Analysis of Complete Contacts Subject to Fatigue

Thesis submitted to the Department of Engineering Science of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

by

Robert C. Flicek

Trinity College

Oxford

Michaelmas Term, 2014

Analysis of Complete Contacts Subject to Fatigue

Robert C. Flicek
Doctor of Philosophy in Engineering Science
Trinity College, Oxford, Michaelmas Term, 2014

Engineering assemblies are very frequently subject to fretting fatigue, which is a damage process that results when very small slip displacements arise at nominally stationary frictional interfaces. Fretting accelerates the initiation and early propagation of fatigue cracks, thereby causing significant reductions in the fatigue performance of many critical engineering components. A majority of the previous research on fretting fatigue has focused on incomplete (i.e. smooth-edged) contacts, while complete (i.e. sharp-edged) contacts have received less attention.

The aim of this thesis is to contribute to the theoretical understanding of complete contacts, especially when they are subject to fatigue conditions. This problem is addressed in two separate ways. First, because fretting failures almost invariably initiate from the edge of contact, a detailed understanding of the conditions in this region should enable more accurate assessments of fatigue performance to be made. Thus, an asymptotic analysis is presented, which provides an accurate description of the contact edge under many conditions. This is done by using the elasticity solution for a semi-infinite notch to represent the state of stress near the contact edge in an asymptotic sense.

Attention is then placed on the fact that cyclically loaded frictional contacts tend toward a steady-state response in which less frictional slip (and energy dissipation) occurs than in the first few load cycles. To investigate this effect, a numerical sub-structuring procedure is described, which significantly reduces the number of degrees of freedom in finite element models of frictional contact. This reduced model is then used to calculate the shakedown limit, i.e. the amplitude of cyclic load above which frictional slip is guaranteed to persist in the steady state. The sensitivity of the steady-state solution to the initial residual displacement state is then investigated, and it is shown that initial conditions can have a large influence on the steady-state behaviour of complete contacts.

For my family

Acknowledgements

I would like to thank my supervisor and friend Professor David Hills for the exceptional guidance and support he has provided throughout my studies. My work has benefited immensely from the countless hours he has spent discussing results, suggesting new ways to approach problems, and providing me with all the help I needed to get the most out of my studies. I would especially like to thank him for being exceptionally supportive during the second year of my studies when I was faced with some challenging personal circumstances.

I would like to express my sincere gratitude to Professor James Barber from the University of Michigan, Dr. Daniele Dini from Imperial College London, and Dr. Matthew Brake from Sandia National Laboratories with whom I've had the privilege of working. I would particularly like to thank Matthew Brake for giving me the opportunity to work with him at Sandia National Laboratories in New Mexico during the first two summers of my research. I would also like to thank James Barber for inviting me to visit him at the University of Michigan and for his invaluable input to my research through numerous discussions and lengthy email exchanges.

I very gratefully acknowledge Rolls-Royce and the Technology Strategy Board for funding my research through the project SILOET-II. I would also like to extend my gratitude to Dr. Ramesh Rajasekaran from Rolls-Royce for arranging for me to work at Rolls-Royce's facility in Derby for several weeks during the third year of my studies.

I would also like to thank Robert Paynter and Rodolfo Fluery from the University of Oxford for very generously spending their time discussing the numerous technical difficulties I came across during the course of my research.

Finally, I would like to thank my parents and my brother for all the love and support they have given me throughout my life. I would also like to thank my godmother Tessa and her son Chris for always being there when I've needed them during my time in Oxford.

Preface

This thesis is a collection of research performed by the author in the Department of Engineering Science at the University of Oxford under the supervision of Professor David A. Hills. This work is original and no part of it has been submitted for a degree at this university or elsewhere.

Some of the work described in this thesis has been published in the following papers:

1. R.C. Flicek, D.A. Hills, and D. Dini. Refinements in the characterisation of mode-mixity and small scale yielding at sharp notch roots. *Eng. Fract. Mech.*, 126:73–86, August 2014. doi:10.1016/j.engfracmech.2014.04.022.

2. R.C. Flicek, D.A. Hills, J.R. Barber, and D. Dini. Determination of the shakedown limit for large, discrete frictional systems. *Eur. J. Mech. A-Solid*, 49:242–250, January-February 2015. doi:10.1016/j.euromechsol.2014.08.001.

3. R.C. Flicek, R. Ramesh, and D.A. Hills. A complete frictional contact: The transition from normal load to sliding. *Int. J. Eng. Sci.*, 92:18–27, July 2015. doi:10.1016/j.ijengsci.2015.03.006.

4. R.C. Flicek, D.A. Hills, and D. Dini. Sharp edged contacts subject to fretting: A description of corner behaviour. *Int. J. Fatigue*, 71:26–34, February 2015. doi:10.1016/j.ijfatigue.2014.02.015.

5. R. Flicek, D.A. Hills, and D. Dini. Progress in the application of notch asymptotics to the understanding of complete contacts subject to fretting fatigue. *Fatigue. Fract. Eng. M.*, 36(1):56–64, January 2013. doi:10.1111/j.1460-2695.2012.01694.x.

6. D.A. Hills, R.C. Flicek, and D. Dini. Sharp contact corners, fretting and cracks. *Frattura ed Integritá Strutturale (Fracture and Structural Integrity)*, 25:27–35, July 2013. doi:10.3221/igf-esis.25.05.

7. A. Thaitirarot, R.C. Flicek, D.A. Hills, and J.R. Barber. The use of static reduction in the finite element solution of two-dimensional frictional contact problems. *P. I. Mech. Eng. C-J. Mech.*, 228(9):1474–1487, June 2014. doi:10.1177/0954406213509086.

8. D.A. Hills, R.C. Flicek, and D. Dini. A discussion of: Development of a complete contact fretting test device by J Juoksukangas et al. *P. I. Mech. Eng. J-J. Eng.*, 228 (1):123–126, January 2014. doi:10.1177/1350650113500946.

9. D.A Hills and R.C. Flicek. A discussion of "Numerical and experimental investigations on fretting fatigue: Relative slip, crack path, and fatigue life" by N. Noraphaiphipaksa, C. Kanchanomai, and Y. Mutoh. *Eng. Fract. Mech.*, 113:52–55, January 2015. doi:10.1016/j.engfracmech.2014.10.024.

