelled during the walking cycle (Jin et al., 1997). Therefore for the purpose of lubrication analysis, it is reasonable to assume synovial fluid to be isoviscous.

There are also several geometries used to describe the hip joint model. In recent times, the geometry used has switched from ball-on-plane models, to ball-in-socket models. This change, is to use a more realistic model geometry for the hip joint. As the geometry has evolved, so has the method of lubrication analysis, with respect to the calculation of elastic deformation. At first, constrained column models were used to model the elastic deformation after which finite element methods were used. This shows some of the ways in which the elasto-hydrodynamic model has changed with improved understanding and computational power.

Although much has changed, the one constant in the model has been the use of the Reynolds equation in its different forms, depending on the type of model, steady-state or transient, and the geometry chosen. Convergence of the model was also imposed by relating the hydrodynamic pressure to the load imposed on the joint. The cup has also

been assumed to be horizontal in the geometries used as opposed to almost 45 ° inclination in its anatomic position.

The journey through the papers written by Jin and his co-authors begins with an analysis of lubrication regimes for hard bearing surfaces in artificial hip joints and a comparison to UHMWPE-metal bearing systems by Jin et al. (1997). A ball-on-plane geometry is used for the model and a linear elastic model is used to describe the deformation of the bearing material. Other important information considered in this model were that a steady-state entrainment model was developed under conditions of constant load and speed; this model was used because predicted film thickness remains relatively constant during walking (Jin et al., 1997).

The main results from this paper on elasto-hydrodynamic lubrication theory for elastic and isoviscous forms of lubrication are that the radial clearance and surface roughness are the most important factors determining lubrication; improved surface roughness from metal and ceramic bearing systems make fluid film lubrication possible and that the contact mechanics are extremely important in replacement design.

Jalali-Vahid et al. (2000) built a model with a ball-in-socket geometry, composed of an UHMWPE cup and a metal or ceramic femoral head. The femoral head was rotated at a steady angular velocity under constant load in a stationary femoral head. The constrained column model was used to describe the elastic deformation. The main purpose of this project was to determine the extent to which elasto-hydrodynamic lubrication plays a role in UHMWPE implants; therefore predicting the lubricating film thickness and pressure distribution. The model built was a steady state model and a ball-on-plane model was also developed to predict minimum film thickness for comparative purposes. Numerical stability and accuracy of this model was obtained by using a combination of half-point finite difference and backward difference schemes; and solved using the Newton-Rhapson method.

An important outcome that resulted was that the predicted film thickness is less than the surface roughness of the UHMWPE. This leads to asperity contact, thereby increasing wear. The model also showed a marked increase in film thickness when elastic deformation occurs. The minimum film thickness increases and pressure decreases when elastic deformation of the UHMWPE cup is considered.

The hydrostatic pressure distribution corresponds to a dry contact pressure distribution, thus indicating heavy loading conditions and which promotes the use of including elastic deformation. A higher pressure distribution is also observed for the ball-in-socket model than for the ball-on-plane model. This results from a pressure distribution which is caused by the vertical load. The pressure distribution is responsible for the load carrying capability; therefore the ball-in-socket geometry has higher pressures than the ball-on-plane geometry.

Jalali-Vahid et al. (2001) then developed an elasto-hydrodynamic lubrication model

built with the ball-in-socket geometry for the modelling of a metal femoral head articulating in an UHMWPE acetabular cup. In the model, the cup was also held stationary and the ball rotated at a steady angular velocity under a constant load. Different design parameters were altered to find the effect it had on the lubrication of the prosthesis. The main outcomes were that an increase in femoral head radius, decrease in radial clearance, increase in UHMWPE thickness, decrease in UHMWPE elastic modulus, all increase the predicted film thickness. However, the predicted film thickness was less than the roughness of the UHMWPE. This lead to the conclusion that mixed lubrication was the most probable lubrication regime for the UHMWPE-metal bearing system.

Jagatia et al. (2001) then attempted to analyse the lubricating film thickness in an UHMWPE hip joint replacement under the action of squeeze film motion. An elasto-hydrodynamic squeeze film lubrication model was built using a ball-in-socket configuration, from which only squeeze-film motion which was under constant load was considered. Only the elasticity of the UHMWPE was considered, as metal and ceramic are essentially rigid in comparison. Elastic deformation was calculated by using both the finite element method as well as the constrained column model, for comparison purposes. Conclusions made were that the elastic models gave similar results for elasto-hydrodynamic lubrication under squeeze-film motion, but that lubricant pockets were formed at the centre of the cup using the finite element method and not the constrained column method. It was also shown that the predicted lubricated film thickness is large, but not large enough to keep the surfaces apart.

Jagatia & Jin (2002) then built an elasto-hydrodynamic lubrication model for metal-on-metal implants. The aim of this study was to develop elasticity models for the two components. A finite-element model was used to determine the effect the pelvic bone has on the contact pressure distributions, and the effect of including the acetabular component fixation via a taper. This was done using a steady state model rotating in the flexion-extension plane under constant load and using a ball-in-socket configuration. A steady-state model was assumed, as the entraining and squeeze film action combined is believed to give relatively constant predicted film thickness under transient conditions.

It was shown that an increase of almost 2000N in the load, has little effect on the maximum pressure and minimum film thickness; this results from the flexibility of the acetabular component which accommodates for the increased load. The elastic deformation was also shown to be dependent on load, and not on the pressure distribution. The predicted film thickness was also much larger than that predicted by the Hamrock and Dowson formula and insensitive to the change in load (Jagatia & Jin, 2002). It was also shown that including the bone in the axisymmetric model has negligible influence on the predicted contact pressure distribution.

There has also been modelling work done on hip-resurfacing prostheses. Udofia & Jin (2003) built an elastohydrodynamic model in spherical co-ordinates for a ball-in-socket

joint under steady state rotation. It should be noted that in hip resurfacing prosthesis, a thin acetabular cup can increase the elastic deformation at the bearing surface.

Results of this study showed that a decrease in viscosity results in an increase in maximum hydrodynamic pressure and decrease in film thickness. The predicted film thickness is also much larger than those predicted for rigid hydrodynamic lubrication. This also shows the importance of modelling elastic deformation.

Jalali-Vahid et al. (2003) investigated the use of the Hamrock and Dowson equation for predicting minimum film thickness. It is important to note that, when solving the elasto-hydrodynamic lubrication problem for synovial fluid, it exhibits isoviscous behaviour. Isoviscous elasto-hydrodynamic systems can have significant elastic deformation, even under conditions of low loading. Film thickness is often solved by using equations in literature. The Hamrock and Dowson formula was based on lightly loaded conditions on a circular point; but this formula has been used in models for heavily loaded hip implants. A lubrication model based on a ball-on-plane model was developed with a modified entraining velocity for flexion-extension motion. The Hamrock and Dowson formula was shown to give good correlations with film thickness for loads less than 2500N and viscosity of $0.0009 Pa.s$.

Yew et al. (2004) did an elasto-hydrodynamic lubrication analysis of McKee-Farrar metal-on-metal hip replacements. The model used a ball-in-socket geometry and modelled steady-state rotation using the Reynold's equation and a finite element method for the elastic deformation. The purpose of the study was to show the effectiveness of elasto-hydrodynamic lubrication and the effect of thin walled acetabular cups on the McKee-Farrar implants. The conclusions reached were that the thin acetabular cup increased the elasticity of the implant,and that the bone and cement were important for the lubrication and contact mechanics analysis. It was shown that the film thickness is larger than that of normal metal-on-metal systems, but the system still operates in the mixed lubrication regime.

Liu et al. (2004) then did an elasto-hydrodynamic lubrication analysis of a Metasul bearing system, which is a metal-on-metal system with a polyethylene backing. Steady-state rotation was modelled under flexion-extension motion. The finite element model was used to model the elastic deformation, but with the use of the fast Fourier transform (FFT), the convergence time and numerical errors were significantly reduced as compared to the previously used direct summation method; this also allowed the use of lower viscosities. The elastic deformation of the cup was shown to be almost seven times that of the head. The film thickness was also shown to be relatively constant in the entraining direction; this is typical of isoviscous elasto-hydrodynamic lubrication contacts. The hydrodynamic pressure distribution was also relatively constant within the contact area; this results from the elastic deformation of the UHMWPE backing which somewhat reduces the pressure in the polar region. It was also concluded for the given system, that

reducing the radial clearance increases the predicted film thickness and that the use of UHMWPE backing improves the probability of fluid film lubrication occurring.

Liu et al. (2006a) proceeded to develop a transient elasto-hydrodynamic lubrication model for a metasul bearing system. It simulated dynamic loading conditions, those of the stance and swing phase with dynamic swing velocity. A ball-in-socket geometry was used and the elasticity equations were solved using a finite element method via FFT. Both entraining and squeeze-film motion were modelled; in which in-vitro and in-vivo simulations were conducted. The dynamic operating conditions were a vertical load and three rotations; in accordance with ISO standards. Only the flexion-extension and vertical loads were considered. Dilute bovine serum was used for in-vitro modelling and peri-prosthetic synovial fluid for in-vivo modelling.

Results showed that the film thickness remained relatively constant, even during the initial change to heavy loading at the beginning of the cycle (as is evident from the Bergmann cycle, an example of which can be seen in figure 5.1, loads are typically of an order magnitude higher than shown in this figure). During the remaining stance phase, the film thickness is greatly increased, due to increased speed and squeeze film action. The film thickness during the swing phase, decreased even though speed was increased and load was decreased. The UHMWPE backing increased the predicted film thickness by 40-50% as compared to full metal backing.

In-vitro simulation results were similar to in-vivo results; except that the film thickness was reduced by about 20%, this mainly due to the decreased viscosity of bovine serum. The transient film thickness obtained was higher than those obtained from quasi-static models and averages. The simplified transient elasto-hydrodynamic lubrication analysis is not necessarily applicable to metal-on-metal bearing systems. Large deformation of the UHMWPE backing promotes large film thickness which could lead to full-fluid film lubrication under smooth surfaces. The in-vivo simulation was dominated by full-fluid film lubrication, whereas the in-vitro simulation was dominated by mixed lubrication.

In another study, Liu et al. (2006b) aimed to find the effect of structural support of a sandwich cup on a transient elasto-hydrodynamic lubrication model. The roles of the UHMWPE and radial clearance of the metal-on-metal bearing system were also investigated. The model built made use of a ball-in-socket configuration with dynamic variation of load and speed and a linear elastic deformation equation. The outputs of the model were the film thickness and pressure distribution; these were used to find 3-d minimum and central film thickness as well as maximum pressure.

The deformation of the polyethylene in the sandwich cup reduces the maximum contact pressure, especially in the stance phase when the load is large, and this causes an increase in the film thickness; the high loads result in more contact area and improves the squeeze film action. This shows the synergistic effect of polyethylene deformation and squeeze film for large loads. Similar outputs were obtained during the swing phase

for both monolithic and sandwich cups. The bearing geometry plays an important role during the swing phase when the dominant lubrication regime is boundary lubrication and surface roughness therefore plays an important role.

2.5 Summary

The hip joint is a synovial, ball and socket joint, between the femur and acetabulum. The bones are covered by hyaline cartilage, commonly referred to as articular cartilage, and the cavity is surrounded by the synovial membrane. The synovial membrane produces synovial fluid, a dialysate of blood plasma containing hyaluronic acid and proteins, which provide lubrication and nutrients for the joint.

The hip joint is known to degenerate due to many factors. This has necessitated the use of artificial hip implants.

The first real exploration in the design of hip joints made by Sir John Charnely; his aim was to construct low friction prostheses. Originally metal bearing systems were used; but these failed due to poor design. Charnely reverted to polymers to solve his friction problem. His first attempt was with PTFE; however this resulted in excessive wear and subsequently failed. He then switched to UHMWPE, at the price of increased friction but better wear properties. The UHMWPE cup system is still in use today; most artificial hip joints have been constructed from metal femoral heads articulating in an UHMWPE acetabular cup.

The use of UHMWPE cups have come under scrutiny. These cups are observed to have been penetrated, initially by creep, but thereafter by wear. Excessive wear debris have ensued despite better processing to improve the wear characteristics by effective cross-linking. The wear debris produced are within the size range which makes them biologically active. The wear debris cause the degeneration of the peri-prosthetic tissue, that leads to osteolysis, and subsequently to the loosening of the implant.

This has lead to the design of prostheses with better wear characteristics. The re-introduction of metal-on-metal bearing systems and the increasing use of ceramic-on-ceramic systems is testament to this fact. The use of cushion form bearings as an articulating surface to help induce fluid-film lubrication has also been investigated.

Second generation metal-on-metal bearing systems are much improved prostheses. They show better performance than their first generation counterparts due to improved design. These metal-on-metal bearing systems don't produce debris that result in loosening; but it has an adverse effect in the high metal ion levels in the blood stream and the granulomas found in the lymph nodes. Hip resurfacing has also been increasingly used to try and prolong the date that necessitates total hip replacement; this is as a result of the active lifestyle of the younger generation, which has increased the wear of hip degeneration.

Ceramic-on-ceramic systems are smoother and produce extremely low amounts of wear debris as compared to other systems. The use of ceramic has been stifled due to the excessive cost of the device; this being the result of difficult manufacturing processes used to improve the ceramic properties.

The modelling of hip joints began with the complex system of the natural joint; this included both the synovial cavity and articular cartilage. The incorporation of the articular cartilage leads to difficulties as it is made of two continua, namely the synovial fluid and the solid porous cartilage matrix. With these aspects in mind, a modified form of the Navier equations of motion were used. Models developed were largely dependent on the physical parameters of the solid matrix. Natural joints were also modelled incorporating more complex fluid models; this showed a significant gelling effect which would improve lubrication.

In later years, the attention was then focussed on hip implants due to increased hip replacement surgeries. Modelling of the implants is largely based on the use of the Reynolds equation of hydrodynamic lubrication as implemented and shown in the many papers of Jin and Dowson. Most models incorporate the elastic deformation of surfaces and therefore model elasto-hydrodynamic lubrication. Elastic deformation was initially incorporated by a linear constrained column model, which was later replaced by the finite element method. Increased understanding of the mathematics and improved computer software and hardware available allowed for more accurate and time saving analysis by implementing the FFT method for deformation calculation.

As with improved calculation of deformation, other just as important aspects such as geometry and movement also improved. The geometries initially used were simple ball-on-plane models which were dropped in favour of the more realistic ball-in-socket model. Many motions were investigated: entrainment, squeeze, rotation. Eventually, the motion now used in most simulations is resembling a realistic walking cycle which includes both the stance and swing phases.

Consensus has been reached that in most of the models, the predominant method of lubrication is in the mixed lubrication regime. The various bearing systems and design parameters used in simulations also suggest that improved lubrication can be obtained if polyethylene backing is used, radial clearance is decreased and femoral head diameter is increased.

CHAPTER 3

Model Development

This chapter will show how the Reynolds equation came to be the foundation upon which the model was built. It leads us from the conservation of mass to the important aspects of the conservation of momentum equation and resultant Navier-Stokes equation. The application of boundary layer and lubrication theory leads to the Reynolds equation; thereafter the equation is manipulated to represent a hip joint, wherein different movements can be modelled.

3.1 Derivation of the Navier-Stokes Equation

Lubricants in many hydrodynamic lubrication analyses are assumed to be Newtonian and in the laminar flow regime. Navier derived the equation of fluid flow based on molecular considerations and Stokes derived the flow of viscous fluids in a different form. This basic equation is called the "Navier-Stokes equation of motion" and is derived from the dynamic equilibrium of a fluid element (Hamrock et al., 2004: 158).

Einstein or Indicial Notation

Einstein notation is a convenient method of representing formulae in multiple coordinates. This convention was introduced by Albert Einstein, and is applied in many fields of mathematics and applied mathematics. This convention is extremely useful when working in Euclidean space; not only does it compact many equations (by simplifying the manner in which vectors are expressed), but also helps to identify hidden relationships usually unseen when utilising normal conventions (Malan, 2006).

One of the most powerful aspects of Einstein notation is the Einstein summation

convention, which is shown below (Malan, 2006).

$$\begin{aligned} \text{Normally} \quad & y = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n \\ \text{which can be represented as} \quad & y = \sum_{i=1}^{n} a_ix_i \\ \text{and in Einstein notation} \quad & y = a_ix_i \quad for \quad i = 1, 2, 3, \cdots, n \end{aligned}$$

This is true, because in Einstein notation the repetition of a subscript in a single term implies summation.

The vector differential operator is expressed as follows (Malan, 2006)

$$\begin{aligned} \nabla &= \frac{\partial}{\partial x_1}\hat{e_1} + \frac{\partial}{\partial x_2}\hat{e_2} + \frac{\partial}{\partial x_3}\hat{e_3} \\ &= \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) \\ &= \frac{\partial}{\partial x_i}\hat{e_i} \quad for \quad i = 1, 2, 3 \end{aligned}$$

The divergence of a vector $\vec{u}(x_1, x_2, x_3) = u_1\hat{e_1} + u_2\hat{e_2} + u_3\hat{e_3}$ can be expressed as (Malan, 2006)

$$\begin{aligned} div\vec{u} &= \nabla \bullet \vec{u} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} \\ &= \frac{\partial u_i}{\partial x_i} \quad for \quad i = 1, 2, 3 \end{aligned}$$

The final expression from the Einstein notation that has been implemented in this work is the Kronecker delta, δ_{ij} (Malan, 2006).

$$\delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$$

In this work, it will also be assumed that the use of the subscripts i, j, k will automatically imply that the subscripts equal to the three Cartesian directions, therefore $1, 2, 3$ (where $x_1 = x$, $x_2 = y$ and $x_3 = z$ in conventional notation).

3.1.1 Reynolds Transport Theorem

A suitable analogy can be made between solid and fluid mechanics. In fluid mechanics, the concern is with fluid flow; in solid mechanics the concern is with particle displacement. Particle displacement is easier to describe whilst following the path of the particle; this is referred to as the *Lagrangian* description. In fluid mechanics this is not always the case,

it is often more likely to describe the velocity field as a function of position and time (as opposed to following each different fluid particle), this is called the *Eulerian* description (White, 2006: 16).

The Eulerian coordinate system is therefore a more appropriate system for fluid mechanics, but the fundamental laws of mechanics: conservation of mass, momentum and energy, are expressed in Lagrangian coordinates.

To solve this problem, the *Reynolds Transport Theorem* is applied,

$$\frac{D}{Dt}\int_{V(t)} \phi \partial V = \int_{V(t)} \left\{ \frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x_j}(\phi u_j) \right\} \partial V \tag{3.1}$$

Equation 3.1 expresses the Lagrangian derivative, LHS of equation 3.1, of a scalar (extensive) property, ϕ, in terms of an Eulerian derivative integrated over an arbitrary and changing volume, V, which has a fluid velocity, u, RHS of equation 3.1 (Crowe et al., 2005: 154). The first term is referred to as the local derivative, while the second expression represents the convective derivatives. This is useful since the conservation of mass and momentum will be derived via a control volume analysis.

3.1.2 Conservation of Mass

The conservation of mass is the starting point for the derivation of our model, which is described by the Lagrangian derivative of density, ρ.

$$\frac{D}{Dt}\int_{V(t)} \rho \partial V = 0 \tag{3.2}$$

Applying Reynolds transport theorem to equation 3.2 (White, 2006: 90), thereby converting from Lagrangian to Eulerian coordinates results in

$$\int_{V(t)} \left\{ \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j) \right\} \partial V = 0$$

and since this is true for any V

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j) = 0 \tag{3.3}$$

If it can be assumed that the density variations are constant, ie. incompressible, then equation 3.3 becomes

$$\frac{\partial u_i}{\partial x_i} = \nabla \bullet \mathbf{u} = div\mathbf{u} = 0$$

3.1.3 Stress in a Fluid

The stress force in a fluid can be described as follows

$$\lim_{\delta S \to 0} \frac{F_n}{\delta S} = \sigma = \text{Stress Force}$$

where F_n is the force normal to the surface S.

In a fluid element there are three mutually perpendicular surfaces with three stresses. The stress, σ, is therefore a tensor with nine components and can be written as

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

The position of the σ_{ij}'s in the array, correspond to applied forces in each coordinate direction. The first subscript refers to the coordinate direction perpendicular to the surface in which the stress acts; the second is the coordinate direction in which the stress acts (Welty et al., 2001: 96). The diagonal elements in the tensor represent the normal stresses and all the other elements represent shear stresses.

Note that $\sigma_{ij} = \sigma_{ji}$, therefore σ is symmetric; symmetry is required to satisfy equilibrium of moments about the three axes of the element. The stress at a point is also defined as follows:

$$n_i \sigma_{ij} = \text{the stress at a point}$$

where n_i is the normal vector of the surface.

3.1.4 Fluid Kinematics

Kinematic properties of a fluid are those such as

- linear and angular velocity
- vorticity
- acceleration
- strain rate

These are properties of the flow field itself, instead of the fluid (White, 2006: 15).

At a fixed point in time, the velocity derivative can be split into two parts, strain rate and angular velocity (one symmetric and one anti-symmetric):

$$\delta u_j = \frac{\partial u_j}{\partial x_k} \delta x_k = \frac{1}{2} \left\{ \frac{\partial u_j}{\partial x_k} + \frac{\partial u_k}{\partial x_j} \right\} \delta x_k + \frac{1}{2} \left\{ \frac{\partial u_j}{\partial x_k} - \frac{\partial u_k}{\partial x_j} \right\} \delta x_k$$

where the first term on the *RHS* represents the shear component of the fluid and the second term represents the rigid body rotation component of the fluid (Malan, 2006).

We define

$$\epsilon_{jk} = \frac{\partial u_j}{\partial x_k} + \frac{\partial u_k}{\partial x_j} \qquad \zeta_{jk} = \frac{\partial u_j}{\partial x_k} - \frac{\partial u_k}{\partial x_j}$$

and therefore

$$\delta u_j = \frac{1}{2}\epsilon_{jk}\delta x_k + \frac{1}{2}\zeta_{jk}\delta x_k$$

where ϵ is referred to as the *rate of strain tensor*; it is the measure of the distortion of the fluid element.

We will now define vorticity, $\vec{\psi}$, which is twice the angular velocity (rate of rotation); the velocity and vorticity are related by:

$$\vec{\psi} = (\psi_1, \psi_2, \psi_3) = \nabla \times \mathbf{u} = curl\mathbf{u}$$

The vorticity vector is *solenoidal* and is closely related to convective acceleration (White, 2006: 20).