Contents

Abstract i
Acknowledgements iii
Preface iv

Contents **v**

1 Introduction **1**
1.1 Fatigue failure 1
1.2 Types of contacts 4
1.3 Literature Review 8
1.3.1 Friction 9
1.3.2 Fretting fatigue 11
1.3.3 Contact mechanics 15
1.3.4 Analysis of complete contacts 16
1.3.5 Steady-state behaviour 19
1.4 Thesis summary 21

2 Semi-infinite notch theory **25**
2.1 Motivation 25
2.2 Williams' solution 26
2.2.1 Stress function 27
2.2.2 Boundary conditions 29
2.2.3 Relevant terms 31
2.2.4 Eigenvalues 34
2.2.5 Eigenvectors 35
2.2.6 Generalized stress intensity factors 37
2.3 Alternative set of parameters 39
2.4 Summary 41

3 Characterization of the process zone of sharp V-notches **42**
3.1 Introduction 42
3.2 Mode-mixity at the process zone 44
3.2.1 Comparison of pure and mixed-mode process zones 46
3.2.2 Mode-mixity results 48
3.3 Small-scale yielding 51
3.3.1 Small-scale-yielding results 53
3.4 Application to sharp V-notches 55
3.4.1 Small-scale yielding of experimental data 58

3.4.2 Mode-mixity of experimental data . . . 63
3.5 Summary . . . 64

4 Asymptotic analysis of complete contacts 66
4.1 Introduction . . . 66
4.2 Implications of Williams' solution for adhered complete contacts . . . 70
4.2.1 Properties of the mode I eigensolution . . . 72
4.2.2 Properties of the mode II eigensolution . . . 75
4.3 Application to clamped-cantilever test rig . . . 77
4.3.1 Application to fretting-fatigue tests . . . 79
4.4 Re-examination of fretting-fatigue data . . . 80
4.5 Discussion . . . 83
4.6 Summary . . . 85

5 Slip and separation at the contact edge 86
5.1 Introduction . . . 86
5.2 Types of contact-edge response . . . 88
5.3 Calculation of the extent of slip and/or separation . . . 91
5.3.1 Case (a): $K_I < 0 < K_{II}$ . . . 91
5.3.2 Case (b): $0 < K_I, K_{II}$ . . . 95
5.3.3 Case (c): $K_I, K_{II} < 0$ . . . 96
5.4 Implications of the internal length dimension . . . 98
5.5 Application to clamped-cantilever test rig . . . 99
5.5.1 Other considerations . . . 101
5.6 Extension to other contact angles . . . 102
5.7 Discussion . . . 105
5.8 Summary . . . 107

6 Asymptotic analysis of fretting-fatigue experiments 108
6.1 Mode-mixity and small-scale yielding of fretting tests . . . 108
6.1.1 Range of applicability . . . 109
6.2 Juoksukangas' fretting tests . . . 112
6.2.1 Mode-mixity and small-scale yielding . . . 112
6.2.2 Implications for fretting performance . . . 114
6.3 Mugadu's fretting tests . . . 116
6.3.1 Experimental configuration . . . 117
6.3.2 Mugadu's analysis of the fretting tests . . . 120
6.3.3 Finite element model . . . 121
6.3.4 Type A and type B contact edges . . . 123
6.3.5 Asymptotic analysis of the test rig . . . 125
6.3.6 Analysis of fretting tests . . . 126
6.3.7 Examination of fatigue performance . . . 130
6.4 Noraphaiphipaksa's fretting tests . . . 134
6.4.1 Asymptotic analysis of test rig . . . 136
6.4.2 Application to fretting tests . . . 137
6.4.3 Implications for fretting performance . . . 139
6.5 Discussion . . . 140

6.6 Summary . . . 142

7 Static reduction of finite element models 143
7.1 Introduction . . . 143
7.2 Single elastic body in contact with a rigid body . . . 145
7.2.1 Sign convention . . . 147
7.2.2 Eliminating the unloaded nodes . . . 148
7.2.3 Force-controlled loading . . . 150
7.2.4 Displacement-controlled loading . . . 151
7.3 Contact between two elastic bodies . . . 152
7.4 Partitioning the contact stiffness matrix . . . 154
7.5 Evolution of the contact interface with time . . . 155
7.5.1 Gauss-Seidel solution algorithm . . . 156
7.6 Incomplete contact problems . . . 157
7.6.1 Modification of the Gauss-Seidel algorithm . . . 159
7.6.2 Modification of the loading vector . . . 159
7.7 Reduction in simulation run-time . . . 161
7.8 Summary . . . 166

8 The transition from normal load to sliding 167
8.1 Introduction . . . 167
8.2 Finite element model . . . 170
8.3 Results . . . 172
8.3.1 High friction . . . 172
8.3.2 Low friction . . . 176
8.4 Discussion . . . 179
8.4.1 Other considerations . . . 180
8.5 Summary . . . 186

9 Calculation of the shakedown limit 187
9.1 Introduction . . . 187
9.2 Formulation . . . 193
9.2.1 The Coulomb friction law . . . 194
9.2.2 Loading regime . . . 195
9.3 First violation of the stick condition . . . 196
9.4 The shakedown limit . . . 196
9.4.1 Calculation of the shakedown limit . . . 197
9.4.2 Improving convergence . . . 199
9.5 Application to example problem . . . 199
9.5.1 First violation of the stick condition . . . 202
9.5.2 The shakedown limit . . . 203
9.5.3 Frictional energy dissipation . . . 204
9.5.4 Optimal initial residual slip displacement . . . 207
9.5.5 Comparison of results . . . 209
9.6 Discussion . . . 210
9.7 Summary . . . 212

10 Conclusion **213**
10.1 Overview . . . 213
10.2 Conclusions . . . 215
10.3 Future work . . . 219

Bibliography **229**

A Additional material **240**
A.1 Additional material for §3 . . . 240
A.1.1 Mode-mixity calculation . . . 240
A.1.2 Small-scale-yielding calculation . . . 242
A.1.3 Processing of experimental data . . . 243
A.1.4 Experimental data . . . 247
A.2 Additional material for §6 . . . 248
A.2.1 Juoksukangas' experimental data . . . 248
A.3 Additional material for §7 and §9 . . . 249
A.3.1 The finite element model . . . 249

Chapter 1

Introduction

The aim of this thesis is to present useful techniques for analysing sharp-edged frictional contacts. In this chapter, the motivation for this thesis is discussed by providing a description of the practical consequences of high-cycle fatigue and fretting fatigue, especially in the aerospace industry. A description of several types of practically occurring frictional contacts is then presented. The literature related to this thesis is briefly reviewed. Finally, a short summary of each chapter is provided.