Consider two arbitrary fixed points in a fluid, A and B, then

$$\begin{aligned} (AB)^2 &= \delta x_j^2 \qquad \text{and} \\ \frac{D}{Dt}(AB)^2 = 2\delta x_j \delta u_j &= \epsilon_{jk}\delta x_k \delta x_j + \zeta_{jk}\delta x_k \delta x_j \\ &= \epsilon_{jk}\delta x_k \delta x_j \end{aligned}$$

since $\zeta_{jk}\delta x_k \delta x_j = 0$ because it is an anti-symmetric (Middelman, 1998: 136). This is true if A and B are fixed points. This shows that the term is dependent only on the rate of strain tensor, ϵ, which is descriptive of the local rate of change of the fluid. Note that when $\vec{\psi} = 0$, the flow is *irrotational*.

3.1.5 Constitutive Equation

Assume that the fluid is Newtonian, and the stress and rate of strain are linearly related; therefore

$$\sigma_{ij} = \alpha_{ij} + \beta_{ijkl}\epsilon_{kl}$$

This assumes that the fluid is continuous.

When $u_i = \vec{0}$, the fluid is at rest, the viscous stresses of the fluid become zero and the normal stresses in the fluid are equal to the hydrostatic pressure (this is the same as the thermodynamic pressure, p) (White, 2006: 65); therefore in the absence of deformation the only force exerted is pressure. The hydrostatic pressure in the fluid is calculated to

be the average of the three normal stress components (Malan, 2006).

$$p = -\frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33})$$

Given $u_i = 0$, which implies $\epsilon_{kl} = 0$, and

$$\begin{aligned} \mathbf{P} &= -pn_i \\ \therefore P_j &= n_i\sigma_{ij} = n_i\alpha_{ij} \qquad \text{since} \qquad u_j = \epsilon_{jk} = 0 \\ \therefore \alpha_{ij} &= -p\delta_{ij} \end{aligned}$$

This shows that α_{ij} is not influenced by the rate of shear strain.

Assume the fluid is isotropic (properties are independent of direction); this allows the deformation to be independent of the coordinate axes (Malan, 2006).

$$\begin{aligned} \therefore \beta_{ijk}\epsilon_{kl} &= \frac{\lambda}{2}\epsilon_{kk}\delta_{ij} + \mu\epsilon_{ij} \\ \therefore \sigma_{ij} &= -p\delta_{ij} + \frac{\lambda}{2}\epsilon_{kk}\delta_{ij} + \mu\epsilon_{ij} \end{aligned}$$

This is the constitutive equation relating stress to deformation. If the fluid is incompressible,

$$\sigma_{ij} = -p\delta_{ij} + \mu\epsilon_{ij}$$

This equation shows that the shear stress depends on the rate at which the fluid is being distorted. More simply,

$$\sigma_{ij} = \mu\epsilon_{ij} = \mu\frac{\partial u_i}{\partial x_j} \qquad i \neq j$$

Assume a steady state 2-d flow where $u_1(x_2)$ and $u_2 = 0$, then

$$\begin{aligned} \mu\epsilon_{12} &= \mu\left\{\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right\} \qquad \text{since} \qquad u_2 = 0 \\ &= \mu\frac{\partial u_1}{\partial x_2} \end{aligned}$$

and apply Newton's law of viscosity (Welty et al., 2001: 96) where

$$\begin{aligned} \text{shear stress} &\propto \frac{\partial u_1}{\partial x_2} \\ \therefore \text{shear stress} &= K\frac{\partial u_1}{\partial x_2} \\ \text{where} \qquad K &= \mu \end{aligned}$$

ϵ_{kk} is the divergence of the velocity vector, this measures the rate at which fluid flows out at each point; it is referred to as the dilation and indicates the expansion of the

fluid. λ is known as the second coefficient of viscosity, and since it is associated only with volume expansion, is referred to as the *coefficient of bulk viscosity* (White, 2006: 66).

The normal stress is defined equal to the mechanical pressure, $\bar{p}$.

$$\begin{aligned} \bar{p} &= -\frac{1}{3}\left(-3p + \frac{3\lambda}{2}\epsilon_{kk} + \mu\epsilon_{kk} \right) \\ p - \bar{p} &= \frac{1}{3}\left(\frac{3\lambda}{2}\epsilon_{kk} + \mu\epsilon_{kk} \right) = \left\{ \lambda + \frac{2\mu}{3} \right\} \frac{\partial u_k}{\partial x_k} \end{aligned}$$

This states that the mean pressure in deforming a viscous fluid is not the same as the thermodynamic pressure. This is not always a huge problem, as for most problems $\frac{\partial u_k}{\partial x_k}$ is very small in magnitude. Stokes hypothesis of $\lambda + \frac{2\mu}{3} = 0$ and the incompressible flow assumption, $\frac{\partial u_k}{\partial x_k} = 0$ negate this problem. In reality this is a controversial topic since the second coefficient of viscosity does need to be applied in shock wave and sound-wave absorption and attenuation problems; since it is apparently frequency-dependent (White, 2006: 66).

3.1.6 Conservation of Momentum

The conservation of momentum requires the application of Newton's second law; this is written in the Lagrangian coordinate system, just as Newton himself envisioned it. In this derivation we will continue using a control volume, V, bounded by a closed surface, S. The law can be expressed in words simply as,

$$\text{Rate of change of momentum} = \text{Force}$$

In this case the momentum principle is used to analyse a problem involving forces and flow. The forces under consideration are (Malan, 2006):

- Body forces; these are forces such as gravity and electromagnetic potential; only the body force will be considered.

$$\text{Total body force} = \int_{V(t)} \rho \mathbf{f} \partial V$$

- surface forces as a result of internal interactions within the fluid and those external stresses applied on the surface of the fluid (element)

$$\text{Total surface force} = \int_{S(t)} \mathbf{P} \partial S$$

The momentum of the fluid can be expressed as,

$$\text{momentum of fluid in} \quad V = \int_{V(t)} \rho \mathbf{u} \partial V$$

Applying Newton's second law,

$$\begin{aligned}
\frac{D}{Dt}\int_{V(t)} \rho \mathbf{u} \partial V &= \int_{V(t)} \rho \mathbf{f} \partial V + \int_{S(t)} \mathbf{P} \partial S \\
\frac{D}{Dt}\int_{V(t)} \rho u_j \partial V &= \int_{V(t)} \rho f_j \partial V + \int_{S(t)} P_j \partial S \\
\frac{D}{Dt}\int_{V(t)} \rho u_j \partial V &= \int_{V(t)} \rho f_j \partial V + \int_{S(t)} n_i \sigma_{ij} \partial S
\end{aligned}$$

Applying Reynolds transport theorem to the *LHS* and the divergence theorem to the *RHS*, this results in equation 3.4

$$\int_{V(t)} \left\{ \frac{\partial}{\partial t}(\rho u_j) + \frac{\partial}{\partial x_k}(u_k \rho u_j) \right\} \partial V = \int_{V(t)} \rho f_j \partial V + \int_{V(t)} \frac{\partial \sigma_{ij}}{\partial x_i} \partial V \tag{3.4}$$

Equation 3.4 holds for any V; therefore the integrals can be removed.

$$\frac{\partial}{\partial t}(\rho u_j) + \frac{\partial}{\partial x_k}(u_k \rho u_j) = \rho f_j + \frac{\partial \sigma_{ij}}{\partial x_i} \tag{3.5}$$

Equation 3.5 is the conservation of momentum equation in conservative form (Malan, 2006). Now, expanding the *LHS* of equation 3.5.

$$\begin{aligned}
\rho \frac{\partial u_j}{\partial t} + u_j \frac{\partial \rho}{\partial t} + u_j \frac{\partial}{\partial x_k}(\rho u_k) + \rho u_k \frac{\partial u_j}{\partial x_k} &= \rho f_j + \frac{\partial \sigma_{ij}}{\partial x_i} \\
\text{since} \qquad u_j \left\{ \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_k}(\rho u_k) \right\} &= 0 \\
\rho \frac{\partial u_j}{\partial t} + \rho u_k \frac{\partial u_j}{\partial x_k} &= \rho f_j + \frac{\partial \sigma_{ij}}{\partial x_i}
\end{aligned} \tag{3.6}$$

Equation 3.6 is the non-conservative form of the momentum equation (Malan, 2006).

3.1.7 The Navier-Stokes Equation

Applying the constitutive equation to the relevant stress term in equation 3.6, leaves the following expression.

$$\begin{aligned}
\frac{\partial \sigma_{ij}}{\partial x_i} &= \frac{\partial}{\partial x_i}\left\{ -p\delta_{ij} + \frac{\lambda}{2}\epsilon_{kk}\delta_{ij} + \mu \epsilon_{ij} \right\} \\
&= -\frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_j}\left\{ \lambda \frac{\partial u_k}{\partial x_k} \right\} + \frac{\partial}{\partial x_j}\left\{ \mu \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right\}
\end{aligned}$$

The momentum equation is therefore,

$$\frac{\partial}{\partial t}(\rho u_j) + \frac{\partial}{\partial x_k}(u_k \rho u_j) = \rho f_j - \frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_j}\left\{\lambda \frac{\partial u_k}{\partial x_k}\right\} + \frac{\partial}{\partial x_j}\left\{\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right\} \tag{3.7}$$

Equation 3.7 is the Navier-Stokes equation in non-conservative form. Applying the conservation of mass leads to the conservative form of the Navier-Stokes equation (Malan, 2006).

In conservative form,

$$\rho \frac{\partial u_j}{\partial t} + \rho u_k \frac{\partial u_j}{\partial x_k} = \rho f_j - \frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_j}\left\{\lambda \frac{\partial u_k}{\partial x_k}\right\} + \frac{\partial}{\partial x_j}\left\{\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right\} \tag{3.8}$$

If the fluid is incompressible and the viscosity is constant, equation 3.8 simplifies to equation 3.9.

$$\rho \frac{\partial u_j}{\partial t} + \rho u_k \frac{\partial u_j}{\partial x_k} = \rho f_j - \frac{\partial p}{\partial x_j} + \mu \frac{\partial^2 u_j}{\partial x_j^2} \tag{3.9}$$

3.2 Derivation of the Reynolds Equation

3.2.1 Assumptions Based on Lubrication Theory

The lubrication of two bodies in close contact is achieved by viscous fluid flow through a narrow gap between the bodies. The derivation begins with the incompressible, Newtonian, isothermal form of the momentum equation, Navier-Stokes, in conservative form, equation 3.9. Applying an assumption from lubrication theory, namely negligible inertia (White, 2006: 180) (Hamrock et al., 2004: 188), reduces equation 3.9 to

$$\rho f_j - \frac{\partial p}{\partial x_j} + \mu \frac{\partial^2 u_j}{\partial x_j^2} = 0$$

Making use of dimensional analysis, this assumptions requires that

$$\frac{\rho u L}{\mu}\left(\frac{h}{L}\right)^2 \ll 1$$

It also shows that the Reynolds number (based on length, L), Re_L, can be large, as long as the gap is small, where

$$Re_L = \frac{\rho u L}{\mu} = \frac{\text{inertia forces}}{\text{viscous forces}}$$

This can be further simplified by assuming that the body force has negligible effect on the fluid flow and pressure distribution. These assumptions result in a much more

simplified equation which describes Stokes' flow:

$$\frac{\partial p}{\partial x_j} = \mu \frac{\partial^2 u_j}{\partial x_j^2}$$

3.2.2 Assumptions Based on Boundary Layer Theory

From this point, the general Reynolds equation will be derived using standard Cartesian coordinates; thereafter the relation will be shown in a more convenient configuration. With the chosen coordinate system the film will be in the x_3 direction.

Applying standard boundary layer theory, the following assumptions can be made based upon work done by Prandtl (White, 2006: 221) (Malan, 2006).

$$\begin{array}{lllll} \dfrac{\partial^2 u_1}{\partial x_1^2} \ll \dfrac{\partial^2 u_1}{\partial x_3^2} & \text{and} & \dfrac{\partial^2 u_1}{\partial x_2^2} \ll \dfrac{\partial^2 u_1}{\partial x_3^2} & \text{also that} \\ \dfrac{\partial^2 u_2}{\partial x_1^2} \ll \dfrac{\partial^2 u_2}{\partial x_3^2} & \text{and} & \dfrac{\partial^2 u_2}{\partial x_2^2} \ll \dfrac{\partial^2 u_2}{\partial x_3^2} & \text{and finally} \\ u_3 \ll u_1 & \text{and} & u_3 \ll u_2 & \end{array}$$

These assumptions are however only valid when $\sqrt{Re_L} \gg 1$. These simplifications result in the following equations to describe the fluid flow for lubrication (in expanded notation)

$$\frac{\partial p}{\partial x_1} = \mu \frac{\partial^2 u_1}{\partial x_3^2} \tag{3.10}$$

$$\frac{\partial p}{\partial x_2} = \mu \frac{\partial^2 u_2}{\partial x_3^2} \tag{3.11}$$

$$\frac{\partial p}{\partial x_3} = 0 \tag{3.12}$$

and the conservation of mass equation for incompressible fluids,

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0 \tag{3.13}$$

Importantly, equations 3.10, 3.11 and 3.12 also shows that $p = f(x_1, x_2, t)$. These assumptions also show that the gradient in the x_3 direction is negligible.

3.2.3 Solution of Fluid Equations to Reynolds Equation

Imposing the following boundary conditions:

- $\mathbf{u} = \vec{0}$ on $x_3 = 0$
- $\mathbf{u} = \left(U_1, U_2, U_3\right)$ on $x_3 = h$

where h is the film thickness, the distance between surfaces; to equations 3.10 and 3.11; the resulting velocity profiles are obtained after integration with respect to x_3,

$$u_1 = \frac{\partial p}{\partial x_1}\left(\frac{x_3^2 - x_3 h}{2\mu}\right) + \frac{U_1 x_3}{h}$$
$$u_2 = \frac{\partial p}{\partial x_2}\left(\frac{x_3^2 - x_3 h}{2\mu}\right) + \frac{U_2 x_3}{h}$$

Conservation of mass is now imposed by integrating equation 3.13 across the layer.

$$\int_0^{h(x_1,x_2,t)} \frac{\partial u_3}{\partial x_3}\partial x_3 = -\int_0^{h(x_1,x_2,t)} \left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2}\right)\partial x_3 \tag{3.14}$$

The film thickness, h, is however dependent on x_1, x_2 and t; it therefore requires adjustment for the changing boundary condition. This is taken into account by applying Liebnitz's theorem of integration (Hamrock et al., 2004: 192); the theorem simplifies the integrals into the following form,

$$\frac{\partial}{\partial x_1}\int_0^{h(x_1,x_2,t)} u_1(x_1,x_2,x_3,t)\partial x_3 = \int_0^h \frac{\partial u_1}{\partial x_1}\partial z + u_1(x_1,x_2,h,t)\frac{\partial h}{\partial x_1} \tag{3.15}$$
$$\frac{\partial}{\partial x_2}\int_0^{h(x_1,x_2,t)} u_2(x_1,x_2,x_3,t)\partial x_3 = \int_0^h \frac{\partial u_2}{\partial x_2}\partial z + u_2(x_1,x_2,h,t)\frac{\partial h}{\partial x_2} \tag{3.16}$$

The integrals on the *LHS* of equations 3.15 and 3.16 are,

$$\int_0^{h(x_1,x_2,t)} u_1(x_1,x_2,x_3,t)\partial z = -\frac{\partial p}{\partial x_1}\frac{h^3}{12\mu} + \frac{U_1 h}{2}$$
$$\int_0^{h(x_1,x_2,t)} u_2(x_1,x_2,x_3,t)\partial z = -\frac{\partial p}{\partial x_2}\frac{h^3}{12\mu} + \frac{U_2 h}{2}$$

The terms $u_1(x_1,x_2,x_3,t)$ and $u_2(x_1,x_2,x_3,t)$ are velocities obtained from the boundary conditions,

$$u_1(x_1,x_2,x_3,t) = U_1$$
$$u_2(x_1,x_2,x_3,t) = U_2$$

and substituting these relations into equation 3.14, results in

$$\begin{aligned}\int_0^{h(x_1,x_2,t)} \frac{\partial u_3}{\partial x_3}\partial x_3 \quad = \quad & -\frac{\partial}{\partial x_1}\left\{-\frac{\partial p}{\partial x_1}\frac{h^3}{12\mu} + \frac{U_1 h}{2}\right\} + U_1\frac{\partial h}{\partial x_1} \\ & -\frac{\partial}{\partial x_2}\left\{-\frac{\partial p}{\partial x_2}\frac{h^3}{12\mu} + \frac{U_2 h}{2}\right\} + U_2\frac{\partial h}{\partial x_2}\end{aligned} \tag{3.17}$$

with further simplification, and rearranging of terms on the *LHS* and *RHS*, equation 3.17 becomes

$$\frac{\partial}{\partial x_1}\left\{\frac{h^3}{12\mu}\frac{\partial p}{\partial x_1}\right\}+\frac{\partial}{\partial x_2}\left\{\frac{h^3}{12\mu}\frac{\partial p}{\partial x_2}\right\}=U_3-\frac{U_1}{2}\frac{\partial h}{\partial x_1}-\frac{U_2}{2}\frac{\partial h}{\partial x_2} \tag{3.18}$$

We can rewrite this relation using kinematic boundary conditions on the surface $x_3 = h$,

$$\begin{aligned} 0=\frac{D}{Dt}\big(x_3-h(x_1,x_2,t)\big) &= \frac{dz}{dt}-\frac{\partial h}{\partial t}+u_1(x_1,x_2,h,t)\frac{\partial h}{\partial x_1}+u_2(x_1,x_2,h,t)\frac{\partial h}{\partial x_2} \\ &= U_3-\frac{\partial h}{\partial t}+U_1\frac{\partial h}{\partial x_1}+U_2\frac{\partial h}{\partial x_2} \\ \therefore U_3 &= \frac{\partial h}{\partial t}+U_1\frac{\partial h}{\partial x_1}+U_2\frac{\partial h}{\partial x_2} \end{aligned} \tag{3.19}$$

Substituting equation 3.19 into equation 3.18, we end up with the Reynolds equation in Cartesian coordinates,

$$\frac{\partial}{\partial x_1}\left\{\frac{h^3}{12\mu}\frac{\partial p}{\partial x_1}\right\}+\frac{\partial}{\partial x_2}\left\{\frac{h^3}{12\mu}\frac{\partial p}{\partial x_2}\right\}=\frac{\partial h}{\partial t}+\frac{U_1}{2}\frac{\partial h}{\partial x_1}+\frac{U_2}{2}\frac{\partial h}{\partial x_2} \tag{3.20}$$

Equation 3.20 can be written in a more general form as

$$\nabla\bullet\left(\frac{h^3}{12\mu}\nabla p\right)=\nabla\bullet\left(\frac{h\vec{U}}{2}\right)+\frac{\partial h}{\partial t} \tag{3.21}$$

where $\vec{U}$ is the surface velocity vector.

3.3 Reynolds Equation Applied to a Hip Joint

For the hip joint to be modelled, certain assumptions will be made to further simplify the system:

- the fluid will be isoviscous (Newtonian)
- the cup will be positioned horizontally
- walking cycle imposed will be based on the Bergmann walking cycle

Synovial fluid and blood plasma are known to have non-Newtonian fluid properties, but studies have concluded that at high shear rates, an isoviscous assumption is valid (Jin, 2006). The acetabular cup is anatomically positioned at 45°; however the contact mechanics allow for the model to be developed in a horizontal position. The Bergmann walking cycle consists of a loading pattern which is double-peaked, this has resulted from studies conducted by Bergmann et al. (1995) on the influence of heel strike on the loading of the hip joint. This loading pattern compares well to the telemetry from

total hip replacement patients obtained by Heller et al. (2001); realistic flexion-extension (rotation) movements were also obtained. For this reason, the Bergmann cycle was used to simulate the walking cycle in the model.

The coordinate system that would best describe the hip joint would be the spherical coordinate system, expansion of equation 3.21 in spherical coordinates is (Goenka & Booker, 1980) (Meyer, 2003)

$$\frac{1}{R^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\frac{h^3}{12\mu}\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial}{\partial\phi}\left(\frac{h^3}{12\mu}\frac{\partial p}{\partial\phi}\right)\right\}= \frac{1}{2R\sin\theta}\left(h\cos\theta U_\theta + h\sin\theta\frac{\partial U_\theta}{\partial\theta}+h\frac{\partial U_\phi}{\partial\phi}+U_\theta\sin\theta\frac{\partial h}{\partial\theta}+U_\phi\frac{\partial h}{\partial\phi}\right)+\frac{\partial h}{\partial t} \quad (3.22)$$

where R is the radius of the acetabular cup and where the surface velocity components, U^θ and U^ϕ are:

$$\begin{aligned} U_\theta &= -R\omega_{x_1}\sin\phi + R\omega_{x_2}\cos\phi \\ U_\phi &= -R\omega_{x_1}\cos\phi\cos\theta - R\omega_{x_2}\sin\phi\cos\theta + R\omega_{x_3}\sin\theta \end{aligned}$$

Substituting these relationships into equation 3.22 and simplifying, yields

$$\sin\theta\frac{\partial}{\partial\theta}\left(h^3\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{\partial}{\partial\phi}\left(h^3\frac{\partial p}{\partial\phi}\right)= 6\mu R^2\sin\theta\Big(\big(-R\omega_{x_1}\sin\phi+R\omega_{x_2}\cos\phi\big)\sin\theta\frac{\partial h}{\partial\theta} + \big(-R\omega_{x_1}\cos\phi\cos\theta - R\omega_{x_2}\sin\phi\cos\theta + R\omega_{x_3}\sin\theta\big)\frac{\partial h}{\partial\phi}+2\sin\theta\frac{\partial h}{\partial t}\Big) \quad (3.23)$$

If the axes are shifted, and the $x_1 - x_3$ plane lies in the horizontal position with x_2 in the vertical direction (Jin & Dowson, 1999), as can be seen in figure 3.1, equation 3.23 can be simplified with **rotation about the x_3 axis** to equation 3.24 (Liu et al., 2006b).

$$\frac{\partial}{\partial\theta}\left(h^3\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{\partial}{\partial\phi}\left(h^3\frac{\partial p}{\partial\phi}\right)=6\mu R^2\sin^2\theta\left(\omega_{x_3}\frac{\partial h}{\partial\phi}+2\frac{\partial h}{\partial t}\right) \quad (3.24)$$

Due to the rotation of the axes, **the loading axis is the x_2 axis**.The next problem is to describe the inward-outward movement of the hip joint. This is done by relating the eccentricity (or eccentricity ratio) of the ball in the socket, to the film thickness of the joint. This can be seen in the relationships shown by equation 3.25, for the eccentricity,

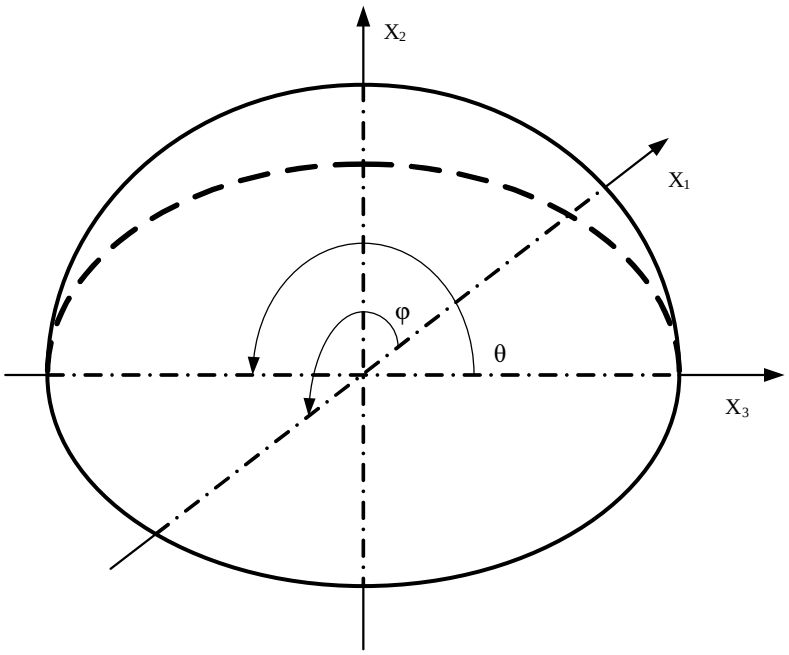

Figure 3.1: Hemispherical shell on rotated axes

and equation 3.26, for the eccentricity ratio (Liu et al., 2006b).

$$h(\phi,\theta) = c - e_{x_1}\sin\theta\cos\phi - e_{x_2}\sin\theta\sin\phi - e_{x_3}\cos\theta \tag{3.25}$$

$$h(\phi,\theta) = c\big(1 - \xi_{x_1}\sin\theta\cos\phi - \xi_{x_2}\sin\theta\sin\phi - \xi_{x_3}\cos\theta\big) \tag{3.26}$$

The model now contains both the angular and inward-outward movement of the joint. The angular component forms part of the Reynolds equation whereas the inward-outward movement is incorporated via the eccentricities or eccentricity ratios in the film thickness relationship.