1.1 Fatigue failure

Engineering components are very frequently subject to fatigue loading as a result of their service conditions, e.g. due to mechanical vibrations. If the resulting cyclic loading is severe enough and persists for a sufficient number of cycles, material failure can eventually occur as a result of the propagation of fatigue cracks. Unfortunately, fatigue failures can be particularly difficult to design against because failure can occur even if the stresses remain well below the yield strength of the material throughout the service life of the component. Consequently, research on the subject of fatigue failure has been ongoing for over 150 years [Schütz, 1996], and although much progress has been made, many difficult problems still remain to be solved.

The primary motivation for fatigue research comes from the numerous engineering structures that have failed in service due to fatigue. These failures invariably cause major financial losses and often also result in fatalities. Even in the absence of such failures,

however, gaps in the current understanding of fatigue still increase the cost of engineering projects. This is because the lack of accurate models is usually compensated for by applying a larger safety factor or by regularly inspecting components for fatigue cracks.

While such conservatism does not always incur an unacceptably high cost, in some applications, e.g. the gas-turbine engines used in the aerospace industry, marginal increases in component weight have knock-on effects (e.g. reducing fuel efficiency and increasing the cost of other engine components) that significantly decrease the efficiency of the overall design. Performing regular inspections of critical components is also very costly and requires the engine to be removed from service for some period of time. Hence, in cases such as these, there is a very strong incentive to improve current approaches for predicting the fatigue performance of engineering components.

Fatigue loading is generally divided into two categories: high-cycle fatigue and low-cycle fatigue. The distinction between these two regimes is defined by the number of load cycles that are required to cause the component to fail at a given cyclic stress amplitude. The boundary is usually taken to be around 10^4 cycles to failure, where high-cycle fatigue corresponds to more than 10^4 cycles to failure, and low-cycle fatigue corresponds to less than 10^4 cycles to failure.

The service life of a component subject to fatigue conditions can generally be divided into (at least) two distinct phases: an initial crack nucleation/initiation phase followed by a crack propagation phase. As noted by Nicholas [1999], a very important difference between high-cycle fatigue and low-cycle fatigue is the proportion of the component's life that is spent in each of these phases. In low-cycle fatigue, cracks usually initiate very early on, and the majority of the component's life is spent in the crack propagation phase. Conversely, in high-cycle fatigue, a significant portion of the component's life is usually spent in the crack nucleation phase, and failure occurs shortly after cracks become large enough to be detected in component inspections.

At present, the crack propagation phase is much more fully understood than the crack nucleation phase. Consequently, failures due to low-cycle fatigue can be prevented more

easily than those from high-cycle fatigue. Ciavarella and Meneghetti [2004] note that a common technique for doing this, which has been quite successful and is currently in use, is the *damage tolerant approach*, which assumes the component possesses an initial flaw, and the fatigue life is then predicted using Paris' law. However, this approach is only possible because the number of cycles spent in the crack initiation phase is small. As a consequence, the resulting fatigue cracks are large enough to be detected in routine inspections for a relatively large portion of the component's service life, so damaged components can be detected and removed from service before failure occurs.

Unfortunately, a damage tolerant approach cannot easily be applied to high-cycle fatigue because a large portion of the component's life is spent outside the crack propagation phase. This is a serious problem because, as noted by Cowles [1989], high-cycle fatigue is the "largest single cause of component failures in modern military aircraft gas turbine engines", and it also causes "significant economic impacts through field inspection and maintenance actions, and reduced readiness reliability." Although several techniques have been developed that successfully predict the plain-fatigue performance of engineering components, e.g. S-N plots and Goodman diagrams, Cowles [1989] notes that a major problem with these approaches for designing against high-cycle fatigue is that they have only a limited ability to predict the life of components that sustain damage from other sources (or a combination of both high-cycle and low-cycle fatigue) during their service life. Nicholas [1999] states that the three most common sources of in-service damage to gas-turbine engines (aside from high-cycle fatigue) are low-cycle fatigue, foreign-object damage, and fretting.

This thesis is concerned with *fretting fatigue*, which occurs when small amounts of relative slip displacement (usually on the order of 25-$100\mu m$ [Szolwinski and Farris, 1996]) arise at the frictional interfaces in jointed structures. Fretting is a particularly severe damage mechanism; for instance, Cowles [1989] reports a reduction in fatigue life of around 40% for both laboratory test specimens and actual engine components subjected to a combination of fretting and high-cycle fatigue in comparison to those subjected to high-cycle

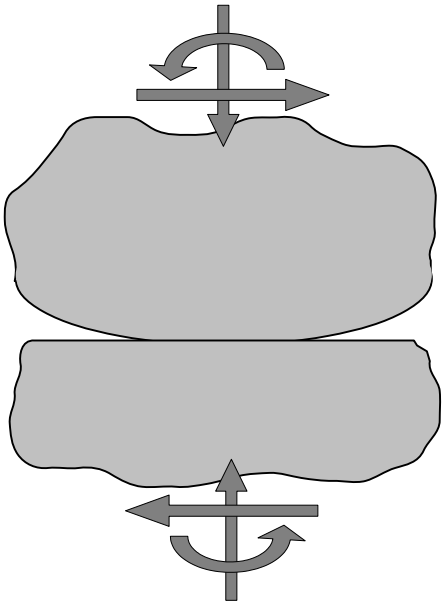
fatigue alone. However, as is emphasized by Cowles [1989] and Nicholas [1999], many of the analytical tools for assessing fretting performance are highly empirical, which makes designing against fretting a very time consuming and expensive task.

The aim of this thesis is to develop techniques that improve the current capability to assess the fretting performance of engineering components subject to high-cycle fatigue. Clearly, the service life of components subject to fretting is determined by the conditions arising in the region where fretting damage accumulates and fatigue cracks initiate. Hence, this thesis focuses on developing computationally efficient techniques that provide an accurate description of the conditions arising at the frictional interfaces at which fretting occurs. Since fretting damage is usually localized at the edges of the contact interface, much of the analysis pursued in this thesis places emphasis on describing contact behaviour in this region, i.e. §2-§6, but tools for examining the behaviour of the contact interface as a whole are also presented in §7-§9.