The model developed can therefore solve for the pressure distributions with specified angular velocity and eccentricity (or eccentricity ratio). The pressure distribution allows the resultant forces to be calculated as shown in equations 3.27, 3.28 and 3.29 (Jin & Dowson, 1999).

$$F_{x_1} = R^2\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} p\sin\theta\cos\phi\sin\theta\partial\theta\partial\phi \tag{3.27}$$

$$F_{x_2} = R^2\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} p\sin\theta\sin\phi\sin\theta\partial\theta\partial\phi \tag{3.28}$$

$$F_{x_3} = R^2\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} p\cos\theta\sin\theta\partial\theta\partial\phi \tag{3.29}$$

The limits of ϕ_1 and θ_1 is $0\,rad$ and for ϕ_2 and θ_2 is $2\pi\,rad$. The use of equations 3.27, 3.28 and 3.29 to calculate the force exerted is based on the assumption that the force caused by the acceleration of the femoral component is minimal; this is believed

to be valid as the radial clearance between the cup and head is small, and therefore the acceleration will be negligible.

CHAPTER 4

Numerical Solution Algorithm

This chapter gives a brief overview of the discretisation schemes implemented to solve for the Reynolds equation. The multigrid solution algorithm was introduced as it was used in solving the intermediary constraint problem. The application of the discrete equation and the preferred numerical method was combined to develop the algorithms needed for investigating the angular motion and walking cycle for the hip joint model.

In order to solve equation 3.24, a stable numerical procedure is required. As a result of the complex nature of the model, geometry and movement, as well as the convergence criteria required, an analytical solution is unlikely.

A provisional analysis of equation 3.24 leads to the following simple deductions:

- the variable to be solved for is pressure
- pressure is solved in two spatial dimensions
- it is a strictly elliptic equation
- it is not a transient numerical problem, with respect to pressure

These initial simple conclusions, show that the complexity of the Reynolds equation has been reduced. The Reynolds equation is not directly transient, since it doesn't contain a temporal term with respect to pressure; it can be mathematically described more generally as a modified form of Poisson's equation.

As the pressure variable is not a function of time, but a transient analysis is required, the steady state condition will be solved at all points in time, with no previous knowledge of the pressure. However, previous knowledge of the film thickness is required, wherein the transient behaviour is contained.

4.1 Discretisation

The first step in the numerical procedure is to choose the grid upon which to solve equation 3.24. There are many possible grids (meshes) to consider for this problem, depending on the solution required. Ideally, an unstructured grid would be sought so that grid adaptation would be possible; thus allowing for more grid points around areas of steeper gradient and therefore better accuracy (Malan & Craig, 2006).

The implementation of such grids, and refinement and adaptation thereof, is a complex procedure in itself; not to mention the increased complexity of solving the Reynolds equation.

It was thus decided to use a grid that would be based on a vertex-centred finite volume discretisation scheme which would easily reduce to the often used finite difference approximation schemes.

The geometry in this problem, spherical coordinates, does pose some problems. This results from discontinuities at the following points

$$\theta = 0 \qquad \theta = \pi$$
$$\phi = 0 \qquad \phi = \pi$$

The problem is eased somewhat as the chosen system is a hemispherical shell, and the coordinate axes have been rotated; this simple rotation puts these discontinuities on the boundary edge. A grid was developed, which relates the hemispherical surface to the spatial (spherical) dimensions, as shown in figures 4.1 and 4.2.

An equally spaced finite difference grid was used for the simulation. Even though this grid is limited in its use, the implementation thereof is much easier.

The equation to be solved is a second order partial differential equation, the discretisation leads to initial difficulties; using the standard approach for the given equation, the second order terms for the primary node would not be influenced by the neighbouring nodes (of the primary node). This can lead to larger errors, but also requires that the first line of interior nodes be modified to allow for the calculation of the primary node values; this problem is solved by using half-point discretisation (Jalali-Vahid et al., 2000).

The spatial derivative of the fluid film thickness was approximated by central differencing; even though the derivative can be obtained analytically, a discrete form was implemented for possible use in the case of elasto-hydrodynamic lubrication. Backward difference approximation was used for the temporal derivative of the fluid film thickness.

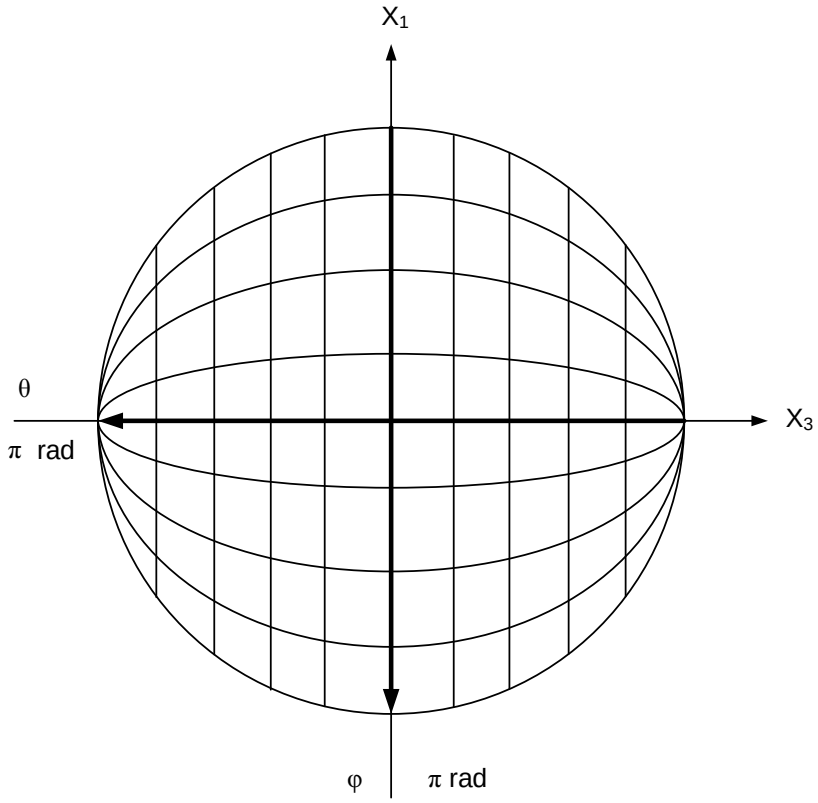

Figure 4.1: 2-dimensional top view of the relation of the angular coordinates to the rotated Cartesian coordinates

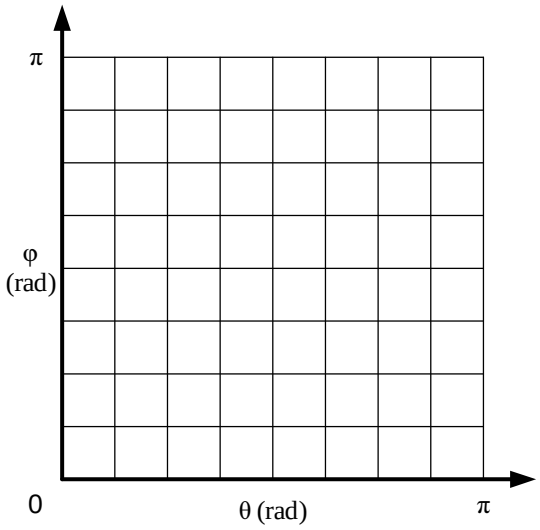

Figure 4.2: Grid developed for numerical simulation

The resulting equation is

$$p_{i,j+1}\left(\frac{\sin\theta_j h^3_{i,j+\frac{1}{2}}\sin\theta_{j+\frac{1}{2}}}{\Delta\theta_{j+\frac{1}{2},j-\frac{1}{2}}\Delta\theta_{j+1,j}}\right)+p_{i,j-1}\left(\frac{\sin\theta_j h^3_{i,j-\frac{1}{2}}\sin\theta_{j-\frac{1}{2}}}{\Delta\theta_{j+\frac{1}{2},j-\frac{1}{2}}\Delta\theta_{j,j-1}}\right)$$
$$-p_{i,j}\left[\frac{\sin\theta_j}{\Delta\theta_{j+\frac{1}{2},j-\frac{1}{2}}}\left(\frac{h^3_{i,j+\frac{1}{2}}\sin\theta_{j+\frac{1}{2}}}{\Delta\theta_{j+1,j}}+\frac{h^3_{i,j-\frac{1}{2}}\sin\theta_{j-\frac{1}{2}}}{\Delta\theta_{j,j-1}}\right)\right]+p_{i+1,j}\left(\frac{h^3_{i+\frac{1}{2},j}}{\Delta\phi_{i+\frac{1}{2},i-\frac{1}{2}}\Delta_{i+1,i}}\right)$$
$$+p_{i-1,j}\left(\frac{h^3_{i-\frac{1}{2},j}}{\Delta\phi_{i+\frac{1}{2},i-\frac{1}{2}}\Delta_{i,i-1}}\right)-p_{i,j}\left[\frac{1}{\Delta\phi_{i+\frac{1}{2},i-\frac{1}{2}}}\left(\frac{h^3_{i+\frac{1}{2},j}}{\Delta_{i+1,i}}+\frac{h^3_{i-\frac{1}{2},j}}{\Delta_{i,i-1}}\right)\right]$$
$$=6\mu R^2\sin^2\theta_j\left[\omega_{x_3}\left(\frac{h_{i+1,j}-h_{i-1,j}}{\Delta\theta_{i+1,i-1}}\right)+2\left(\frac{h^k_{i,j}-h^{k-1}_{i,j}}{\Delta t_{k,k-1}}\right)\right] \quad (4.1)$$

where i is representative of the node in the ϕ direction and j is representative of the node in the θ direction. The RHS of equation 4.1 can be seen as the source term when solving the equation.

Calculation of the forces is done numerically via the trapezoidal rule and the equations used for numerical approximation are equations 4.2, 4.3 and 4.4 (Jin & Dowson, 1999).

$$F_{x_1}=\frac{R^2}{4}\sum_{i=1}^{n}\sum_{j=1}^{m}\Big(p_{i,j}\sin\theta_j\cos\phi_i\sin\theta_j+p_{i+1,j}\sin\theta_j\cos\phi_{i+1}\sin\theta_j$$
$$+p_{i,j+1}\sin\theta_{j+1}\cos\phi_i\sin\theta_{j+1}+p_{i+1,j+1}\sin\theta_{j+1}\cos\phi_{i+1}\sin\theta_{j+1}\Big)\big(\theta_{j+1}-\theta_j\big)\big(\phi_{i+1}-\phi_i\big) \quad (4.2)$$

$$F_{x_2}=\frac{R^2}{4}\sum_{i=1}^{n}\sum_{j=1}^{m}\Big(p_{i,j}\sin\theta_j\sin\phi_i\sin\theta_j+p_{i+1,j}\sin\theta_j\sin\phi_{i+1}\sin\theta_j$$
$$+p_{i,j+1}\sin\theta_{j+1}\sin\phi_i\sin\theta_{j+1}+p_{i+1,j+1}\sin\theta_{j+1}\sin\phi_{i+1}\sin\theta_{j+1}\Big)\big(\theta_{j+1}-\theta_j\big)\big(\phi_{i+1}-\phi_i\big) \quad (4.3)$$

$$F_{x_3}=\frac{R^2}{4}\sum_{i=1}^{n}\sum_{j=1}^{m}\Big(p_{i,j}\sin\theta_j\cos\theta_j+p_{i+1,j}\sin\theta_j\cos\theta_j$$
$$+p_{i,j+1}\sin\theta_{j+1}\cos\theta_{j+1}+p_{i+1,j+1}\sin\theta_{j+1}\cos\theta_{j+1}\Big)\big(\theta_{j+1}-\theta_j\big)\big(\phi_{i+1}-\phi_i\big) \quad (4.4)$$

The matrix of pressure values at the specific nodes can be vectorised. Given a boundary condition, this results in a set of linear equations which is more commonly written in the form

$$A\mathbf{u}=\mathbf{f} \quad (4.5)$$

u is the pressure vector, **f** is the source term and A is the matrix that results from discretisation. The solution to equation 4.5 is denoted by **u** and the value of **u** in x_j is u_j. u_j approximates the solution in the vertex x_j; therefore equation 4.1 is called *vertex-centred discretisation* (Wesseling, 1992: 5).

The matrix A, is a sparse matrix, which has a largely banded form, as shown in figure 4.3. As a result of the sparse nature of the matrix, efficient solvers are available to solve the equation, this results in less computational time.

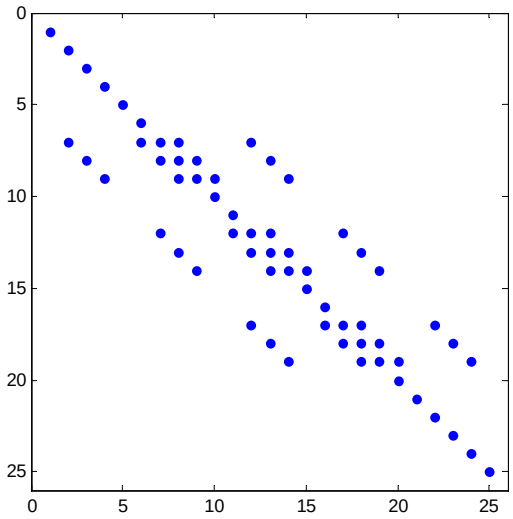

Figure 4.3: Sparse nature of matrix A for a 5×5 grid

4.2 Iterative Techniques

Equation 4.1 is also solved with a constraint, namely

$$if \qquad p_{ij} \leq 0 \qquad then \qquad p_{ij} = 0$$

where p_{ij} represents a point on the 2-dimensional grid. This is done by applying an iterative technique, whereby the constraint is applied at every iteration of every interior node.

4.2.1 Gauss-Seidel

In this section a brief description will be given of the iterative Gauss-Seidel method.

The Gauss-Seidel method is a classical technique that is not often used for systems of small dimensions as a result of accuracy demands which would lead to slower computation than direct methods. For systems of large sparse matrices, iterative techniques prove to

be much more efficient; as in the case of partial differential equations and boundary valued problems (Burden & Faires, 2005: 436).

The description will begin by discussing the Jacobi method. The Jacobi iterative method involves solving the i^{th} equation in $A\mathbf{u} = \mathbf{f}$ for u_i to obtain

$$u_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(- \frac{a_{ij} u_j}{a_{ii}} \right) + \frac{f_i}{a_{ii}} \qquad for \qquad i = 1, 2, \cdots, n$$

$\mathbf{u}$ is the exact solution to the system of equations and $\mathbf{v}$ is the approximation to the exact solution generated by some iterative method. u_j and v_j denote the j^{th} components of $\mathbf{u}$ and $\mathbf{v}$.

Generating v_i^k from components of $\mathbf{v}^{k-1}$ for $k \geq 1$ by

$$v_i^k = \frac{\sum_{\substack{j=1 \\ j \neq i}}^{n} (-a_{ij} v_j^{k-1}) + f_i}{a_{ii}} \qquad for \qquad i = 1, 2, \cdots, n$$

A possible improvement of the Jacobi method is the Gauss-Seidel method. In the Jacobi method $\mathbf{v}^{k-1}$ is used to calculate $\mathbf{v}^k$; however $v_1^k, \cdots, v_{i-1}^k$ has already been calculated, and these are probably better approximations to the solution than $v_1^{k-1}, \cdots, v_{i-1}^{k-1}$. The Gauss-Seidel method takes advantage of these updated values and are used to calculate v_i^k as follows

$$v_i^k = \frac{\sum_{j=1}^{i-1} (-a_{ij} v_j^k) + \sum_{j=i+1}^{n} (-a_{ij} v_j^{k-1}) + f_i}{a_{ii}} \qquad for \qquad i = 1, 2, \cdots, n$$

If the matrix A is a strictly diagonally dominant matrix; convergence of $\mathbf{v}^k$ to $\mathbf{u}$ will result from any choice of $\mathbf{v}^0$. In most cases, Gauss-Seidel will converge faster than the Jacobi method; and is therefore preferred (Burden & Faires, 2005: 440). The Gauss-Seidel method will be convergent if A is symmetric and positive-definite (Quateroni & Valli, 1991: 31).

4.2.2 Gauss-Seidel Applied to 2-dimensional Grids

For the Gauss-Seidel iterations, it has been assumed that $\mathbf{v}$ has been updated in ascending order. For the Jacobi method, the order is of no consequence since previous approximations are not immediately overwritten. Using Gauss-Seidel, updating plays an important role. The values at the nodes do not always have to be calculated in ascending order, but the order in which they are visited is largely problem dependent.

A few variations of the order in which the nodes are visited were investigated; however the most efficient method for the problem at hand was the alternating direction method. This method involves sweeping alternately by rows and columns (Briggs, 1987: 61).

4.3 Multigrid

The problem at hand can be described by a linear system of equations as shown in equation 4.5.

The exact solution to $A\mathbf{u} = \mathbf{f}$ is unknown, but an approximation $\mathbf{v}$, can be computed and is therefore known. There are two main methods of how $\mathbf{v}$ approximates $\mathbf{u}$. One method is the calculation of the algebraic error (Briggs, 1987: 7), $\mathbf{e}$, as shown in equation 4.6.

$$\mathbf{e} = \mathbf{u} - \mathbf{v} \tag{4.6}$$

The error is commonly measured by the second Euclidean norm (Briggs, 1987: 8) (Malan & Craig, 2006)

$$||\mathbf{e}||_2 = \left\{ \sum_{j=1}^{N} e_j^2 \right\}^{\frac{1}{2}} \tag{4.7}$$

Unfortunately, the error is just as much an unknown as $\mathbf{u}$. However, an attainable measure of how $\mathbf{v}$ approximates $\mathbf{u}$ is the residual, $\mathbf{r}$ (Briggs, 1987: 8).

$$\mathbf{r} = \mathbf{f} - A\mathbf{v} \tag{4.8}$$

The residual is an indication by which $\mathbf{v}$ fails to satisfy the original equation, equation 4.5. The residual is a vector and can be measured, similar to the error vector in equation 4.7, by the second Euclidean norm (Malan & Craig, 2006).

The original equation can be rearranged to include the definition of the residual

$$A\mathbf{v} = \mathbf{f} - \mathbf{r} \tag{4.9}$$

Subtracting equation 4.9 from equation 4.5, we obtain an important relationship between the error and the residual (Malan & Craig, 2006),

$$A\mathbf{e} = \mathbf{r} \tag{4.10}$$

Equation 4.10 is called the *residual equation*. By the uniqueness of the solution $\mathbf{u}$, $\mathbf{r} = 0$ only if $\mathbf{e} = 0$. The residual equation also states that the error satisfies the same set of equations as $\mathbf{u}$ when $\mathbf{f}$ is replaced by $\mathbf{r}$ (Briggs, 1987: 8).

It is relatively easy to compute the residual, using equation 4.8. To improve the approximation $\mathbf{v}$, it is possible to solve the residual equation for $\mathbf{e}$, then calculate the updated approximation by using the definition of the error as shown in equation 4.11(Briggs, 1987: 9).

$$\mathbf{u} = \mathbf{v} + \mathbf{e} \tag{4.11}$$

This leads to the more intricate properties of multigrid. Most standard iterative methods possess the smoothing property. The smoothing property of these relaxation methods are effective at eliminating high frequency errors, but leave low frequency errors unchanged (Malan & Craig, 2006). A method of improving the relaxation scheme is to use a good initial guess. A method of obtaining a good initial guess is by doing calculations on a coarser grid. Calculations on coarser grids are less expensive because there are fewer unknowns to update (Briggs, 1987: 33).

It is important to note that smooth (Fourier) modes on a fine grid look less smooth on a coarse grid. When relaxation begins to slow down on a fine grid, in the presence of smooth error modes, it is then when relaxation is done on coarser grids where the smooth error appears more oscillatory (Malan & Craig, 2006). Relaxing on the these oscillating errors on coarser grids can be done by relaxing on the error directly, by making use of the *residual equation*. It is also apparent that relaxing on the original equation, $A\mathbf{u} = \mathbf{f}$, with an arbitrary initial guess $\mathbf{v}$, is the same as relaxing on the residual equation with $\mathbf{e} = 0$ (Briggs, 1987: 35).

An effective strategy to implement would be to relax on the fine grid, approximating $\mathbf{v}^{\Delta x}$, where the superscript denotes the spacing of the grid (Briggs, 1987: 36). Then calculate the residual $\mathbf{r} = \mathbf{f} - A\mathbf{v}^{\Delta x}$. Relax on the residual equation, $A\mathbf{e} = \mathbf{r}$ on $\Omega^{2\Delta x}$, where Ω signifies the grid, thereby obtaining $\mathbf{e}^{2\Delta x}$. Thereafter, the approximation on $\Omega^{\Delta x}$ is corrected by including the error on $\Omega^{\Delta x}$, ie. $\mathbf{v}^{\Delta x} \longleftarrow \mathbf{v}^{\Delta x} + \mathbf{e}^{2\Delta x}$.

This is called coarse grid correction; it however requires the transfer of information from the fine grid to the coarse grid and vice versa. This is done by the methods of restriction and prolongation (Malan & Craig, 2006).

For inter-grid transfers from coarse to fine grids, prolongation almost always takes the form of interpolation; it is represented in the following manner $\mathbf{I}_{\Delta x}^{2\Delta x}$ where information is transferred from $\Omega^{\Delta x}$ to $\Omega^{2\Delta x}$ (Briggs, 1987: 37). Intergrid transfers from fine grids to coarse grids are restriction operators and the most obvious restriction operator is *injection*. This means that the coarse grid vector takes directly the value of the corresponding fine grid points. Restriction can also take the form of *weighting*; whereby neighbouring nodes give weight to the value of the node of interest. *Injection* will be used in the numerical procedure to be developed and is symbolised by $\mathbf{I}_{2\Delta x}^{\Delta x}$ where information is transferred from $\Omega^{2\Delta x}$ to $\Omega^{\Delta x}$ (Briggs, 1987: 39).

The basic concepts of multigrid have been discussed, all that is now necessary is to effectively combine these concepts to generate a multigrid algorithm. The aim of the multigrid algorithm is to smooth as many of the error modes as possible by using coarser grids; naturally this will utilise many grids in the procedure. This is the reason for the

use of multigrid; the rate of convergence of basic iterative methods can be improved by multigrid methods. Basic iterative methods have the property of decaying slowly to long Fourier modes, but reduce rapidly to short wavelength Fourier modes (Wesseling, 1992: 6). The principle of multigrid is to approximate the longer wavelength part of the error on coarser grids.

A problem then arises as to how to calculate the error of the residual equation which is the second finest grid; this is solved by realising that the residual equation is in the same form as the original equation, therefore a residual equation is used based on the equation for $\Omega^{2\Delta x}$.

For simplification of notation, $\mathbf{r}^{2\Delta x}$ will be represented by $\mathbf{f}^{2\Delta x}$ and similarly with $\mathbf{e}$ and $\mathbf{u}$. Note however, that apart from the finest grid, all the equations are residual equations and subsequent residual equations thereof. ν_1 and ν_2 represent how many relaxation sweeps should be done via the iterative technique. An algorithm with L grids is shown below to illustrate how the coarse grid correction scheme is applied to multiple grids (Briggs, 1987: 46).

Relax on $A^{\Delta x}\mathbf{u}^{\Delta x} = \mathbf{f}^{\Delta x}$ ν_1 times with initial guess $\mathbf{v}^{\Delta x}$.
Compute $\mathbf{f}^{2\Delta x} = \mathbf{I}^{2\Delta x}_{\Delta x}\mathbf{r}^{2\Delta x}$.
Relax on $A^{2\Delta x}\mathbf{u}^{2\Delta x} = \mathbf{f}^{2\Delta x}$ ν_1 times with initial guess $\mathbf{v}^{2\Delta x} = 0$.
Compute $\mathbf{f}^{4\Delta x} = \mathbf{I}^{4\Delta x}_{2\Delta x}\mathbf{r}^{4\Delta x}$.
Relax on $A^{4\Delta x}\mathbf{u}^{4\Delta x} = \mathbf{f}^{4\Delta x}$ ν_1 times with initial guess $\mathbf{v}^{4\Delta x} = 0$.
Compute $\mathbf{f}^{8\Delta x} = \mathbf{I}^{8\Delta x}_{4\Delta x}\mathbf{r}^{8\Delta x}$.
$\vdots$
Solve $A^{L\Delta x}\mathbf{u}^{L\Delta x} = \mathbf{f}^{L\Delta x}$.
$\vdots$
Correct $\mathbf{v}^{4\Delta x} \longleftarrow \mathbf{v}^{4\Delta x} + \mathbf{I}^{4\Delta x}_{8\Delta x}\mathbf{v}^{8\Delta x}$.
Relax on $A^{4\Delta x}\mathbf{u}^{4\Delta x} = \mathbf{f}^{4\Delta x}$ ν_2 times with initial guess $\mathbf{v}^{4\Delta x}$.
Correct $\mathbf{v}^{2\Delta x} \longleftarrow \mathbf{v}^{2\Delta x} + \mathbf{I}^{2\Delta x}_{4\Delta x}\mathbf{v}^{4\Delta x}$.
Relax on $A^{2\Delta x}\mathbf{u}^{2\Delta x} = \mathbf{f}^{2\Delta x}$ ν_2 times with initial guess $\mathbf{v}^{4\Delta x}$.
Correct $\mathbf{v}^{\Delta x} \longleftarrow \mathbf{v}^{\Delta x} + \mathbf{I}^{\Delta x}_{2\Delta x}\mathbf{v}^{2\Delta x}$.
Relax on $A^{\Delta x}\mathbf{u}^{\Delta x} = \mathbf{f}^{\Delta x}$ ν_2 times with initial guess $\mathbf{v}^{\Delta x}$.

It is also not always necessary to solve the equation on the coarsest grid; often a pre-defined number of relaxation sweeps is done, ν_3, where $\nu_3 = \nu_1 + \nu_2$. This was done in the algorithm implemented.

The algorithm tunnels down to the coarsest grid, and then works its way back up to the finest grid as seen in figure 4.4. The order in which the grids are visited resembles a V and is thus called the *V-cycle* (Malan & Craig, 2006). There are numerous other cycles in the multigrid family, but the *V-cycle* has been implemented for the problem at hand and will therefore be the only one discussed.

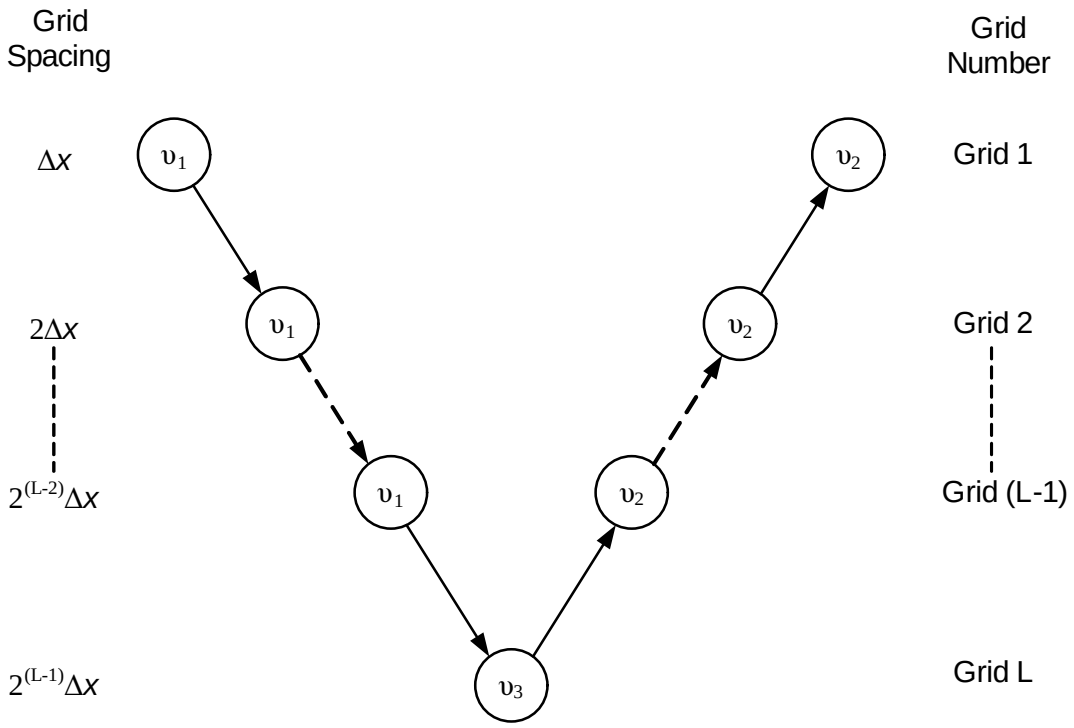

Figure 4.4: Schedule of grids for a *V-cycle*

4.4 Algorithms Implemented

For the purpose of this study, the two known movements applied to hip joint models, steady-state rotation and the transient walking cycle will be studied. The specified inputs and the desired outputs for the two simulations are different, as a result, two different algorithms are required.

The simpler model is that of pure angular motion and the corresponding algorithm can be seen in figure 4.5.

The algorithm for the angular motion is a simple procedure where the system is fully specified when the angular velocity and eccentricity ratio are chosen. At this point the pressure distribution needs to be solved using equation 4.1, and if necessary, the pressure constraint may be implemented whereby the pressure is restricted by not being allowed to be negative. The resultant forces are calculated by equations 4.2, 4.3 and 4.4.

The algorithm used for the transient model which simulates the walking cycle is shown in figure 4.6.

This algorithm has one degree of freedom whereby it is necessary to choose an eccentricity ratio to solve for the pressure distribution and resultant forces. This was done by choosing the initial eccentricity ratio of the current time step, to have a value equal to the solved eccentricity ratio value of the previous time step. The eccentricity ratio was changed by a built-in MATLAB function which was used to minimise the squared error between the input and calculated load.

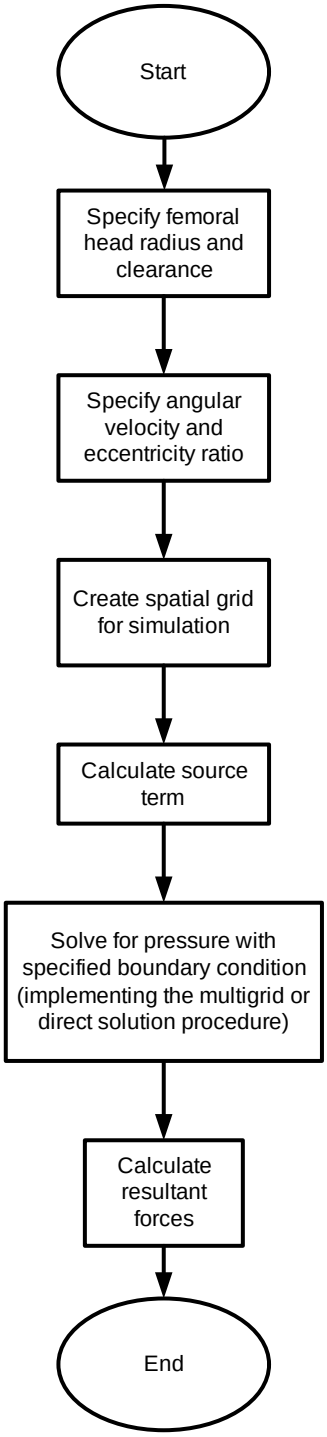

Figure 4.5: Algorithm implemented to obtain results for the angular motion simulation

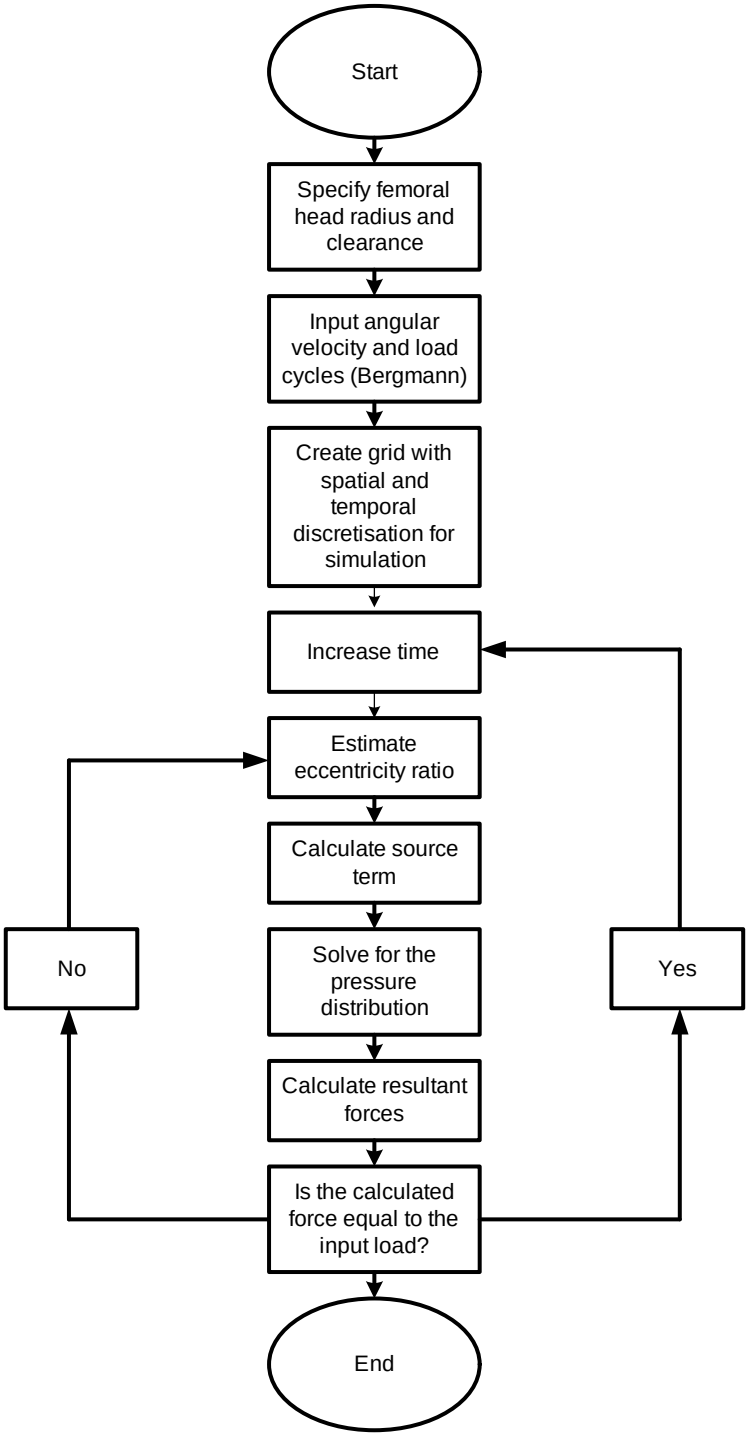

Figure 4.6: Algorithm implemented to obtain results for the transient walking cycle simulation

CHAPTER 5

Planning and Methods

The main purpose of this study is to investigate the hydrodynamic lubrication nature of an artificial hip joint. With this in mind, the necessary planning and methods of implementation were considered.

5.1 Planning

Having developed a model in chapter 3 that describes hydrodynamic lubrication in an artificial hip joint (with certain assumptions), further investigation was required to obtain an understanding of how different aspects of the model affect the lubrication.

This would be conducted on a steady-state model. This model was chosen as it gives insight to the lubrication by eliminating the transient behaviour, thereby also simplifying the solution procedure and subsequent analysis. Steady-state analysis is not uncommon, an example thereof is that conducted by Jalali-Vahid et al. (2001).

The main aspects considered in the investigation were the:

- boundary conditions
- eccentricity ratio
- intermediary pressure constraint

From the literature review on modelling, the boundary condition implemented was the Dirichlet boundary condition ($p = 0$) (Jagatia et al., 2001), or in tandem with the Neumann boundary condition ($\frac{\partial p}{\partial \phi} = \frac{\partial p}{\partial \theta} = 0$) (Jalali-Vahid et al., 2001). As a result, both conditions were investigated separately.

In the analyses made by Jin & Dowson (1999), it was shown that the eccentricity ratio changed during the transient analysis. Similar behaviour would be expected in the

transient analysis to be conducted on the model developed for this study. Due to this expectancy, the effect of the eccentricity ratio on the system was duly investigated and varied accordingly.

The final concept to be investigated by applying the steady-state model is the implementation of an intermediary constraint, namely that the interior nodes must be positive (Jin & Dowson, 1999). The implementation of this constraint is somewhat perplexing, but its effect will be investigated.

The outputs from all the investigations will be the pressure distribution and hydrodynamic load or resultant supportive force.

For the transient analysis, it is required to obtain realistic inputs. This is done by implementing the Bergmann walking cycle which was obtained from an investigation of the load that a hip joint would experience and which was verified by telemetry from artificial hip joint patients. Another good reason to make use of this cycle is because it also forms part of the ISO standard $14242-1$ regarding wear rate testing. The model, which does not incorporate elastic deformation, is not expected to be able to withstand high loads; as a result the Bergmann profile is that with a moderate/low maximum load. The input used is that implemented by Pieterse (2005). This is a Bergmann load cycle

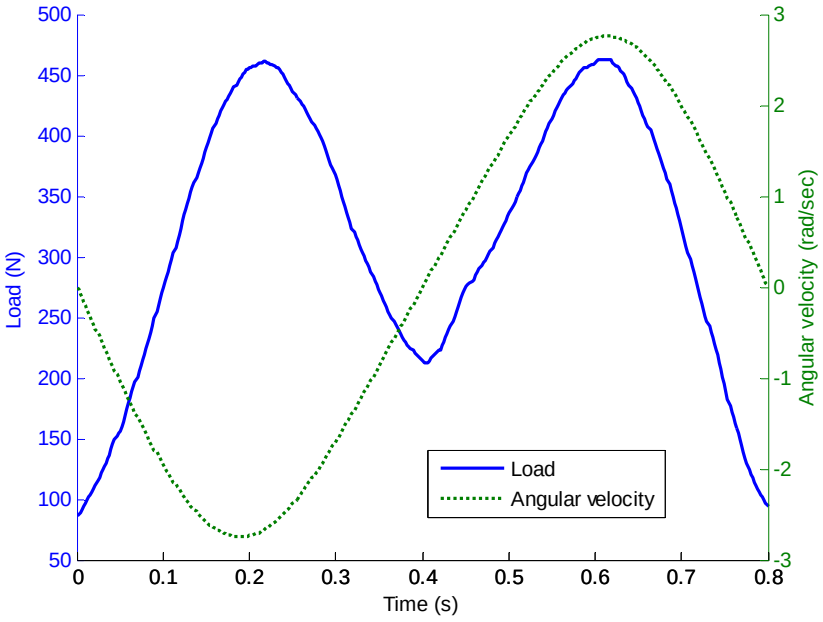

Figure 5.1: Load cycle and angular velocity used for transient simulation

with a frequency of 1.25Hz.

For consistency, the physical parameters used were also those implemented by Pieterse (2005) and can be seen in table 5.1. The viscosity used is an intermediate value between 0.0025 and $0.0009 Pa.s$ which have been used as estimates to peri-prosthetic synovial fluid.

Parameter	Value
Femoral head radius	$28mm$
Radial clearance	$0.25mm$
Viscosity	$0.0015Pa.s$

Table 5.1: Parameters for transient walking cycle simulation

5.2 Methods

The steady-state analyses were all conducted by implementing the algorithm shown in figure 4.5.

The boundary conditions were directly imposed on the grid generated. If the grid generated has n nodes in the ϕ direction and m nodes in the θ direction then the Dirichlet boundary condition is simple to implement and is specified as

$$p_{i,m} = p_{n,j} = p_{i,1} = p_{1,j} = 0$$

On the other hand, the Neumann boundary condition is approximated by

$$\begin{aligned} p_{i,1} &= \frac{p_{1,2} + \cdots + p_{n,2}}{n} \\ p_{i,m} &= \frac{p_{1,m-1} + \cdots + p_{n,m-1}}{n} \\ p_{1,j} &= \frac{p_{1,j+n} + p_{1,j-n} + p_{2,j}}{3} \\ p_{n,j} &= \frac{p_{1,j+n} + p_{1,j-n} - p_{2,j}}{3} \end{aligned}$$

For the analyses based on varying eccentricity ratio, the method of implementation is relatively simple. Since the eccentricity ratio is an input in the steady-state algorithm, the value was merely specified.

To implement the intermediary constraint, the Gauss-Seidel local relaxation technique was implemented. During the calculation of the value at every grid point the value was determined and adjusted if necessary. This was done simply by using Boolean logic.

$$if \qquad p_{ij} \leq 0 \qquad then \qquad p_{ij} = 0$$

The transient analyses were conducted by implementing the transient algorithm shown in figure 4.6. The specified input is the Bergmann cycle and iterations are executed for the eccentricity ratios until the calculated load is that imposed by the Bergmann cycle.

CHAPTER 6

Results and Discussion

This chapter provides analysis into the results obtained from simulations conducted on the model developed in chapter 3. The purpose of this study was to investigate the hydrodynamic nature of lubrication for the transient walking cycle in the hip joint; this was done by implementing the Bergmann cycle as an input in the model.

Investigations into both steady-state angular rotation and transient walking cycle models were conducted. The steady-state model was formulated and simulated to develop an understanding of how different aspects of the model affect the outputs, pressure and force (hydrodynamic load). The transient model was used to investigate the lubrication of the hip joint under realistic walking conditions. Parameters in the model were altered to ascertain the effect on lubrication. The effect of grid size for all the simulations was also investigated.

6.1 Grid Refinement

An important aspect when solving partial differential equations using numerical techniques, is the accuracy and sensitivity of the problem to a change in the grid size. Using a coarse grid can result in an inaccurate representation of the solution to the problem obtained. Finer grids often result in better solutions; however finer grids may also cause certain other numerical problems. Finer grids can lead to numerically unstable solutions, so the choice of grid is also dependent on the discretisation and numerical technique implemented in obtaining the solution.

Based on the discretisation scheme chosen, a finite difference grid is generated. The grid is equally spaced in both directions. The grid was also developed to have an equal amount of nodes in both the ϕ and θ directions.