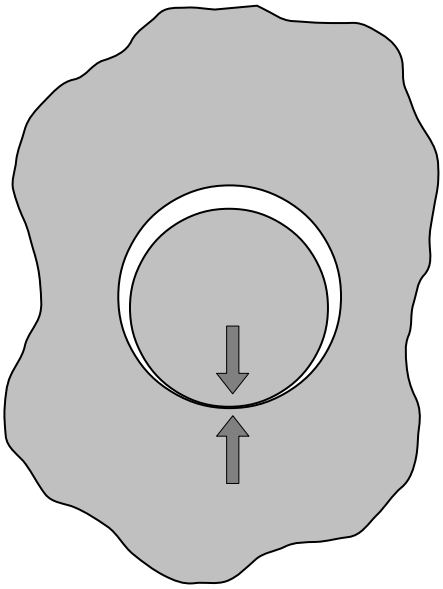
1.2 Types of contacts

Modern engineering structures frequently incorporate numerous highly complex frictional contact geometries. However, these contacts can generally be grouped into two categories: *conforming* and *non-conforming*. Conforming contacts are defined as those for which the contact area is finite when no external loads are applied, which implies that the contact is defined by the geometry of (at least) one of the bodies. Conversely, non-conforming contacts, which are also frequently referred to as *incomplete contacts*, are defined as those for which the contact area is infinitesimal in the absence of external loads. These definitions usually imply that incomplete contacts are formed between two bodies that have smoothly varying surface profiles, whereas conforming contacts result when one (or more) of the bodies has a sharp edge in its profile.

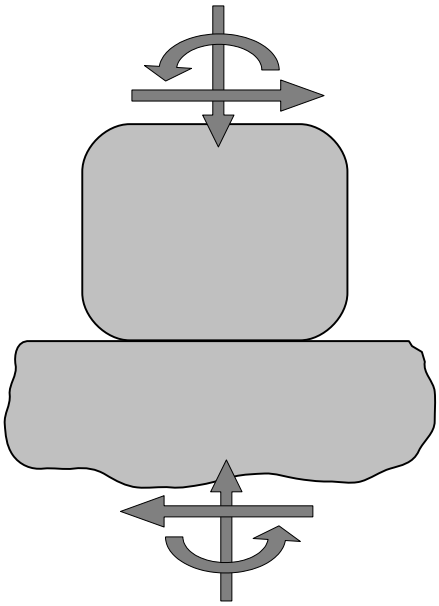
Several types of idealized frictional contacts are shown in Figure 1.1. The contacts shown in Figure 1.1(a-b) are incomplete contacts, and those shown in Figure 1.1(c-e) are conforming contacts. Because incomplete contacts have a smooth surface profile, the



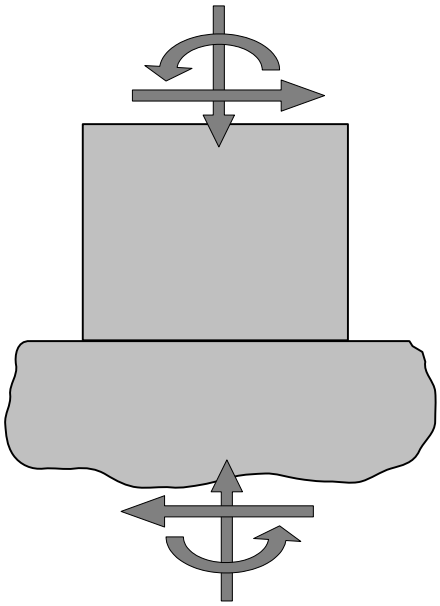
(a) Hertzian contact.



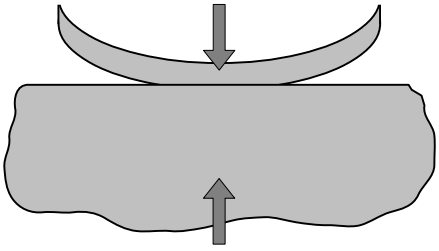
(b) Pin-in-hole contact.



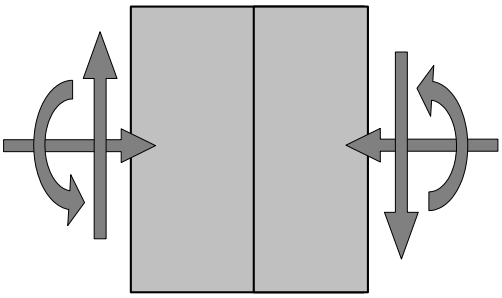
(c) Flat-and-rounded contact



(d) Complete contact



(e) Receding contact.



(f) Common-edged contact.

Figure 1.1: Types of contacts.

contact pressure distribution for these contacts always tends to zero as the contact edge is approached. For both the Hertzian and the pin-in-hole contacts shown in Figure 1.1(a-b), the contact pressure takes on its maximum value at the centre of the contact. The main difference between these two types of contacts is that the Hertzian contact can be modelled by treating both bodies as half-planes because the contact area is small in comparison to the size of the bodies, and the curvature of the bodies is small. Conversely, neither body can be treated as a half-plane for the pin-in-hole contact since the contact area is comparable to the size of the pin, and the curvature of the bodies is too great.

Strictly speaking, the flat-and-rounded contact shown in Figure 1.1(c) is a conforming contact because the contact area is partially defined by the upper body. However, since this contact has a smooth profile near its edges, its contact area still depends on the normal load (although to a lesser extent). This type of contact also produces a more complex contact pressure distribution than the two incomplete contacts just described, but which still locally tends to zero at the contact edge. For this reason, the flat-and-rounded contact might be said to be locally 'incomplete' at the contact edges (since the contact pressure distribution resembles that of an incomplete contact in this region). Thus, in some ways, the flat-and-rounded contact represents a transition between conforming and non-conforming contacts.

The contact shown in Figure 1.1(d) is referred to as a *complete contact*, which is defined as a contact whose contact area is independent of the applied normal load. For this type of contact, the application of a normal load produces a singularity in the elastic state of stress at the contact edge (and hence in the contact pressure distribution). The order of this singularity depends on the contact-edge angle, the degree of elastic dissimilarity of the bodies, and the coefficient of friction. When a complete contact is subjected to a shear or some other applied load, the contact-edge state may change, such that new slip zones may form, or the edges may 'lift off' (i.e. separate). Thus, the contact area of a complete contact may *decrease* as a result of the applied loads, but it cannot increase.

The contact shown in Figure 1.1(e) is referred to as a *receding contact*, which is here

defined as a contact whose contact area *decreases* on the application of the normal load. Dundurs and Stippes [1970] have shown that when a normal load is monotonically applied to this type of contact, both contact edges lift off, such that zones of separation and adjacent slip form. As the normal load is monotonically increased further, the sizes of the slip and separation zones remain fixed, and the displacements scale linearly with the applied load. Since significant separation can occur at the edges of such contacts, the contact's edges become locally incomplete (as with the flat-and-rounded contact). Thus, as with the incomplete contacts shown in Figure 1.1(a-b), the contact pressure for receding contacts locally tends to zero at the contact edges.

It should be noted that receding contacts are sometimes defined, e.g. by Dundurs and Stippes [1970], as a contact whose contact area cannot increase, implying that all complete contacts are also receding contacts. One benefit of this definition is that whether a contact is complete or receding (according to the definitions used here) depends not only on the contact geometry but also on the way in which the normal load is applied,[1] whereas this alternative definition accounts for both kinds of response.

The contact shown in Figure 1.1(f) is referred to as a *common-edged contact*, which is defined as a contact whose contact area is defined by both bodies simultaneously, i.e. when the two bodies share common edges. For these contacts, the contact pressure distribution takes on a finite value at all points – including at the contact edges. This is a unique property of this type of contact since the contact pressure tends either to zero or infinity for all the other contacts shown in Figure 1.1. However, common-edged contacts are perhaps not as relevant in practical engineering applications because it is very unlikely that the contact edges will ever be perfectly aligned in practice. And if any minor misalignment is present, the region local to the contact edge then begins to resemble the edge of the complete contact shown in Figure 1.1(d). Thus, when analysing common-edged contacts, some amount of misalignment should probably be assumed, and they should be

[1] For example, the contact shown in Figure 1.1(e) only recedes because the normal load is applied at the centre of the contact as a concentrated force. If instead the normal load were distributed uniformly along the entirety of the upper surface of the upper body (i.e. the thin strip), the contact would remain complete.

treated as complete contacts.

To date, a majority of the research in fretting-fatigue has focused on the Hertzian contact geometry. This is probably because an analytical solution for the contact pressure distribution for this geometry (for the frictionless case) was discovered by Hertz [1882] over 130 years ago, and several other important solutions have also been obtained since then, e.g. [Cattaneo, 1938, Mindlin, 1949, Mindlin and Deresiewicz, 1953]. In addition, the smooth nature of the contact profile makes experiments using this geometry much less sensitive to minor misalignments in the test configuration, so the tests are much easier to conduct, control, and analyse. More recently, research has also focused on the flat-and-rounded geometry, e.g. [Rajasekaran and Nowell, 2006], since it can be used to model some of the joints that are used in gas-turbine engines, e.g. the dovetail joint.

Complete and receding contacts have received relatively less attention. However, certain critical components in gas-turbine engines result in complete contact, e.g. the splined couplings that are used to join the central split shafts in the engine. Limmer et al. [2001] note that pressure to design more fuel efficient aero-engines has caused these splined couplings to be subjected to increasingly severe conditions, and it is expected that these joints will be one of the main features that will limit shaft life in the future. This thesis is concerned with developing improved tools for the analysis of complete contacts such as these, which are subject to the severe combination of fretting and high-cycle fatigue.

1.3 Literature Review

In this section, a brief review of the literature is presented. First, the Coulomb friction model is briefly discussed since it is used for all of the frictional analyses in this thesis. Several important developments in the study of fretting fatigue are then briefly reviewed. A review of some important developments in contact mechanics and in the analysis of complete contacts is then presented. Finally, the process of *frictional shakedown* is discussed along with its application to determining the steady-state response of cyclically loaded frictional contacts.

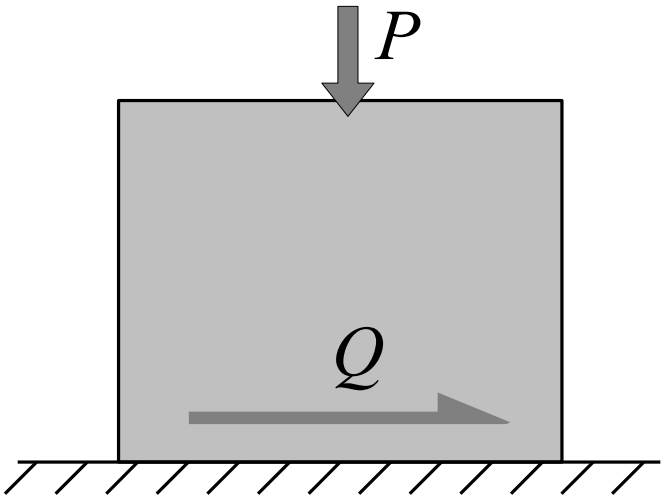


Figure 1.2: Elastic block in sliding frictional contact with a rigid surface.

1.3.1 Friction

Some of the earliest studies of contact between materials focused on friction. The laws of sliding friction were first discovered, but unpublished, by Leonardo da Vinci in the 15$^{\text{th}}$ century, and were then rediscovered in 1699 by Amontons [1699]. These laws were later confirmed and expanded on in 1785 by Coulomb [1785]. The friction model that is based on this early work is referred to as *Coulomb friction*, and it simply states that the frictional force is directly proportional to the normal load. Thus, for the sliding contact shown in Figure 1.2, it is assumed that $Q = fP$, where Q is the total shear force, P is the total normal force, and f is the coefficient of friction. Thus, Coulomb friction assumes the frictional force is independent of the contact's area and of its sliding velocity.

Although the Coulomb friction law is very simple and was developed over two centuries ago, it remains very popular and is still used in many practical contact analyses to this day, e.g. [Ding et al., 2007]. In fact, the mathematical simplicity of this friction law, while still providing a reasonably good approximation of the actual physical behaviour in many circumstances (at least in metals), is probably the main reason for its popularity. For these reasons, all of the frictional analyses performed in this thesis use the Coulomb model of friction.

At this point, it is important to distinguish between two regimes of frictional slip:

partial-slip and *gross-sliding*. When the Coulomb friction model was first developed, its intended application was for the gross-sliding regime, i.e. when the two bodies are in rigid-body motion with respect to one another. Conversely, the partial-slip regime refers to the case when some particles along the contact interface remain stuck throughout the loading regime, while other regions of the contact interface experience relative tangential displacement. Note that partial-slip results in fretting fatigue, whereas gross-sliding results in *fretting wear*. As wear generally does not result in failure from the propagation of fretting cracks, this thesis only concerns the partial-slip regime.

The only modification to the original law that is made here is that it is assumed to apply both at the macro level and also in a point-wise sense along the contact interface. In other words, at any point along the contact interface, x, that is in partial-slip, it is assumed that $q(x) = fp(x)$, where $q(x)$ is the shear traction distribution, and $p(x)$ is the contact pressure distribution. Clearly, this type of friction model can easily be extended to accommodate both spatial and temporal variations in the friction coefficient, and a dependence on various other parameters can also be introduced in this way (see e.g. Kondo et al. [2004]). However, such generalizations are not pursed in this thesis.