The model developed calculates the resultant force, or hydrodynamic load, that can be supported for every specified eccentricity ratio and angular velocity. To try and develop a method whereby it is possible to evaluate the sensitivity of the model, a value will be calculated to give an indication of the force, hydrodynamic load, at that particular point. Due to the inputs of the model chosen, the force itself could not be used as the resultant force was negligible; instead equation 6.1 and 6.2 were used.

$$ISP = R^2 \int_0^{\pi} \int_0^{\pi} p^2 \partial\theta\partial\phi \tag{6.1}$$

$$IAP = R^2 \int_0^{\pi} \int_0^{\pi} |p| \partial\theta\partial\phi \tag{6.2}$$

ISP is the integral of the square of the pressure and IAP is the integral of the absolute pressure. Equations 6.1 and 6.2 are discretised and calculated as follows

$$ISP = \frac{R^2}{4} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{i,j}^2 + p_{i+1,j}^2 + p_{i,j+1}^2 + p_{i+1,j+1}^2 \right)$$

$$IAP = \frac{R^2}{4} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(|p_{i,j}| + |p_{i+1,j}| + |p_{i,j+1}| + |p_{i+1,j+1}| \right)$$

The convergence criteria, grid size which leads to no appreciable numerical improvement, were based on the values of these integrals and were applied to a steady state model with a specified angular velocity and Dirichlet boundary condition. The constraint of allowing only positive pressures was not enforced. The femoral head radius was specified as $28mm$ and the clearance was specified as $0.25mm$. A convergence graph (integral values), for specified angular velocity ($\omega_{x_3} = 2\,rad/s$) and eccentricity ratio ($\xi = 0.99$) can be seen in figure 6.1.

A value of $\xi = 0.99$ was chosen as this would result in the highest pressures possible for this model; $\xi = 0.99$ was chosen as the uppermost limit so as to take into account surface roughness, this value was chosen as it would easily illustrate the difference in results obtained. As can be seen in figure 6.1, it is not the value of the integral that gives an indication of what grid size to use, but more correctly, the variation of the integral with different grid nodes.

The IAP criterion seem to show negligible variation after 129×129 node matrix had been used. If the purpose of the model tested is a qualitative result, then a matrix in the region of 100×100 should be sufficient. However if the value of the calculated load that can be supported is also an important aspect to consider, the ISP criterion would be more indicative. The ISP is much more sensitive to variation, and will lead to more

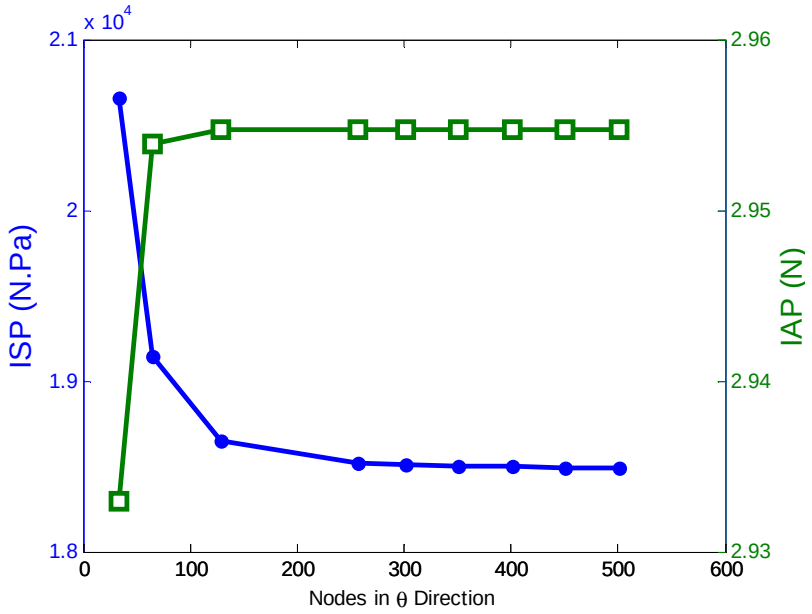

Figure 6.1: Variation in ISP and IAP with different grid sizes for $\omega_{x_3} = 2$ and $\xi = 0.99$

informative choices when needing to calculate forces greater than order 1. This would most likely be in the region from 150×150 to 250×250 nodes, as anything above this will require extensive computer power and storage without much increase in accuracy.

6.2 Angular Motion

A provisional study was done by using only angular motion. This will provide better understanding for the transient cycle which will follow later in the chapter. It has also been reported earlier, that in some cases, only the angular motion has been modelled; this will provide insight into what contribution angular motion has on spherical bearing systems such as a hip joint.

The following aspects will be investigated:

- different boundary conditions
- the non-negative pressure constraint
- different eccentricity ratios

For the modelling of strictly angular motion, equation 3.24 reduces to equation 6.3.

$$\frac{\partial}{\partial\theta}\left(h^3 \sin\theta \frac{\partial p}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(h^3 \frac{\partial p}{\partial\phi}\right) = 6\mu R^2 \sin^2\theta \omega_{x_3} \frac{\partial h}{\partial\phi} \tag{6.3}$$

6.2.1 Boundary Conditions

The boundary of the system are the points that lie in the $X_1 - X_3$ plane shown in figure 3.1 ($x_1 = x$, $x_2 = y$ and $x_3 = z$ in conventional notation). The conditions implemented to obtain, or specify, the values of the nodes which lie on the boundary were investigated.

In the papers reviewed, the typical boundary constraint applied was the Dirichlet boundary condition ($p = 0$). This implies that there is a constant gauge pressure on the system boundary. This also helps to define the pressure boundary similar to that of a pressure "reservoir", whereby equilibrium can be reached.

It was decided to also investigate the use of the Neumann boundary condition ($\frac{\partial p}{\partial \theta} = \frac{\partial p}{\partial \phi} = 0$). This implies that at the boundary, no pressure induced flow occurs, and that the flow at the boundary is caused by the movement of the femoral component via the no-slip boundary condition. From the review of current literature of hip joint modelling, this is incorporated in some way with the Dirichlet boundary condition.

The model was tested with parameters shown in table 6.1.

Parameter	Value
Femoral head radius	$28mm$
Radial clearance	$0.25mm$
Viscosity	$0.0015Pa.s$
ω_{x_3}	$2rad.s^{-1}$
ξ	0.75
Grid size	129×129

Table 6.1: Parameters for steady-state angular motion simulation

Dirichlet Boundary Condition

The model was solved by applying the numerical procedure described in figure 4.5. The resultant pressure distribution and contour plot are shown in figures 6.2 and 6.3.

From the pressure distribution and contour plot, it can clearly be seen that the pressures calculated are clearly anti-symmetric about the X_3 axis. Due to the anti-symmetry, it is unsurprising that the calculated resultant forces in the Cartesian directions are negligible.

Since ω_{x_3} is positive, the motion of the femoral head component would be from the positive X_1 axis to the negative X_1 axis. Given that the flow of fluid would also be expected to be in the same direction as the imposed motion of the femoral head, some interesting aspects have been encountered.

The most obvious phenomena are the presence of a source and sink as is shown in figure 6.4, where the arrows show direction of flow. The presence of a source and sink is interesting; this implies that there is flow outward from a certain point within the control volume and flow towards a point in the control volume which is not at the control volume

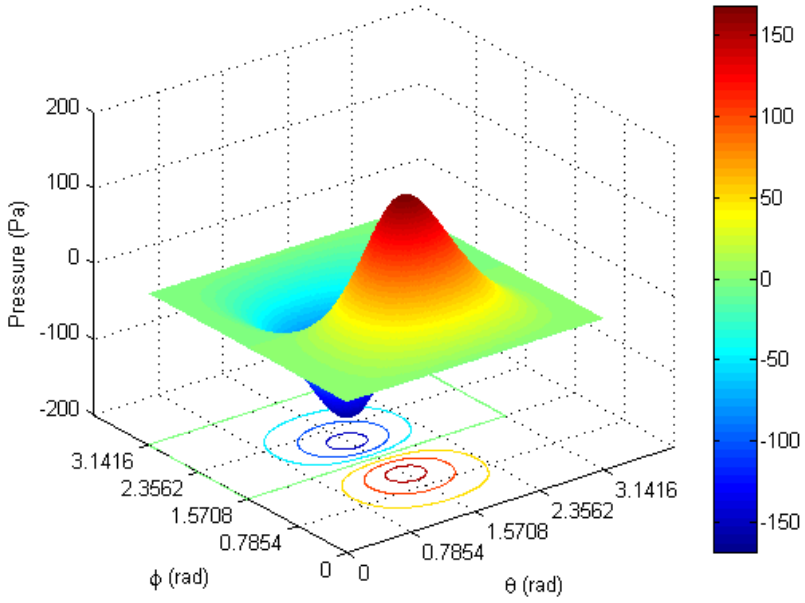

Figure 6.2: Pressure distribution when imposing the Dirichlet boundary condition

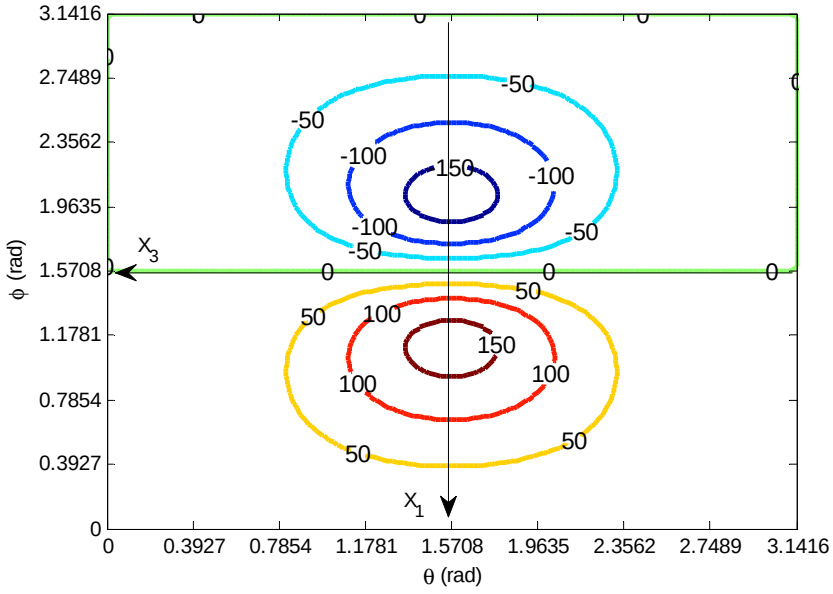

Figure 6.3: Pressure contour when imposing the Dirichlet boundary condition

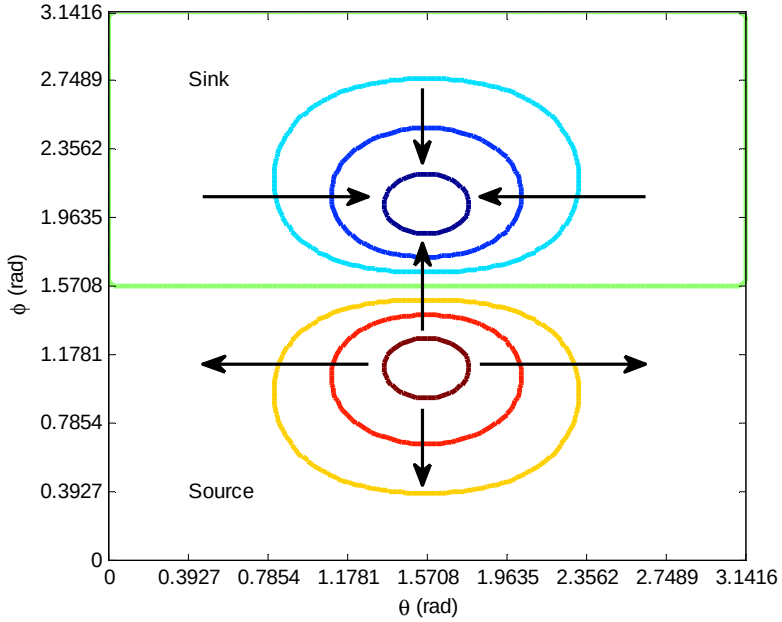

Figure 6.4: Source and sink from angular rotation with Dirichlet boundary condition

boundary. The areas of positive and negative pressures are unsurprising, but that the pressure gradient along the X_1 axis does change sign is a surprise. This development is explained from the imposed boundary condition. This necessitates that the high interior pressure must be reduced the closer the node is to the boundary, similarly the low interior pressure must be increased the closer the node is to the boundary. It should also be noted that flow is also imposed from movement of the femoral component via the no slip boundary condition; this would counteract the direction of flow as suggested by the pressure profile; therefore the pressure calculated represents only the force that is required to cause the flow of the fluid past the most constricted points between the cup and femoral head.

From an external point of view, the $p = 0$ boundary condition seems reasonable, despite the presence of the source and sink.

Neumann Boundary Condition

Due to the results obtained from applying the Dirichlet boundary condition, it was decided to investigate applying the Neumann boundary condition by implementing the numerical procedure described in figure 4.5.

The resultant pressure distribution and contour plot are shown in figures 6.5 and 6.6. Similar to the results obtained from applying the Dirichlet boundary condition, these results also seem to be almost anti-symmetric about the X_3 axis and imposing no load carrying capability. The profile itself seems to be more realistic as fluid will predominantly flow in the negative X_1 direction.

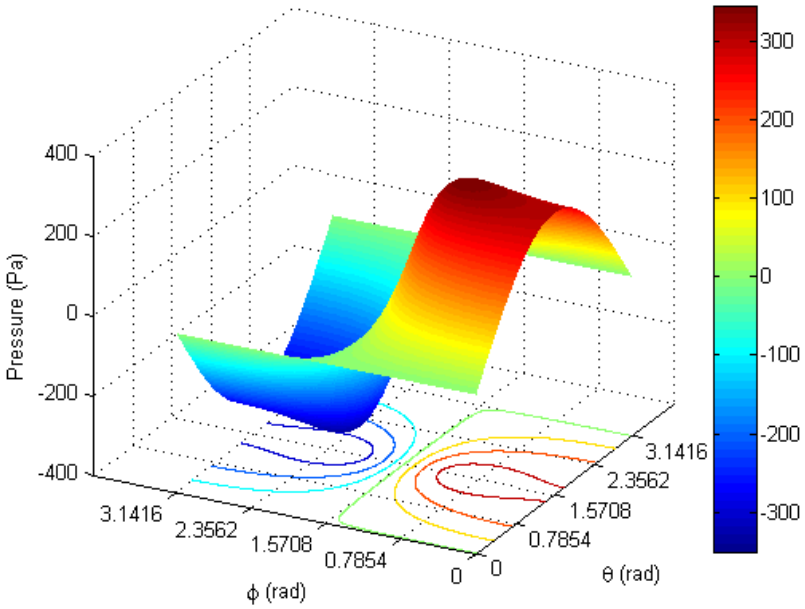

Figure 6.5: Pressure distribution when imposing the Neumann boundary condition

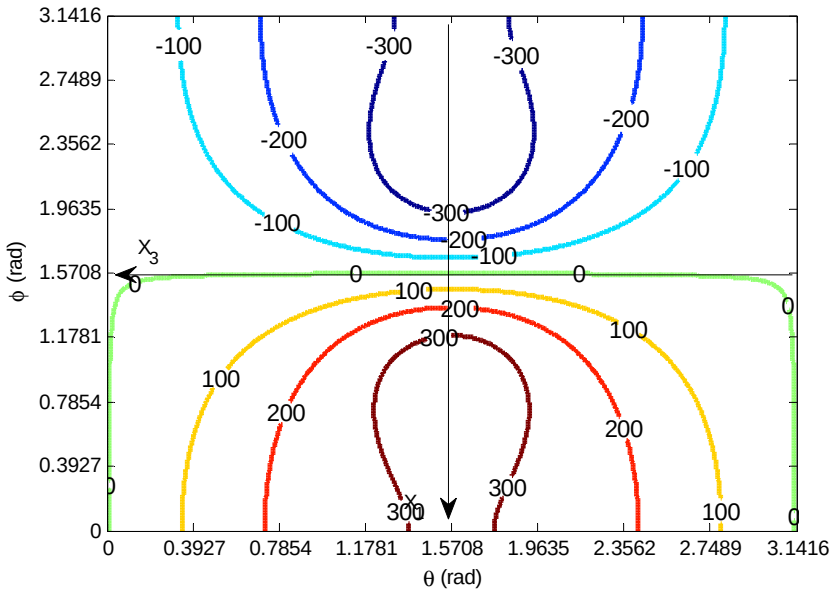

Figure 6.6: Pressure contour when imposing the Neumann boundary condition

This pressure profile seems to be more representative of the characteristics of the source term calculated (RHS of equation 6.3). The source term profile is shown in figure 6.7.

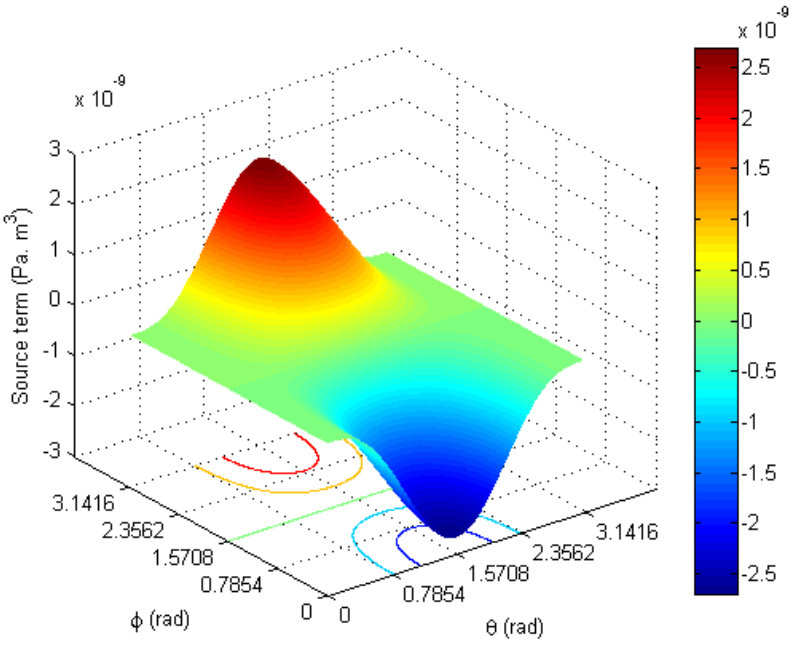

Figure 6.7: Profile of the source term for angular motion

The profile of the source term should be the same as the second order partial differential equation, and as a result some understanding can be gathered with respect to the pressure. Given that the source term is symmetrical about the X_3 axis, figure 6.8 will help in understanding the effect of the boundary conditions on the source term and the resultant pressure.

Applying the different boundary conditions forces the derivative and function to have certain characteristics. As can be seen in figure 6.8, the Dirichlet boundary condition directly imposes the necessity to have a local minimum and maximum in the resultant pressure profile. Similarly the Neumann boundary condition forced the minimum and maximum to be at the boundary edge. However the modelled second order partial differential equation contains information about the film thickness as well, therefore these statements will be valid when the change in film thickness between the surfaces is negligible. This effect is much more appreciable when applying the Neumann boundary condition.

6.2.2 Pressure Constraint

The pressure constraint has been applied to some lubrication problems, this constraint necessitates that the pressure be positive. The pressure being solved is the gauge pressure,

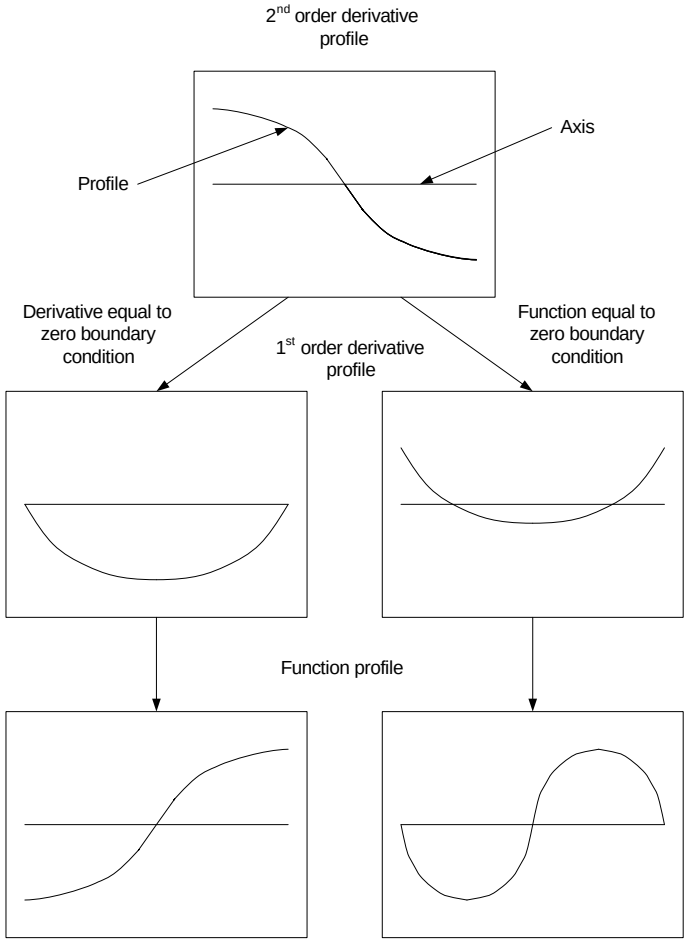

Figure 6.8: Effect of the profile of the source term for angular motion

thus the pressure need not be positive, merely an attainable pressure. Having already obtained the direct solution to the angular rotation problem, an indirect method (shown in figure 4.5) will be used whilst applying the intermediary constraint. Simulations applying the Dirichlet boundary condition were run; the pressure distribution and contour from the resultant simulation can be seen in figures 6.9 and 6.10.