It should be noted that the inadequacy of the Coulomb model under some conditions has produced a great deal of interest in recent years to develop a better model of friction, which would be based on a better understanding of the underlying physical mechanisms that result in friction. However, developing such a model has proven to be an extremely challenging task, and a model that acts as a wholesale replacement for Coulomb friction has not yet been discovered. This may partly be because the mechanisms contributing to friction arise at several different length scales ranging from the grain level to the length scale associated with the roughness of the contact surface. In addition, friction is highly sensitive to the surface conditions, and this also contributes to the difficulty of developing a physics-based friction model.

1.3.2 Fretting fatigue

As previously mentioned in §1.1, fretting fatigue is a damage process that results when frictional contacts are subjected to partial-slip conditions as a result of fatigue loading. Although investigations into the process of plain fatigue began in the mid-1800's (see [Schütz, 1996]), Waterhouse [1992] notes that fretting was not reported in the literature until 1911. In this year, Eden et al. [1911] reported finding oxide particles on the surface of a steel fatigue specimen where the specimen was clamped to the fatigue test rig.

Waterhouse [1992] notes that the first tests that were performed specifically to investigate fretting were carried out in 1927 by Tomlinson [1927], who designed two fatigue test machines for this purpose. These early experiments revealed the important finding that fretting damage only occurred when partial-slip was present but that slip amplitudes as small as $0.1\mu m$ were sufficient to cause fretting. Although Tomlinson [1927] initially incorrectly attributed fretting to some type of chemical process, later work by Tomlinson et al. [1939] revealed that fretting is a purely mechanical process that is associated with partial-slip. According to Waterhouse [1992], the origin of the term 'fretting' (or more specifically 'fretting corrosion') can be attributed to this later paper by Tomlinson et al. [1939] in which the term 'fretting' appears in the title.

Warlow-Davies [1941] was the first researcher to investigate the influence of fretting on fatigue performance [Waterhouse, 1992]. To do this, Warlow-Davies [1941] conducted experiments in which he first subjected test specimens to fretting and then later to fatigue loading so as to isolate the influence of fretting (i.e. to avoid simultaneous fretting and fatigue damage). Warlow-Davies [1941] performed tests using both a medium-carbon steel and a nickel-chromium-molybdenum alloy steel and found that fretting reduced the fatigue life of the specimens by 13% and 18%, respectively. Waterhouse [1992] notes that more experiments were carried out over the next several years, many of them by McDowell (e.g. [McDowell, 1953, McDowell et al., 1954]), that examined the combined effect of fretting and fatigue, and these tests revealed much more significant reductions in

fatigue life (i.e. exceeding 50%). Another very important finding was made in 1958 by Fenner and Field [1958], who discovered that fretting speeds up the rate at which cracks initiate [Waterhouse, 1992].

The next major advance came from a set of 6 papers published by Nishioka and Hirakawa, viz. [Nishioka and Hirakawa, 1969a,b,c,d, 1972, Nishioka et al., 1968], which present a very systematic and fundamental investigation of fretting fatigue. In the first paper, a shrink fit contact that is subjected to a cyclic rotating bending load is examined to investigate the reduction in fatigue strength due to fretting in press-fit railway axle assemblies. Some of the results of these tests include that the slip amplitude decreased with the number of load cycles and that it was unaffected by the frequency of the cyclic loading.

In the following 5 papers, Nishioka and Hirakawa performed a more fundamental investigation using a special fretting-fatigue test rig that they devised. A very important feature of this test rig was that it used cylindrical contact pads, and thus a Hertzian contact geometry, which made the contact behaviour much easier to analyse and control. Some of the factors investigated in these papers include the effects of: slip amplitude, coefficient of friction, surface pitting, non-propagating cracks, mean bulk stress, surface treatment, contact pressure, and surface hardness.

One particularly interesting result reported in [Nishioka and Hirakawa, 1969a] is that fatigue life was most greatly reduced at an intermediate level of relative slip displacement (approximately $20\mu m$ in this case). This is because fretting damage is reduced both for small slip amplitudes (because minimal damage accumulates) and for large slip amplitudes (because wear becomes the dominant damage process). Another useful result reported by Nishioka and Hirakawa [1969c] is that while the mean bulk stress did have a major influence on the propagation of existing cracks, it did not significantly influence the initiation of fretting cracks.

Another important result was discovered by Hoeppner and Goss [1974], who performed several sets of fretting-fatigue tests on two aluminium alloys and demonstrated the existence of a *fretting damage threshold*. In other words, they found that if tests were

conducted in which the "fretting source" (i.e. the contact pads) was removed after some number of load cycles, there was a number of cycles below which the fatigue life of these specimens became no different to that of plain-fatigue specimens. Specifically, Hoeppner and Goss [1974] found that if the contact pads were removed within 20% of the number of cycles corresponding to the plain-fatigue life of the specimen, fretting did not influence the fatigue life of the specimen. This result led Hoeppner and Goss [1974] to conclude that fretting acts mainly "as a flaw generator" and that this is the main mechanism by which fretting reduces fatigue life, i.e. via speeding crack initiation.

Similar results were also reported by Endo and Goto [1976], who found a fretting damage threshold of 20-25% for tests conducted on carbon steel. However, these authors also found that fretting not only speeds crack initiation but also speeds up the early part of the crack propagation phase. A short discussion article by Hoeppner [1977] and the response by Endo and Goto [1977] further illuminate the effect of fretting on early crack propagation. These results helped form the modern understanding that the fretting-fatigue life of a specimen can be divided into three phases:

1. A crack nucleation/initiation phase in which the presence of fretting significantly accelerates the initiation of cracks.

2. An early crack propagation phase in which the contact stresses strongly influence and accelerate the propagation of existing fatigue cracks.

3. A late crack propagation phase in which the crack propagation rate is controlled by the alternating bulk stresses in the specimen and is no longer influenced by the presence of fretting.

Hills et al. [1988] note that once a crack has developed and is in the propagation phase, its propagation rate is generally well predicted using Paris' law, although extra care must be taken to account for the contact stresses when determining the stress intensity factor for the early propagation phase. Conversely, the crack initiation phase is much less well understood and is a topic where much of the current research effort is being focused.

The two main approaches for quantifying the crack nucleation phase are: *critical-plane* and *short-crack* approaches. Critical-plane approaches involve trying to find empirical combinations of parameters that can be correlated against crack initiation, and some of the more commonly used parameters are due to Ruiz et al. [1984], Smith et al. [1970], and [Fatemi and Socie, 1988]. Conversely, the short-crack approach assumes that the threshold stress intensity factor for propagating a fretting crack is dependent on the size of the crack and that shorter cracks require a smaller stress intensity factor range to propagate. The initial work in this area is due to Kitagawa and Takahashi [1976], but later extensions were presented by El Haddad et al. [1979a,b].