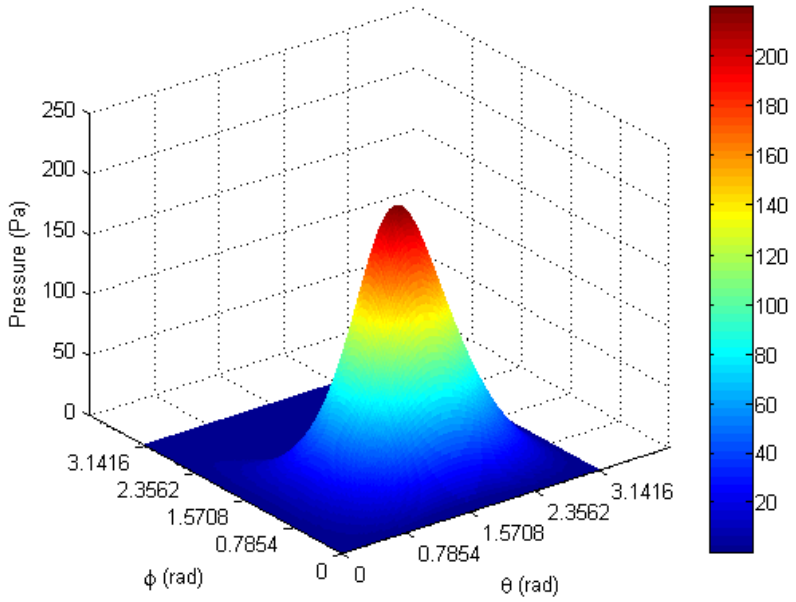

Figure 6.9: Pressure distribution when imposing the Dirichlet boundary condition and pressure constraint

The resulting pressure profiles differ significantly as compared to when the pressure constraint was not imposed. The shape of the profile in figure 6.9 is purely convex. There is a maximum pressure in the interior of the grid, in a similar region to where the maximum pressure is of the profile in figure 6.2. The maximum pressure is closer to the centre point of the system as a result of the positive pressures. A small positive force, supportive load, was also calculated. The magnitude of the force could be more significant if this was imposed on system that results in much higher pressures (as in the case of high eccentricity ratios).

The validity of the profile itself, could be doubtful, as the residual calculated had not been minimised significantly. This is evident in figure 6.11. The residual reaches an asymptote, and no significant improvement on the pressure profile is attained. Given that the source term values are of order (-9) (in the range of 10^{-9}) and the calculated Euclidean norm of the residual is of order (-7) (in the range of 10^{-7}), no significant reduction in the calculated norm is achieved to assume a feasible solution.

Applying the intermediary constraint, does not solve the problem at hand, it merely approximates a solution and is not entirely representative of the given problem. Using the

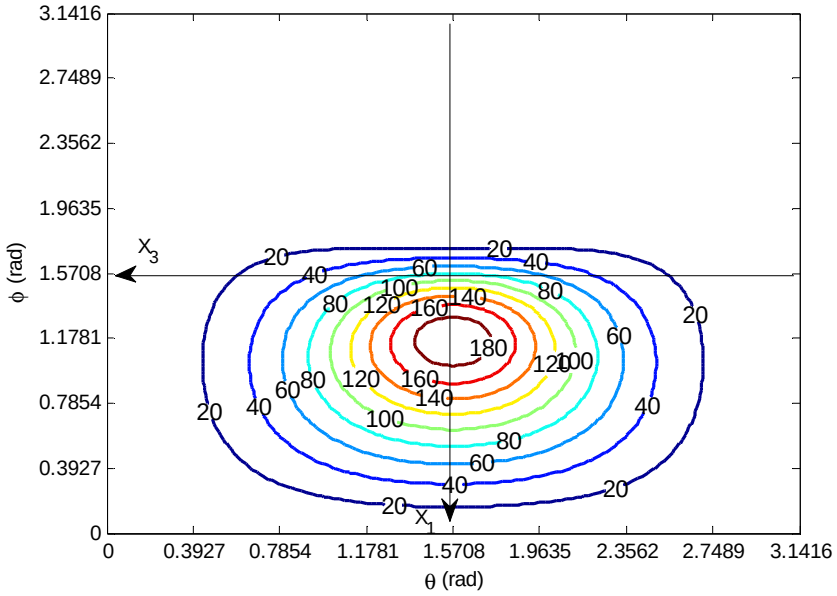

Figure 6.10: Pressure contour when imposing the Dirichlet boundary condition and pressure constraint

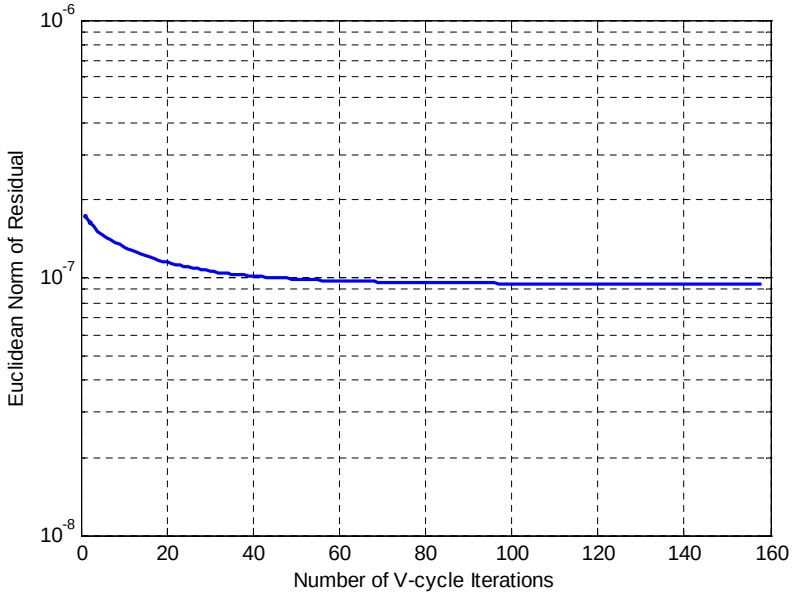

Figure 6.11: Residual norm variance during iteration process

direct method, it was observed that negative pressures were encountered, however these values were not low enough so as to have infeasible solutions from a realistic perspective.

As a result of applying the pressure constraint, the numerical solution to the LHS of equation 6.3 will not equal the RHS of equation 6.3; because of this, the use of the pressure constraint on the model developed is, somewhat debatable. A more appropriate solution would be a solution with realistic pressure values that solves the model at hand; if this is not possible, the model should be altered, not restricted with application of constraints.

6.2.3 Eccentricity Ratio

The effect of the eccentricity ratio on the pressure profiles will also be investigated. As the eccentricity ratio increases, the film thickness decreases in the specified direction. For this purpose, only the eccentricity ratio in the X_2 direction will be varied as this will be the major axis of the imposed force in the walking simulation.

The variance of the pressure will be quantified via the maximum and minimum pressures calculated. The values obtained are shown in table 6.2.

eccentricity	Dirichlet		Neumann	
	Min Pressure	Max Pressure	Min Pressure	Max Pressure
0.6	-72.1786	72.1786	-188.5937	189.6784
0.7	-123.1916	123.1916	-279.2068	276.9066
0.75	-168.0942	168.0942	-341.5518	345.3822
0.8	-241.4179	241.4179	-439.9598	435.32
0.9	-697.7073	697.7073	-929.6144	933.5327
0.99	-21316.9696	21316.9696	-21504.6302	21502.2624

Table 6.2: Change in pressure with eccentricity ratio (Min = Minimum, Max = Maximum)(All pressures in Pa)

As is observed in table 6.2, the magnitude of the maximum and minimum pressures increases as the eccentricity ratio increases. This is not surprising as the RHS of equation 6.3 is a function of the derivative of the film thickness; as the eccentricity ratio increases, the film thickness decreases on the interior nodes. At the centre of the cup, the film thickness is a minimum, and higher pressures are required to impose the flow of fluid. The higher the eccentricity ratio, the closer the peaks of minimum and maximum pressure are to the centre of the cup. This can clearly be seen by comparing the results from figure 6.2, for an eccentricity ratio of 0.75 and 6.12, for an eccentricity ratio of 0.99, whilst applying the Dirichlet boundary condition.

The pressure profiles when applying the Neumann boundary condition and increasing eccentricity ratios, are also interesting. The pressure profiles with eccentricity ratios of 0.9 and 0.99 can be seen in figures 6.13 and 6.14.

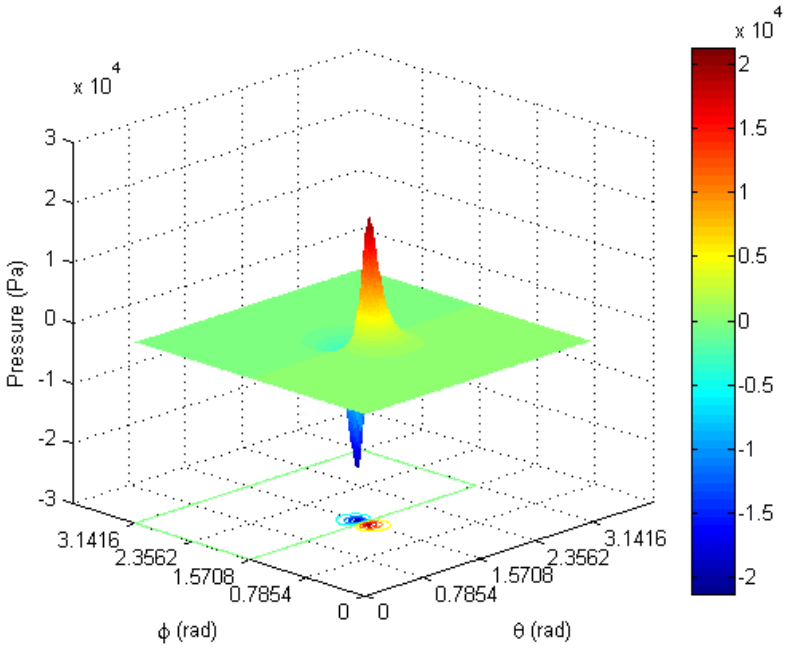

Figure 6.12: Pressure distribution when imposing the Dirichlet boundary condition with an eccentricity ratio of 0.99

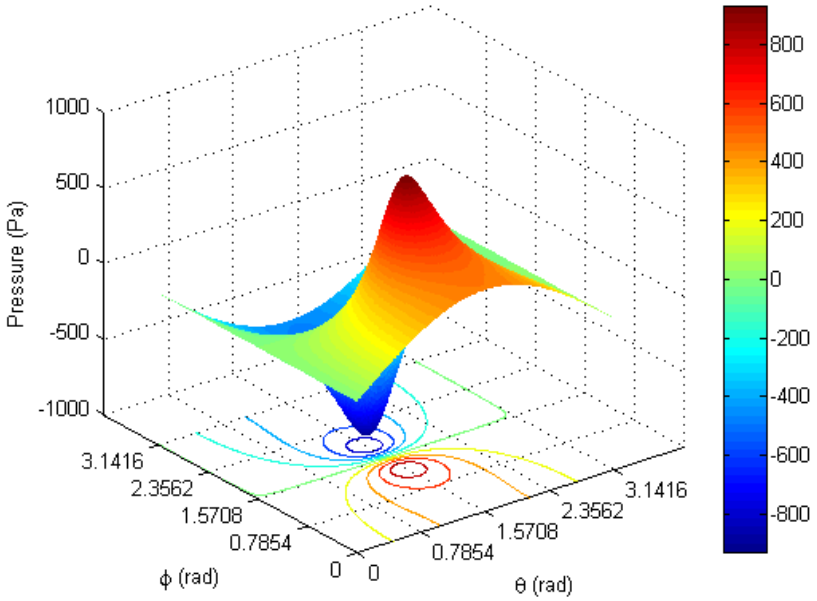

Figure 6.13: Pressure distribution when imposing the Neumann boundary condition with an eccentricity ratio of 0.9

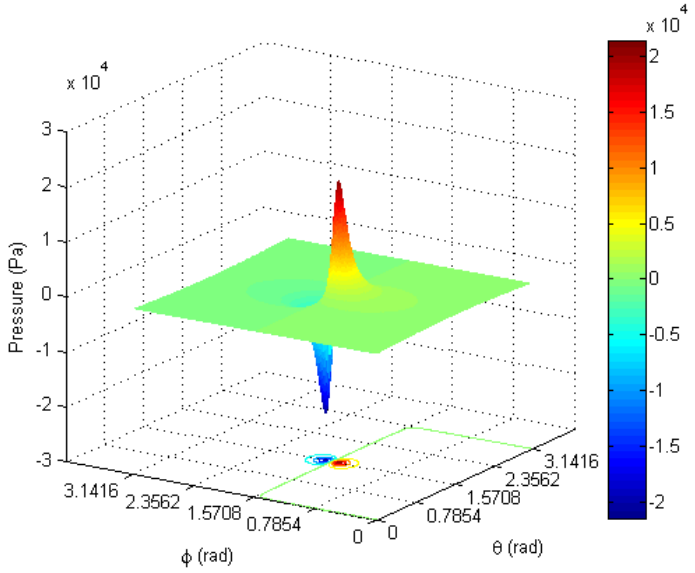

Figure 6.14: Pressure distribution when imposing the Neumann boundary condition with an eccentricity ratio of 0.99

As can be seen from figures 6.5, 6.13 and 6.14, as the eccentricity ratios increase, the peaks of the pressure profile become more prominent, similar to that observed when applying the Dirichlet boundary condition.

As the eccentricity ratio increases, the spatial derivative of the film thickness becomes more prominent, and dominates at extremely high ratios. As a result of this, the pressure profiles start to resemble the profiles obtained when using the Dirichlet condition.

This implies that at high loadings, an insignificant pressure differential occurs at the boundaries and that the pressure itself is almost negligible in comparison to the extreme pressures encountered.

6.3 Transient Model

The transient model used solves equation 3.24 via the numerical procedure shown in figure 4.6. The model utilised had an input load and angular velocity as shown in figure 5.1.

Analysing equation 3.24, the angular rotation and spatial and temporal derivatives both contribute to the solution. Provisional analysis has resulted in no load carrying capability of the angular rotation of the model. If the model is assumed to be linear functions of the temporal and spatial derivatives, the solutions could be superimposed. This would lead us to believe that the temporal derivative is the load carrying term in the equation; as a result, only positive derivatives would allow for positive loads, and therefore the eccentricity ratio would continually increase until the load cycle is completed or until there is contact between the femoral head and acetabular cup; in the simulation, *time to*

contact or *touch* was defined at either $\xi_{x_1} = 0, \xi_{x_2} = 0$ or $\xi_{x_3} = 0$.

For this simulation, the physical parameters shown in table 5.1 were used and the Dirichlet boundary condition was implemented. The initial eccentricity ratios, $\xi_{x_1}, \xi_{x_2}, \xi_{x_3}$, we all set to zero for this simulation. Ideally, if no touching had occurred, an iterative procedure would have been employed to determine the true initial conditions; this would be done by setting the initial conditions of the eccentricity ratio to those at the end of the previous cycle modelled.

6.3.1 Grid Refinement with Time

From the grid refinement discussed in section 6.1, only spatial grids in the region between 150×150 and 250×250 nodes were considered. Considering this, choosing the grid for this model, was based on the time required for the surfaces to touch. Table 6.3 shows the time at which contact occurs for the different grid sizes used.

Spatial grid size	Temporal nodes	Time of contact (s)
151×151	189	0.0298
151×151	1600	0.0205
151×151	10000	0.0189
151×151	16000	0.0188
151×151	20000	0.0188
151×151	24000	0.0187
151×151	28000	0.0187
151×151	32000	0.0187
181×181	189	0.0298
181×181	1600	0.0205
181×181	10000	0.0189
181×181	16000	0.0188
181×181	20000	0.0187
181×181	24000	0.0187
181×181	28000	0.0187
181×181	32000	0.0187
241×241	189	0.0298
241×241	1600	0.0205
241×241	10000	0.0189
241×241	16000	0.0188
241×241	20000	0.0187
241×241	24000	0.0187
241×241	28000	0.0187
241×241	32000	0.0187
271×271	189	0.0298

Table 6.3: Variability of time of contact with different spatial grids and number of temporal nodes

For the grid refinement of this simulation, it was decided that four decimal accuracy

would be more than adequate due to the short time required before contact. It is noticeable that at equal temporal discretisation, and with different grid sizes, the time required to reach contact was similar. This shows that improved spatial discretisation does not affect the time required to reach contact. The number of temporal nodes chosen were for the 0.8s cycle, which limited the accuracy of the transient model; as a result, very few temporal nodes were used during the simulation before contact occurred.

The number of temporal nodes was then increased, to improve the accuracy. This was vilified by the decrease in the contact time. A small difference is observed between contact time of the simulations with 10000 and 20000 temporal nodes. During the simulation using a 241×241 grid with 20000 temporal nodes, approximately 500 temporal nodes had associated non-zero values and only for these values had computer memory been allocated. As a result, the chosen grid had 241×241 spatial nodes and 20000 temporal nodes.

6.3.2 Model Results

The main outputs from the model are the calculated forces (supported load) and eccentricity ratios which result in the pressure distributions that correspond to the applied load. These outputs can be seen in figures 6.15 and 6.16.

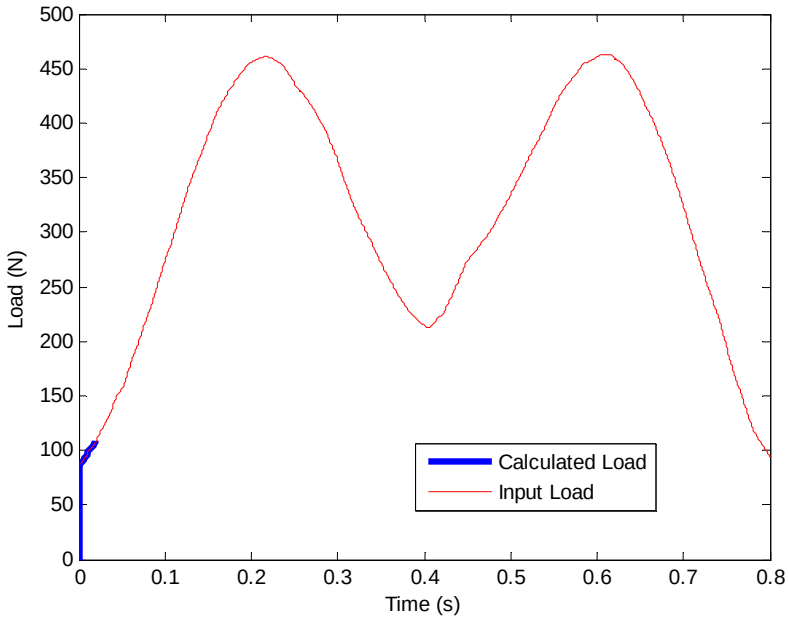

Figure 6.15: Comparison of input load and calculated load from simulation

From figure 6.15 it is clearly visible that the lubricant cannot withstand a significant load. The articulating surfaces touch after approximately $0.0187s$, where the load is approximately $107N$. This shows that the fluid cannot develop a high enough pressure distribution to counteract the load from the walking cycle.

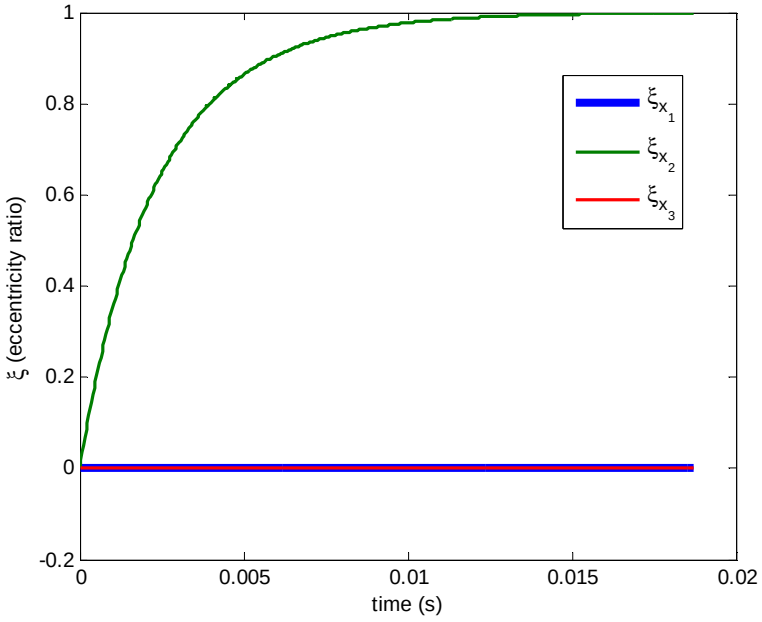

Figure 6.16: Change of eccentricity ratio during simulation

It is also evident from figure 6.16 that the eccentricity ratio continuously increases until the point of contact. This point being the time at which failure is measured. This model shows that asperity contact will almost be a certainty and that fluid is unable to withstand the pressures required, this leads to the belief that such a low viscosity fluid is unable to maintain the lubrication of the joint in the hydrodynamic regime without elastic deformation. It should be taken into consideration that this model is not valid in the region of extreme pressures, as the higher pressures would cause deformation of the surfaces (due to low geometrical conformity) and this would allow for better load carrying capability; an elasto-hydrodynamic model would be a valid alternative in this regime.

The continuously increasing eccentricity ratio is verification that only decreasing film thickness can hold a load. This suggests that at some point in multiple cycles, contact will always be reached due to increasing nature of the eccentricity ratio. From a more practical perspective however, after a single walking cycle, the muscles lower the femoral component of the leg during the swing phase of the cycle, and therefore the cycle begins from the original position again.

What is also visible in figure 6.16, is that there is no change in eccentricity ratios in the x_1 and x_3 directions. This is caused by negligible or no resultant forces being generated in the corresponding directions. The most obvious reason for this is the orientation of the acetabular component in the model. The acetabular cup lies in the horizontal plane as opposed to the anatomically correct 45°.

Switching to a more numerical perspective, as discussed before, the source term should be the same shape as the second order partial differential equation; the shape of the source

term for the transient simulation can be seen in figure 6.17.

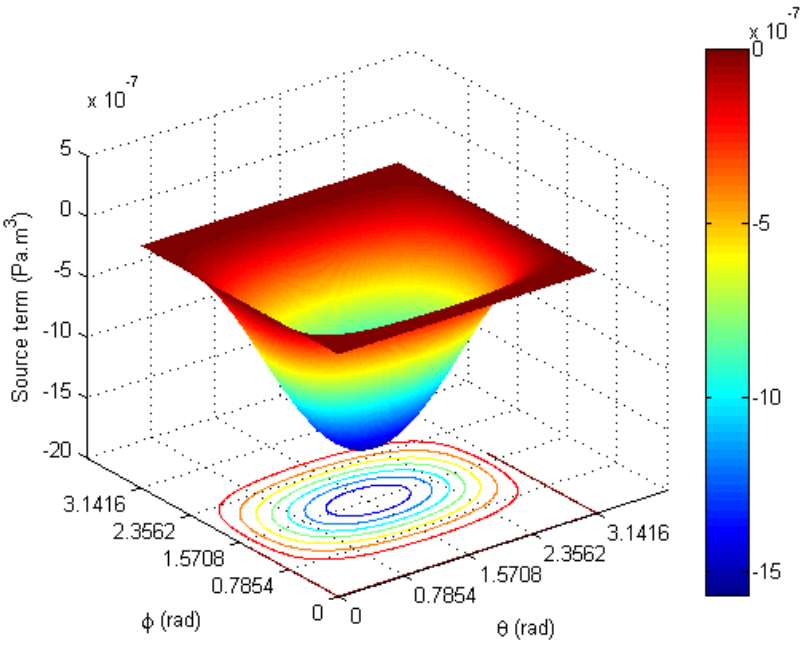

Figure 6.17: Profile of the source term for transient motion

The source term in figure 6.17, is for the first time increment in the walking cycle. It is quite obvious that as compared with the steady angular motion, the source term for transient movement is at least 100 times larger; it is also completely negative, therefore resulting in only positive pressures being calculated. One other important aspect to observe is the total lack of similarity in the shape of the source terms. This merely confirms the fact that the angular motion, which is included in the source term, has almost no contribution to the shape of the pressure distribution and its corresponding values.