Another important result, which was first discovered by Bramhall [1973] and later confirmed and expanded on by Nowell [1988], is the existence of a 'size effect' for fretting-fatigue tests. Both of these authors conducted tests using cylindrical contact pads in which the size of the contact was varied, but the contact traction distribution (and hence the zones of slip and stick) was kept identical. It was found that contacts below a critical size did not fail within 10^7 load cycles, whereas all contacts greater than this size failed. Hills [1994] points out that these results "illustrate a completely unexpected dimensional dependence" on the size of the contact, which has not yet been convincingly explained.

The discovery of this size effect is also important because it provided part of the motivation for the development of asymptotic approaches, which provide parameters that implicitly incorporate a length dimension. Recent work by Hills et al. [2012b] presents several asymptotic forms that can be used to correlate against the fretting-fatigue life of both Hertzian and flat-and-rounded contacts, which collapse the test results for contacts of different sizes (and for both types of contacts) onto the same fretting-fatigue life curve. These asymptotic parameters also accurately predict the boundary between finite and infinite life for contacts of different sizes. In a sense, much of the work in §2-§6 of this thesis could be viewed as an attempt to use an analogous asymptotic approach to quantify the fatigue life of complete contacts.

Finally, it should be mentioned that a majority of the experimental studies of fret-

ting fatigue since the 1970's have used Hertzian contact pads due to their many benefits, which have been described previously. Although some fundamental investigations into the fretting fatigue behaviour of complete contacts have been carried out, e.g. [Juoksukangas et al., 2013, Kondo and Bodai, 2001, Kondo et al., 2004, 2005, Mugadu, 2002, Noraphaiphipaksa et al., 2013, Sabsabi et al., 2011], such investigations have only begun fairly recently. It should also be noted that fretting-fatigue tests have also been conducted for component specific geometries and loading regimes, e.g. [Araújo and Nowell, 2009, Ding et al., 2007, Gean and Farris, 2009, Rajasekaran and Nowell, 2006], especially for the dovetail joints and spline couplings in gas-turbine aero-engines.

1.3.3 Contact mechanics

The study of fretting fatigue has been greatly aided by theoretical advances in the field of contact mechanics. Indeed, many of these advances were developed to address specific unresolved engineering problems that were encountered at the time. In the early days, much of this work was motivated by rolling contact problems encountered in the rail industry. According to Knothe [2008], the first person in Germany to study the stresses resulting from wheel-rail contact was Redtenbacher [1855], who published a book that considered some of these problems in 1855. However, the birth of the field of contact mechanics is widely attributed to Hertz [1882], who published the solution for the stresses resulting from contact between two elastic spheres in 1882. To do this, Hertz [1882] idealized both bodies as smooth (i.e. frictionless) elastic half-spaces with quadratic surface profiles and obtained the resulting contact pressure distribution.

Several other important solutions for semi-infinite elastic domains (i.e. half-planes or half-spaces) were also obtained shortly thereafter. For instance, Boussinesq [1885] published the solution for the stress state in an elastic half-space resulting from the application of a point load and also for a cylindrical punch indenting a half-space. Additionally, Flamant [1892] published the solution for a concentrated line force applied to the sharp corner of a semi-infinite elastic wedge. This solution is of fundamental importance be-

cause the special case when the wedge angle is 180° corresponds to a line force applied to a half-plane, which can be used as a Green's function for solving problems in which a half-plane is subjected to more complex load distributions. For example, Barber [2009] §12 shows that the solution for a rigid punch of arbitrary (but known) profile in frictionless contact with a half-plane can be obtained using this approach.

A hugely important result for the study of fretting fatigue was the partial-slip solution for frictional contact between a cylinder and an elastically similar half-plane subjected to a static normal load and an oscillatory shear load, which was independently published by Cattaneo [1938] and Mindlin [1949]. Solutions for more complex load cases were also developed a few years later by Mindlin and Deresiewicz [1953]. Another very important extension to the original partial-slip solution was obtained by Hills et al. [1988], who showed how the contact tractions are modified by the presence of a superimposed bulk tension. This solution is clearly helpful for analysing fretting-fatigue experiments since a bulk load is usually applied to the specimen.

More recently, other important generalizations of solutions based on half-plane theory have been published by Jäger [1998] and Ciavarella [1998a,b]. These results have shown how partial-slip solutions may be developed for contacts of any profile to which a half-plane idealization can be applied. This enabled Ciavarella et al. [1998] to obtain the partial-slip solution for contact between a flat-and-rounded punch and a half-plane, which has important applications for modelling dovetail connections in gas-turbine engines.

1.3.4 Analysis of complete contacts

The previous section briefly mentions some contact solutions that have been developed, which are very helpful for studying fretting fatigue. However, a feature that is common to most of these results is their application of half-plane theory, i.e. they assume both bodies may be idealized as elastic half-planes. Although very useful results have been obtained for both incomplete contacts and flat-and-rounded contacts using this approach, even the simplest complete contacts cannot be treated in this way because the contact

defining body cannot reasonably be idealized as a half-plane.

Consider the two-dimensional contact problem of an elastic rectangle pressed into an elastically similar half-plane in the presence of friction. For this problem, the contact stresses will depend on the finite dimensions of the rectangle, i.e. on the ratio of its length to its height, and a closed-form solution for even this simple case has not yet been discovered. However, some results have been obtained for certain special cases. For example, Khadem and O'Connor [1969a,b] considered the problem of an elastic square being compressed by two elastically similar half-planes and of an elastic circular cylinder compressed by two elastically similar half-spaces. These authors obtained numerical solutions for both of these contact problems for the case of a fully adhered contact interface and for a frictionless interface.

More recently, Mugadu and Hills [2003] obtained the solution for a rigid punch pressed into an incompressible elastic half-plane subjected to a constant normal load, an oscillatory shear load, and an oscillatory bulk tension in the half-plane. However, the solution technique employed by these authors can only be used for this special combination of material properties, and deviations from this special case, e.g. the elastically similar problem discussed above, result in very different contact behaviour. Thus, in many cases, a purely numerical approach must be used, e.g. the finite element method.