The resulting pressure distribution is shown in figure 6.18. In a manner similar to that explained in section 6.2, the shape of the source term is opposite in shape to the pressure distribution; the same applies to the source term and pressure distribution for the transient simulation. The magnitude of the pressures are significantly higher, and the hydrodynamic pressure distribution in figure 6.18 can withstand a load of approximately $86N$. It is evident that from such a distribution it is not needed to implement the cavitation pressure constraint whereby no negative pressure values are allowed, as the calculated pressures are all positive. This also allows for use of the quicker direct solution procedure for the current model.

During the cycle, as the required load to withstand becomes higher, the eccentricity ratios increase and the shape and magnitudes of the pressure distributions change as can be seen in figures 6.19 and 6.20. The pressure distributions become more concentrated at the centre of the cup and the magnitudes increase as the eccentricity ratio increases.

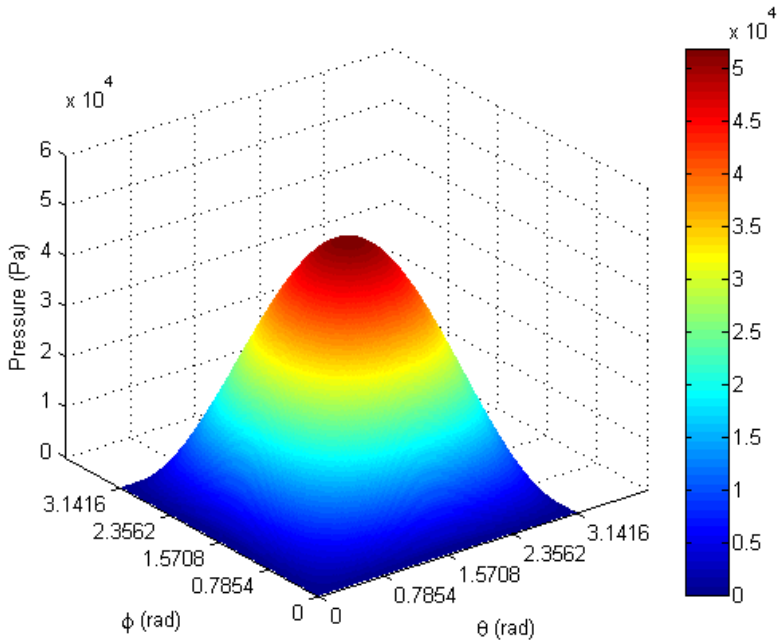

Figure 6.18: Pressure distribution for the first time step in the simulation

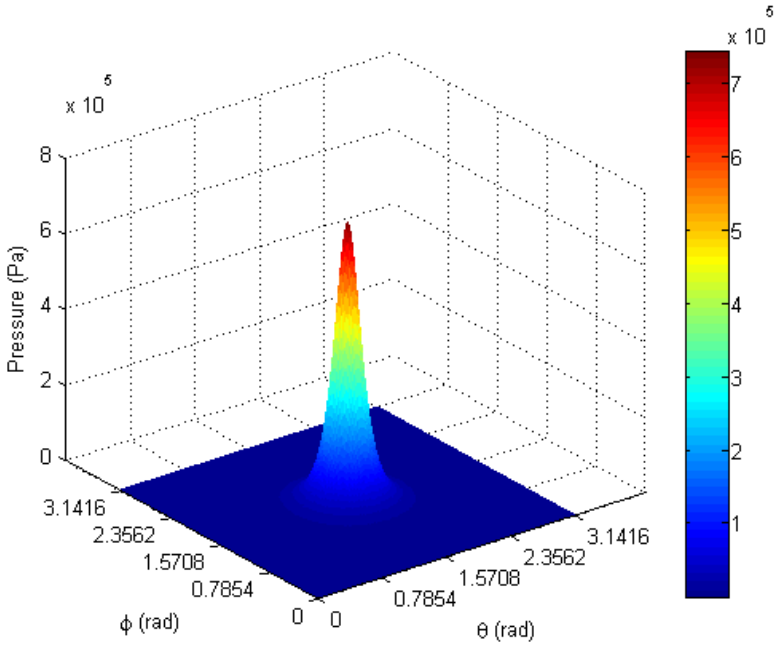

Figure 6.19: Pressure distribution for time = 0.0093s

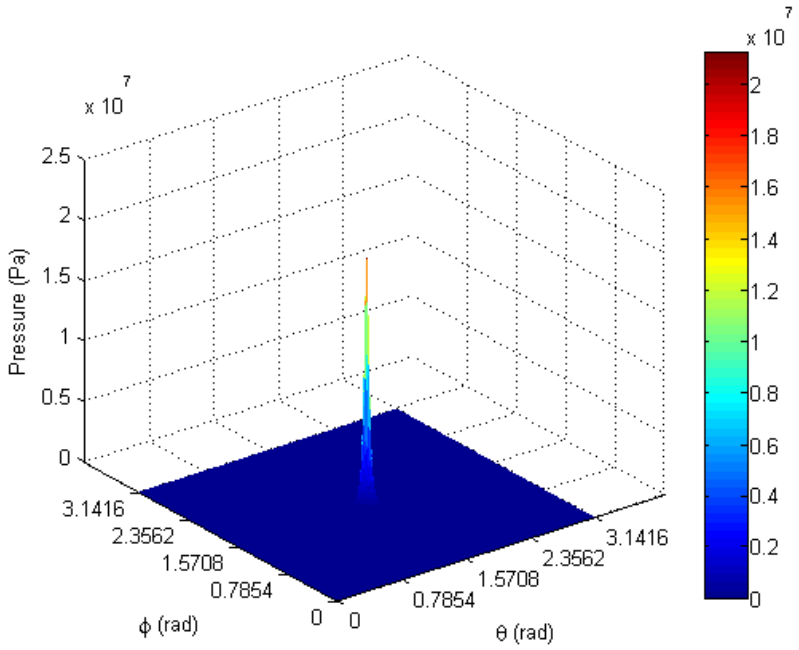

Figure 6.20: Pressure distribution for time = 0.0187s

This happens as a result of the surfaces coming closer and closer to contact. This explains the concentrated nature of the profiles but not the magnitudes. It is understood that since no elastic deformation is being modelled, the rate of change of the film thickness with time is directly related to the rate of change of the eccentricity ratio with time; but the rate of change of the eccentricity ratio is decreasing as the simulation proceeds, and it would be expected that as a result, the source term would be smaller in magnitude and therefore a pressure profile lower in magnitude would be expected. This is not the case since the expression relating film thickness to eccentricity ratio, equation 3.26, shows that the film thickness is most affected at the centre of the cup; and any small difference is exaggerated the closer the surfaces are to each other and hence the derivative of film thickness with time is just as exaggerated as the simulation proceeds.

The pressures generated in figures 6.19 and 6.20 are also extremely high, this would almost certainly be high enough to cause significant deformation of the acetabular cup if the material used is UHMWPE. If for a rigid analysis, elastic deformation is more likely for ceramic or metal surfaces at pressure values between the two exhibited pressure distributions.

In this model, the only lubricant parameter is the viscosity of the fluid. The viscosity chosen was 0.0015$Pa.s$ which is almost the same as water. The peri-prosthetic synovial fluid or blood plasma, has other molecules and components which contribute to the lubrication, however the only pertinent property in this study is viscosity. This proves to be a limiting factor as water is known not to be an effective lubricant, and the corresponding hydrodynamic pressure distribution is insufficient to complete an entire walking cycle.

This is easy to believe as it is a normal force that it cannot withstand, and is not a lubricant which is preventing asperity contact in shear flow.

The results obtained are unsurprising as it has been reported by Jin (2006) that a hydrodynamic analysis would only be applicable to conforming high elastic modulus materials. However even in the operating conditions of an artificial hip joint, elastic deformation is significant, even for metal and ceramic components.

6.3.3 Effect of Viscosity

It was clearly shown that with a low viscosity such as $0.0015Pa.s$ the walking cycle could not be completed. This leads to the investigation of what viscosity would lead to the completion of the cycle? This is an appropriate question to ask since there are no other lubricant parameters in the model developed. This gives a feel as to the effect of viscosity to load carrying capability. The measure of the effectiveness was measured by the time required for contact.

Simulations were run with the same physical parameters as those shown in table 5.1. The results of the simulations with varying viscosity yielded the results seen in table 6.4.

Viscosity $(Pa.s)$	Time of contact (s)
0.0015	0.0187
0.015	0.1056
0.15	0.5703
1.5	Did not touch

Table 6.4: Effect of viscosity on the time of contact

From table 6.4 it is clearly visible that with increasing viscosity, the time of contact is significantly improved. The results are proof that higher viscosity fluids tend to be better lubricants than lower viscosity fluids. In this simple model, only a fluid with a viscosity of $1.5Pa.s$ was able to withstand the load from the accompanying walking cycle. A much more efficient visual measure of the effectiveness of the lubricant is the manner in which the eccentricity ratio changes with time. This can be see in figure 6.21.

Figure 6.21 distinctly shows that with increasing viscosity the rate of change of the eccentricity ratio decreases. The rate of change of the eccentricity ratio is markedly different between the curve of $\mu = 1.5Pa.s$ and the lower viscosity curves. The three low viscosity curves have the same shape, a steep increase at the beginning which eventually tends to even out as the eccentricity ratio approaches one. The high viscosity curve has a gradient that increases, then decreases twice during the cycle. The corresponding gradient curve would have local maxima which coincide with the peaks in the Bergmann load cycle. This is not evident in the lower viscosity curves; for viscosity of $\mu = 0.0015Pa.s$ and $\mu = 0.015Pa.s$ this occurs because the peaks of the cycle had not been reached, in

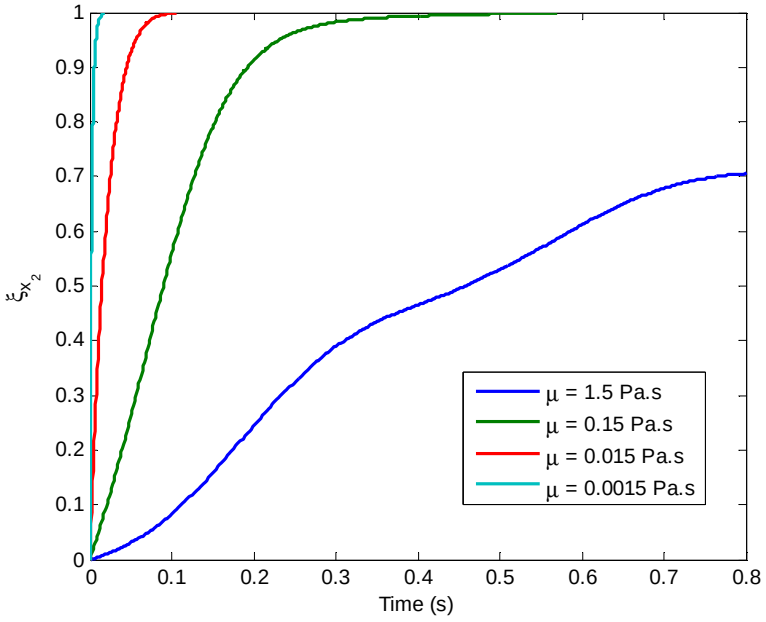

Figure 6.21: Change of eccentricity ratio during simulation with different viscosity

the case of $\mu = 0.15Pa.s$ the eccentricity ratio is too close to one to see any noticeable difference.

6.3.4 Effect of Radius and Clearance

The final investigation is to investigate the effect that the physical parameters of the hip joint has on the lubricating properties as described by the model developed. This is important as it provides valuable information with respect to the aspects in the design of artificial hip joints. The physical parameters that can be adjusted are the radius of the femoral component and the radial clearance. As in section 6.3.3, the measure of the lubricity will be measured by the time required for contact. Using this measure, it was decided to choose a suitable viscosity whereby the effect could be best noticed; therefore a viscosity of $0.15Pa.s$ was chosen for these simulations. Grid refinement needed to be done to ascertain a grid for accurate results. This was based upon section 6.3.1; based on the results from the chosen grid whereby 500-600 nodes had been used, a confirmation of using similarly as many nodes was chosen. The refinement was done with different grids using a radial clearance of $250\mu m$ and a femoral head radius of $26mm$ and the results can be seen in table 6.5;

It is evident that the grid size and spatial discretisation do not have great effect on the time of required to reach contact. To optimise for time and accuracy it was chosen to use the 241×241 grid with 600 nodes.

The radial clearance and radius were changed simultaneously and the observed changes in time to contact are shown in table 6.6.

Spatial grid size	Temporal nodes	Time of contact (s)
151×151	600	0.4407
151×151	700	0.4395
151×151	800	0.4385
181×181	600	0.4394
181×181	700	0.4383
181×181	800	0.4375
241×241	600	0.4381
241×241	700	0.4372
241×241	800	0.4365

Table 6.5: Grid refinement for model parameter investigation

Radial clearance (mm)	Femoral head radius (mm)				
	26	27	28	29	30
0.2	0.6304	0.7185	Did not touch	Did not touch	Did not touch
0.225	0.5342	0.5957	0.6611	0.7853	Did not touch
0.25	0.4381	0.5102	0.5703	0.6277	0.7025
0.275	0.3432	0.4180	0.4902	0.5503	0.6050
0.3	0.2912	0.3352	0.4033	0.4755	0.5356

Table 6.6: Time of contact (in s) variation with different radii and radial clearances

Table 6.6 shows two distinctive trends:

- increasing radius increases time to contact
- increasing radial clearance decreases time to contact

Increasing the radius of the femoral head, provides a larger surface area over which the pressure acts, resulting in the ability to hold larger loads. The larger head also allows for a larger sliding distance which can result in increased wear.

For decreasing radial clearances, there is better conformity between the acetabular and femoral components; this improves the contact area and results in more even pressure distributions, and this allows for higher loads to be withstood.

It can also be seen that there is a strong synergistic effect that takes place when radius is increased and the radial clearance is decreased. This suggests that the best implant, based on lubrication properties from this model, would be that with the largest femoral head and smallest radial clearance that would not result in asperity contact and associated wear and friction.

CHAPTER 7

Conclusions

This chapter is a summary of the main outcomes obtained from this study. It contains the major aspects of the model developed as well as the results of the simulations using both the steady-state and transient models.

7.1 Hip Joint Model

A model was developed to simulate the lubrication of an artificial hip joint. The model developed was based upon the Navier-Stokes equation of motion. This was simplified to the Reynolds equation of lubrication; this equation simulates hydrodynamic (fluid film) lubrication. The assumptions made in the derivation of this model are as follows:

- the bearing system is inelastic
- the fluid is Newtonian (isoviscous)
- the inertial forces are negligible
- the velocity gradients across the lubricant film dominate
- pressure change across the film is negligible
- the system is isothermal

These assumptions simplify to produce a velocity profile which is caused by combined Couette-Poiseulle flow (flow induced by a moving boundaries and pressure gradient).

The model developed was built in the spherical coordinate system for a hemispherical cup positioned horizontally. The film thickness was calculated based on the geometry of the system and the resultant eccentricities of the femoral component in each direction.

For the different movements and situations in which the model was used, grid refinement was performed to assess the sensitivity of the system. The grid refinement was used as a tool to obtain the grid which would give the most accurate results (with specified numerical significance) while reducing the required time for computation.

7.2 Angular Motion

An investigation into the effects that angular motion has on the current model was conducted. It was concluded that angular motion produces a perfectly anti-symmetric pressure distribution and can therefore support no load. This leads to the belief that the angular, sweeping, motion only exists to induce flow of fluid into the cavity between the two components. Under conditions of inelastic surfaces this seems to be valid, however when elastic deformation is significant, such as in an artificial hip joint bearing system some load-carrying-capability is possible (Jin, 2006).

Both Dirichlet and Neumann boundary conditions were implemented and showed pressure distributions which differed significantly in shape, especially at low eccentricity ratios. With higher eccentricity ratios they produce similar pressure profiles. These pressure distributions contain a pressure source and sink (showing the point of the global maximum and minimum in the pressure distribution); as the eccentricity ratio increases, the source and sink converge towards the centre of the cup. Not only do the source and sink converge, but the magnitude of the pressure increases as well. At extremely high eccentricity ratios, pressures obtained from the sink are so low, that the results are for all practical purposes infeasible.

The intermediary pressure constraint was also implemented, whereby the pressure was limited by only allowing positive values. The resultant pressure distribution was dome-like and could support a small load. The results of this simulation, are unrealistic as the residual from the solution procedure had not been significantly reduced to produce reasonable results; this was directly as a result of implementing the constraint. The application of this constraint is not suggested, as limiting the pressure does not always allow for the exact solution (or representative solution) to be obtained; especially when negative gauge pressure values are required to solve the given problem.

7.3 Walking Cycle

The transient model was simulated with a Bergmann cycle as its input. The model clearly showed that the angular motion had no significant contribution to the resultant pressure distribution. The inward-outward movement, which causes the transient nature of the film thickness was the dominant term. There was only movement of the femoral

component in the direction of the applied load, and the eccentricity ratio only increased in the simulation, as it is the only way to produce a positive resultant force; resultant forces in the other directions were negligible. The simulation demonstrated that the artificial hip joint, with fluid of viscosity $0.0015Pa.s$, cannot support any significant load under inelastic fluid-film conditions, as the simulation terminated with a final load of $107N$. With the viscosity used it is concluded that the system cannot in any way operate in the hydrodynamic lubrication regime without accounting for deformation. This is supported by the studies from Jin (2006).

The only lubricant parameter in the model, namely viscosity, was varied in the simulation. With increased viscosity, the weight that could be supported was increased and subsequently was able to complete more of the cycle. A substance of viscosity $1.5Pa.s$ was able to complete the entire cycle. Physical parameters that were also varied in the model were that of the femoral head radius and radial clearance. It can be concluded, that based on the hydrodynamic model developed, an increase in the femoral head radius and decrease in the radial clearance would contribute to better lubrication.

7.4 Recommendations

It is recommended that this model be extended to include elastic deformation of the acetabular and femoral components. An elasto-hydrodynamic model would give more insight into the extent of fluid-film lubrication in this model. It would therefore be necessary to conduct a contact mechanics study and build a finite element model. If possible, the simulation should be conducted without making all of the assumptions which result in the Reynolds equation; therefore, a computational fluid dynamics model be developed with appropriate software. If possible, these two models, fluid and solid models, should be coupled to solve the elasto-hydrodynamic simulation.

Attention should also be spent on trying to judge the effect of boundary lubrication on the model. This would necessitate a much better understanding of the proteins present and a mathematical concept of how these molecules would affect boundary lubrication.

The linking of an elasto-hydrodynamic and boundary lubrication model, to form a mixed lubrication model, would be the best indication yet of the extent of lubrication in an artificial hip joint. This could help significantly in the design of prostheses and the eventual development of synovial fluid.

BIBLIOGRAPHY

Bergmann, G.; Kneggendorf, H.; Graichen, F. and Rohlmann, A. (1995) "Influence of shoe and heel strike on the loading of a hip joint", *Journal of biomechanics, 28* (7).

Bhushan, B. (Ed.) (2002) *Introduction to Tribology*, John Wiley and Sons, New York.

Bigsby, R. J. A.; Auger, D. D.; Jin, Z. M.; Dowson, D.; Hardaker, C. S. and Fisher, J. (1998) "A comparative tribological study of the wear of composite cushion cups in a physiological hip joint simulator", *Journal of Biomechanics, 31*, 363–369.

Briggs, W. L. (Ed.) (1987) *A Multigrid Tutorial*, Society for Industrial and Applied Mathematics, Philadelphia.

Buford, A. and Goswami, T. (2004) "Review of wear mechanisms in hip implants: Paper 1 - General", *Materials and Design, 25*, 385–393.

Burden, R. L. and Faires, J. D. (Eds.) (2005) *Numerical Analysis*, Thomson Brooks/Cole, United States of America.

Burger, N. D. L. (2005) *Failure Analysis of Ultra-High Molecular Weight Polyethylene Acetabular Cups*, PhD thesis, University of Pretoria,.

Collins, R. (1982) "A model of lubricant gelling in synovial joints", *Journal of applied mathematics and physics, 33*, 93–123.

Crowe, C. T.; Elger, D. F. and Roberson, J. A. (Eds.) (2005) *Engineering Fluid Mechanics*, John Wiley and Sons, United States of America.

Dijkgraaf, L. C.; de Bont, L. G. M.; Boering, G. and Liem, R. S. B. (1996) "Structure of the normal synovial membrane of the temporomandibular joint: A review of the literature", *Journal of Oral Maxillofacial Surgery, 54*, 332–338.

Dowson, D. (1995) "A comparative study of the performance of metallic and ceramic femoral head components in total replacement hip joints", *Wear,* pages 171–183.

Dowson, D.; Hardaker, C.; Flett, M. and Isaac, G. H. (2004) "A Hip Joint Simulator Study of the Performance of Metal-on-Metal Joints", *The Journal of Arthroplasty, 19*, 124–130.

Dumbleton, J. H. and Manley, M. T. (2005) "Metal-on-Metal Total Hip Replacement", *The Journal of Arthroplasty, 20*, 174–188.

Ebied, A. and Journeaux, S. (2002) "Metal-on-metal hip resurfacing", *Current Orthopaedics, 16*, 420–425.

Fisher, J. and Dowson, D. (1991) "Tribology of Total Artificial Joints", *Journal of Engineering in Medicine, Part H*, 73–79.

Furey, M. J. and Burkhardt, B. M. (1997) "Biotribology: Friction, Wear and Lubrication of Natural Synovial Joints", *Lubrication Science, 9–3*, 256–271.

Germain, M. A.; Hatton, A.; Williams, S.; Williams, J. B.; Stone, M. H.; Fisher, J. and Ingham, E. (2003) "Comparison of the cytotoxity of clinically relevant cobalt-chromium and alumina wear particles in-vitro", *Biomaterials, 24*, 469–479.

Goenka, P. K. and Booker, J. (1980) "Spherical Bearings: Static and Dynamic Analysis Via the Finite Element Method", *Journal of Lubrication Technology, 102*, 308–319.

Hall, R. M.; Banks, M. J. K. and Blunn, G. (2001) "Biotribology for joint replacements", *Current Orthopaedics, 15*, 281–290.

Hamrock, B. J.; Schmid, S. R. and Jacobson, B. O. (Eds.) (2004) *Fundamentals of Fluid film Lubrication*, Marcell Dekker Inc., New York.

Hatton, A.; Nevelos, J. E.; Nevelos, A. A.; Banks, R. E.; Fisher, J. and Ingham, E. (2002) "Alumina-alumina artificial hip joints. Part 1: A histological analysis and characterisation of wear debris by laser capture microdissection of tissues retrieved at revision", *Biomaterials, 23*, 3429–3440.

Heller, M. O.; Bergmann, G.; Deuretzbacher, G.; Durselen, L.; Pohl, M.; Claes, L.; Haas, N. P. and Duda, G. N. (2001) "Musculo-skeletal loading conditions at the hip during walking and stair climbing", *Journal of biomechanics, 34*, 883–893.

Humanoid Animation Working Group "http://ovrt.nist.gov/projects/vrml/h-anim/jointInfo.html", Internet Webpage November (2006).

Hutchings, I. M. (Ed.) (1992) *Tribology: Friction and Wear of Engineering Materials*, Edward Arnold, London.

Jagatia, M.; Jalali-Vahid, D. and Jin, Z. M. (2001) "Elastohydrodynamic lubrication analysis of ultra high molecular weight polyethylene hip joint replacements under squeeze film motion", *Proceedings of the Institute of Mechanical Engineers Part H, 215*, 141–152.

Jagatia, M. and Jin, Z. M. (2002) "Analysis of elasthydrodynamic lubrication in a novel metal-on-metal hip joint replacement", *Proceedings of the Institute of Mechanical Engineers Part H, 216*, 185–193.

Jalali-Vahid, D.; Jagatia, M.; Jin, Z. M. and Dowson, D. (2000) "Elastohydrodynamic lubrication analysis of UHMWPE hip joint replacements", *Thinning film and tribological interfaces,*.

Jalali-Vahid, D.; Jagatia, M.; Jin, Z. M. and Dowson, D. (2001) "Prediction of lubricating film thickness in UHMWPE hip joint replacements", *Journal of Biomechanics, 34*, 261–266.

Jalali-Vahid, D.; Jin, Z. M. and Dowson, D. (2003) "Isoviscous elastohydrodynamic lubrication of circular point contacts with particular reference to metal-on-metal hip implants", *Proceedings of the Institute of Mechanical Engineers Part J, 217*, 397–402.

Jin, Z. M. (2006) "Theoretical studies of elastohydrodynamic lubrication of artificial hip joints", *Proceedings of the Institute of Mechanical Engineers Part J, 220*, 719–727.

Jin, Z. M. and Dowson, D. (1999) "A full numerical analysis of hydrodynamic lubrication in artificial hip joint replacements constructed from hard materials", *Proceedings of the Institute of Mechanical Engineers Part C, 213*, 355–370.

Jin, Z. M.; Dowson, D. and Fisher, J. (1997) "Analysis of fluid film lubrication in artificial hip joint replacements with surfaces of high elastic modulus", *Proceedings of the Institute of Mechanical Engineers Part H, 211*, 247–256.

Jin, Z. M.; Stone, M.; Ingham, E. and Fisher, J. (2006) "Biotribology", *Current Orthopaedics, 20*, 32–40.

Liu, F.; Jin, Z. M.; Hirt, F.; Rieker, C.; Roberts, P. and Grigoris, P. (2006)a "Transient elastohydrodynamic lubrication analysis of metal-on-metal hip implant under simulated walking conditions", *Journal of Biomechanics, 39*, 905–914.

Liu, F.; Jin, Z. M.; Roberts, P. and Grigoris, P. (2006)b "Effect of bearing geometry and structure support on transient elastohydrodynamic lubrication of metal-on-metal hip implants", *Journal of Biomechanics,*.

Liu, F.; Wang, F. C.; Jin, Z. M.; Hirt, F.; Rieker, C. and Grigoris, P. (2004) "Steady-state elastohydrodynamic lubrication analysis of a metal-on-metal hip implant employing a metallic cup with an ultra-high molecular weight polyethylene backing", *Proceedings of the Institute of Mechanical Engineers Part H, 218*, 261–270.

Mabuchi, K. and Sasada, T. (1990) "Numerical analysis of elastohydrodynamic squeeze film lubrication of total hip prosthesis", *Wear, 140*, 1–16.

Malan, A. G. "Advanced Fluid Mechanics MGM732", University of Pretoria (2006).

Malan, A. G. and Craig, K. "Numerical Thermoflow MSM732", University of Pretoria (2006).

Martini, F. H. (Ed.) (2001) *Fundamentals of Anatomy and Physiology*, Prentice-Hall, New Jersey 5th edition.

Meyer, D. (2003) "Reynolds Equation for Spherical Bearings", *Journal of Tribology, 125*, 203–206.

Middelman, S. (Ed.) (1998) *An Introduction to Fluid Dynamics*, John Wiley ans Sons Inc, New York.

Naudie, D.; Roeder, C.; Parvizi, J.; Berry, D.; Eggli, S. and Busato, A. (2004) "Metal-on-Metal Versus Metal-on-Polyethylene Bearings in Total Hip Arthroplasty", *The Journal of Arthroplasty, 19*, 35–41.

Nevelos, J.; Ingham, E.; Doyle, C.; Streicher, R.; Nevelos, A.; Walter, W. and Fisher, J. (2000) "Microseparation of the Centers of Alumina-Alumina Artificial Hip Joints During Simulator Testing Produces Clinically Relevant Wear Rates and Patterns", *Journal of Arthroplasty, 15*, 793–795.

Nevelos, J. E.; Ingham, E.; Doyle, C.; Nevelos, A. B. and Fisher, J. (2001) "Wear of HIPed and non-HIPed alumina-alumina under standard and severe simulator testing conditions", *Biomaterials, 22*, 2191–2197.

Nigg, B. M. and Herzog, W. (Eds.) (1995) *Biomechanics of the musculo-skelketal system*, John Wiley and Sons, New York.

Pieterse, N. (2005) "Development of a dynamic hip joint simulation", Master's thesis, University of Pretoria,.

Quateroni, A. and Valli, A. (Eds.) (1991) *Numerical Approximation of Partial Differential Equations*, Springer, Germany.

Saladin, K. S. (Ed.) (2007) *Anatomy and physiology: the unity of form and function*, McGraw-Hill, New York 4th edition.

Sinha, P.; Singh, C. and Prasad, K. R. (1982) "Lubrication of human joints - a micro-continuum approach", *Wear, 80*, 159–181.

Sokoloff, L. (Ed.) (1978) *The joints and synovial fluid. Vol I*, Academic Press, New York.

Tipper, J. L.; Firkin, P. J.; Besong, A. A.; Barbour, P. S. M.; Nevelos, J.; Stone, M. H.; Ingham, E. and Fisher, J. (2001) "Characterisation of wear debris from UHMWPE on zirconia ceramic, metal-on-metal and alumina ceramic-on-ceramic hip prostheses generated in a physiological anatomical hip joint simulator", *Wear, 250*, 120–128.

Tipper, J. L.; Hatton, A.; Nevelos, J. E.; Ingham, E.; Doyle, C.; Streicher, R.; Nevelos, A. B. and Fisher, J. (2002) "Alumina-alumina artificial hip joints. Part 2: Characterisation of the wear debris from in vitro hip joint simulations", *Biomaterials, 23*, 3441–3448.

Tipper, J. L.; Ingham, E.; Jin, Z. M. and Fisher, J. (2005) "The science of metal-on-metal articulation", *Current Orthapaedics, 19*, 280–287.

Torzilli, P. A. and Mow, V. C. (1976)a "On the fundamental fluid transport mechanism through normal and pathological articular cartilage during function - I formulation", *Journal of Biomechanics, 9*, 541–552.

Torzilli, P. A. and Mow, V. C. (1976)b "On the fundamental fluid transport mechanism through normal and pathological articular cartilage during function - II the analysis, solution and conclusions", *Journal of Biomechanics, 9*, 587–606.

Udofia, I. J. and Jin, Z. M. (2003) "Elastohydrodynamic lubrication analysis of metal-on-metal hip resurfacing prosthesis", *Journal of Biomechanics, 36*, 537–544.

Walter, W. L.; Insley, G. M.; Walter, W. K. and Tuke, M. A. (2004) "Edge Loading in Third Generation Alumina Ceramic-on-Ceramic Bearings", *Journal of Arthroplasty, 19*, 402–413.

Wang, A.; Yue, S.; Bobyn, J. D.; Chan, F. W. and Medley, J. B. (1999) "Surface characterization of metal-on-metal hip implants tested in a hip simulator", *Wear, 225–229*, 708–715.

Welty, J. R.; Wicks, C. E.; Wilson, R. E. and Rorrer, G. (Eds.) (2001) *Fundamentals of Momentum, Heat and Mass Transfer*, John Wiley and Sons, United States of America 4th edition.

Wesseling, P. (Ed.) (1992) *An Introduction to Multigrid Methods*, John Wiley and Sons, Chichester.

White, F. M. (Ed.) (2006) *Viscous Fluid Flow*, McGraw Hill, Singapore.

Wikipedia "http://en.wikipedia.org/wiki/Bone", Internet Webpage November (2006)a.

Wikipedia "http://en.wikipedia.org/wiki/Image:Illu_synovial_joint.jpg", Internet Webpage November (2006)b.

Wimmer, M. A.; Loos, J.; Nassutt, R.; Heitkemper, M. and Fischer, A. (2001) "The acting wear mechanisms on metal-on-metal hip joint bearing: *in vitro* results", *Wear, 250*, 129–139.

Wimmer, M. A.; Sprecher, C.; Hauert, R.; Täger, G. and Fischer, A. (2003) "Tribochemichal reaction on metal-on-metal hip joint bearings A comparison between in-vitro and in-vivo results", *Wear, 255*, 1007–1014.

Yew, A.; Udofia, I.; Jagatia, M. and Jin, Z. M. (2004) "Analysis of elastohydrodynamic lubrication in McKee-Farrar metal-on-metal hip joint replacements", *Proceedings of the Institute of Mechanical Engineers Part H, 218*, 27–34.

APPENDIX A

Additional CD material

Accompanying this dissertation is a CD-ROM which includes the following:

- **Matlab code**
  This section contains all the MATLAB code for performing any of the simulations used in this dissertation.
- **Simulation results**
  Animations are available for some of the transient analysis conducted.

APPENDIX B

Description of MATLAB Code

This chapter will give a brief overview of the code used to generate the results which forms the basis of this dissertation. The manner in which the program was built and the mfiles coded to build the program will be discussed briefly. The manner in which the programs were structured with the given mfiles is shown in figures B.1, B.2 and B.3.

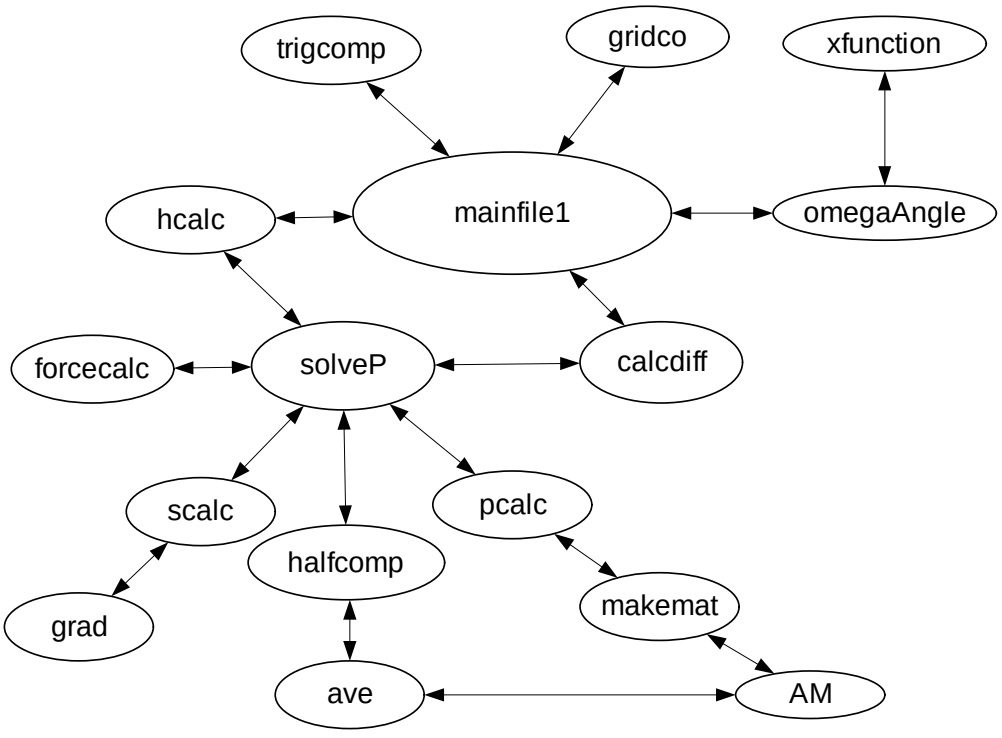

Figure B.1: Program structure for walking cycle simulation

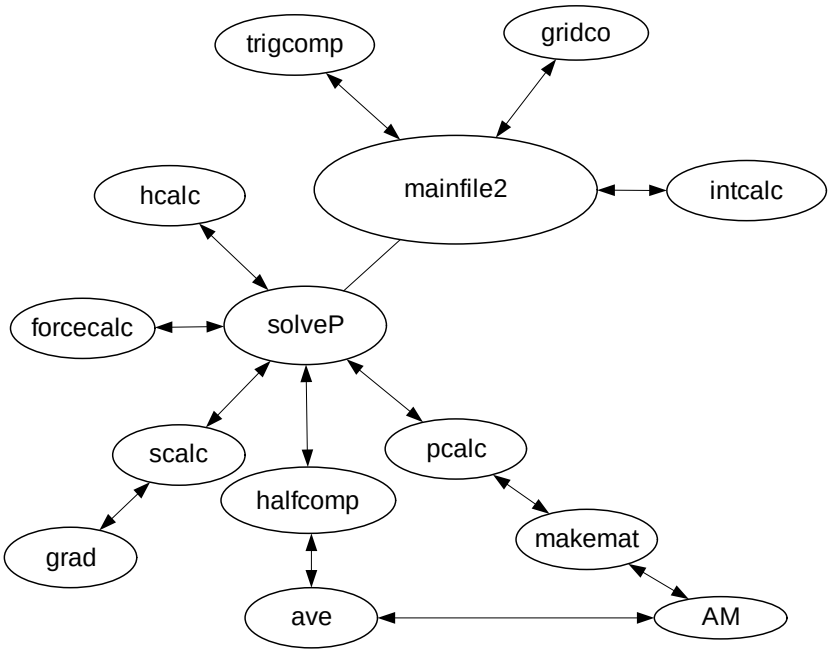

Figure B.2: Program structure for angular rotation simulation without intermediary constraint

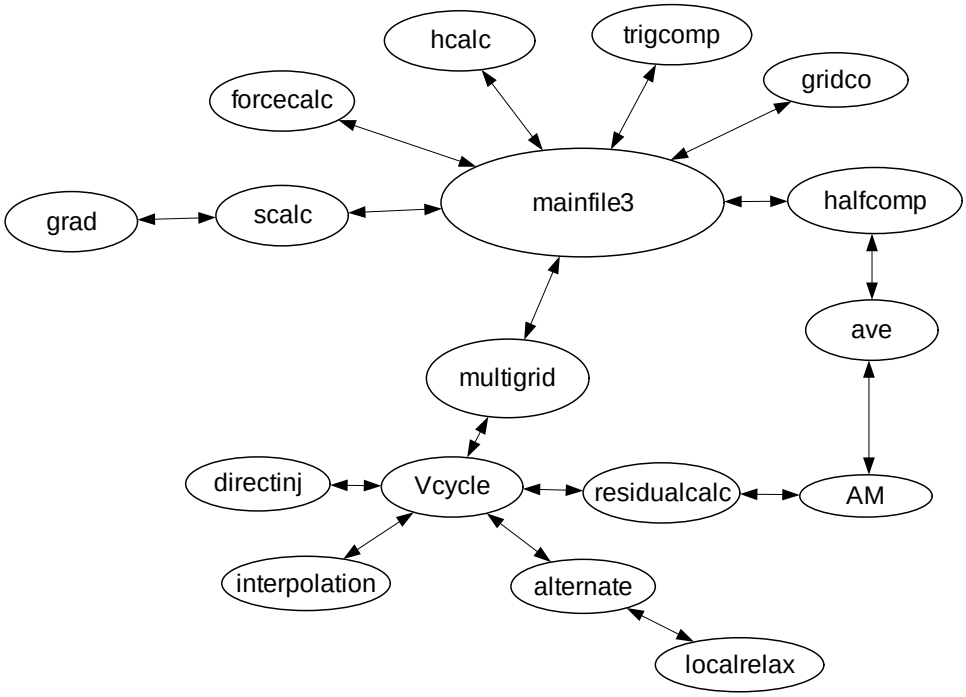

Figure B.3: Program structure for angular rotation simulation with intermediary constraint

mainfile1

This is the main file used to run the program for transient analysis. In this file the physical parameters (viscosity, femoral head radius, clearance) and grid size are specified. The angular velocity and input load are obtained. The grid coordinates and associated trigonometric identity values are obtained for use in subsequent functions. A for loop with constraints is used to run through the walking cycle obtained. The eccentricity ratio is obtained to minimise the difference between the applied and calculated load. This is done by using the built-in MATLAB function *lsqnonlin*. *lsqnonlin* is an optimising function used to solve nonlinear least-squares curve fitting problems. The obtained eccentricity ratio is then used to calculate the representative pressure distribution and is stored amongst the variables.

gridco

This function calculates two matrices representing the coordinates of the grid. One matrix is for the θ values for the given node, the other is for the ϕ values.

trigcomp

The function *trigcomp*, calculates $\sin\theta$, $\cos\theta$, $\sin\phi$ and $\cos\phi$ for each of the coordinates in the grid.

hcalc

This function calculates the film thickness using equation 3.26.

calcdiff

The purpose of this function is to calculate the difference between the obtained force (from the specified eccentricity ratio) and the applied load. This is the function used by *lsqnonlin*, and the difference is to be minimised. Information is sent to the function *solveP*, and receives back the calculated pressure distribution and resultant force.

solveP

This function sends through the calculated pressure distribution and resultant force. With the specified eccentricity ratios the film thickness is solved using *hcalc*. The half-components from the applied discretisation are obtained using the *halcomp* function. The source values (RHS of equation 6.3 or 3.24) are then calculated. The calculated film thickness, half components and source terms are sent to the function

pcalc, and the calculated pressure distribution is obtained. With the given pressure distribution, the resultant force is then calculated.

forcecalc

forcecalc calculates the resultant forces in the three Cartesian direction by implementing equations 4.2, 4.3 and 4.4.

pcalc

This function sets the values for the boundary nodes and calculates the pressure at every point in the grid. The pressure is solved by creating a matrix to solve for the pressure directly, the matrix is obtained by using the function *makemat*. The pressure is solved by using a set of linear equations in the same form as equation 4.5.

makemat

Function *makemat*, creates the matrix used to solve for the pressure. It specifies coordinates in this linear matrix and specifies the values stored in it; this is done because the matrix is very sparse and doing it in such a manner safes space and computational time.

scalc

This function calculates the source term based on equations 6.3 or 3.24. The partial derivatives of the film thickness are calculated, using function *grad*, and for equation 3.24, the temporal derivative as well. Using this information, the source term is calculated.

grad

grad calculates the gradient between two points.

halfcomp

This function calculates all the half components from the result of the discretisation procedure, $h_{i,j+\frac{1}{2}}$, $h_{i,j-\frac{1}{2}}$, $h_{i+\frac{1}{2},j}$, $h_{i-\frac{1}{2},j}$, $\sin\theta_{j+\frac{1}{2}}$ and $\sin\theta_{j-\frac{1}{2}}$.

ave

ave calculates the average between two values. This function was only built so that both the arithmetic and geometric mean could be tested in the model.

AM

Function *AM* calculates the arithmetic mean between two values.

omegaAngle

This function calculates the angular velocity based on the information obtained from a hip joint simulator, and its geometry.

xfunction

This is the function called up function *omegaAngle* to calculate the angular velocity.

mainfile2

This is the main file used to the run the program for steady-state rotation without the intermediary constraint. In this file, the physical parameters and grid sizes are specified. The input parameters of eccentricity ratios and angular velocity are also specified. The grid coordinates and associated trigonometric identity values are obtained for use in *solveP*. The pressure distribution and resultant forces are then obtained from *solveP*. The integrals used for analysis in section 6.1 are then calculated.

intcalc

Function *intcalc* calculates the integrals used to evaluate grid refinement by implementing equations 6.1 and 6.2.

mainfile3

This is the main file used to the run the program for steady-state rotation with the intermediary constraint. In this file, the physical parameters and grid sizes are specified. The input parameters of eccentricity ratios and angular velocity are also specified. The grid coordinates and associated trigonometric identity values are obtained for using *gridco* and *trigcomp*. The half components, film thickness and source terms are also obtained using functions *halfcomp*, *hcalc* and *scalc*. The pressure distribution calculated with the intermediary constraint is then obtained using *multigrig*. From the pressure distribution, the resultant forces are then obtained with *forcecalc*.

Vcycle

This function follows the algorithm discussed in section 4.3 and shown in figure 4.4 to

obtain the pressure distribution. It does this making use of the user-made function: *directinj*, *interpolation*, *alternate* and *residualcalc*.

directinj

This function transfers information directly from a coarse grid to a fine grid.

alternate

alternate is the function used to visit each node in the grid. It makes use of the alternating procedure described in section 4.2.2.

localrelax

This function calculates the new value of the pressure of the primary node in the grade based on equation 4.1 implementing the Gauss-Seidel method.

residualcalc

Function *residualcalc* calculates the residual of every node in the grid by calculating the difference between the RHS and LHS of equation 4.1

interpolation

This is the function that does the two dimensional interpolation to transfer information from the coarse grids to the fine grids (prolongation).