An alternative approach is to notice that fretting damage invariably accumulates near the edges of the contact interface, so it is the contact stresses and behaviour in this region that control the accumulation of fretting damage. Thus, asymptotic forms can be applied to obtain an accurate description of this region. The concept behind this approach is to focus attention very close to the contact edge (as shown in Figure 1.3), such that all other geometric features can be considered to be remote. And if the contact edge is assumed to be perfectly sharp, the stress state can be represented using elasticity solutions for semi-infinite notches, e.g. for contact between a half-plane and a quarter-plane.

The pioneering work in this field was published in 1952 by Williams [1952], who found the solution for a semi-infinite elastic wedge subjected to in-plane, remote loading.

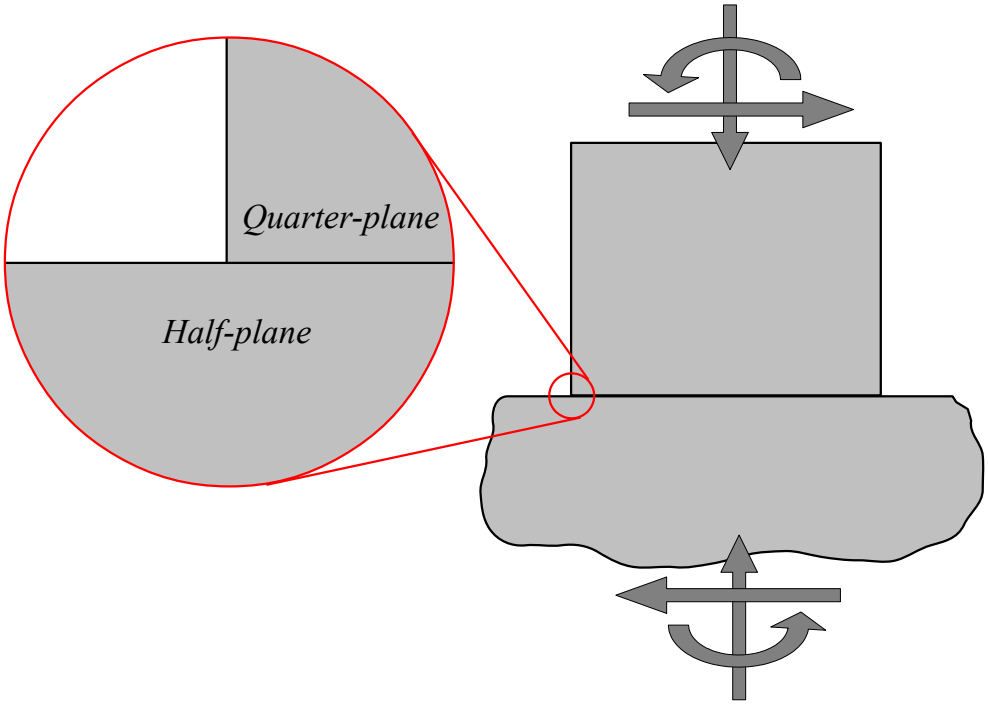


Figure 1.3: Region at the edge of a complete contact where asymptotic forms can be applied.

Notice that if the interface of the contact shown in Figure 1.3 is assumed to be fully adhered, the displacement state across the contact interface is continuous. Hence, if the bodies are elastically similar, the influence of the contact interface can be ignored, the contact edge can be idealized as a 270° notch, and Williams' solution can be used to represent the stress state in this region. As much of the work in this thesis is based on Williams' solution, it is examined in detail in §2.

Another important solution was obtained in 1968 by Bogy [1968] for fully adhered contact between elastically dissimilar wedges. In the following year, Dundurs [1969] published a short discussion of [Bogy, 1968] that showed that the solution could be simplified considerably by taking advantage of a reduced dependence on the elastic constants for this problem. An improved solution was later published by Bogy [1971] based on this exchange. Shortly thereafter, the solution for sliding frictional contact between a wedge and an elastically dissimilar half-plane was published by Gdoutos and Theocaris [1975] and Comninou [1976].

Some initial work on applying these asymptotic forms to analyse elastically similar complete contacts was presented by Mugadu and Hills [2002] and was also considered in Mugadu's doctoral thesis [Mugadu, 2002]. This approach was also pursued and extended by Churchman [2006] and Karuppanan [2007] in their doctoral research. For instance, Churchman and Hills [2006b] studied the behaviour of an elastic square pressed onto an elastically similar half-plane subjected to a constant normal load and an oscillatory shear load. These authors used a combination of detailed finite element modelling and notch asymptotes to provide additional detail near the contact edge. Karuppanan and Hills [2008a] later extended the results for this case and also analysed two other trapezium-shaped complete contacts in this way.

More recently, the *distributed dislocation technique*, which is described in detail by Hills et al. [1996], has been applied to analyse the edge of complete contacts in an asymptotic sense. In order to apply this technique, the kernel for a dislocation in the relevant domain must first be known. For a square punch in contact with an elastically similar half-plane, a three-quarter plane is the relevant domain, and Churchman et al. [2006] published the solution for this case. Churchman and Hills [2006a] then showed how this solution can be applied to obtain the stress intensity factors for a crack arising at the edge of a three-quarter plane. Churchman and Hills [2006c] later used this approach to calculate the extent of the slip zone arising at the edge of a complete contact. More recently, Paynter et al. [2010] also used this approach to calculate the extent of slip and separation arising at the contact edge for this same geometry.

1.3.5 Steady-state behaviour

Contact problems involving friction are inherently non-linear and history dependent. However, it is frequently observed that cyclically loaded contacts tend toward a steady-state response after a sufficient number of load cycles, e.g. [Banerjee and Hills, 2006, Churchman and Hills, 2006b, Ding et al., 2007]. In addition, in many cases, only a very small number of cycles is required before even very complex contacts reach a steady-state

response, e.g. Ding et al. [2007] report that 6 cycles were sufficient for a full model of a spline coupling with numerous contacts to reach a steady state. Therefore, in the context of high-cycle fatigue, it is reasonable to assume that such contacts spend effectively their entire fatigue life in approximately steady-state conditions. Consequently, it is of practical interest to develop efficient techniques for calculating the steady-state response of cyclically loaded frictional contacts (even if this means ignoring the initial transient behaviour), and this is examined in §9.

The process by which frictional contacts tend toward a steady-state response is referred to as *frictional shakedown*, which occurs when frictional slip modifies the contact tractions in such a way as to reduce the tendency for further slip in subsequent cycles. This concept has been derived by analogy from the concept of *elastic shakedown* in the theory of plasticity, where elastic shakedown refers to the process by which yielding in the initial load cycles results in the self-development of (internal) residual stresses that inhibit further yielding in subsequent load cycles, bringing the loading trajectory back within the elastic limit. Research on frictional shakedown has only been carried out quite recently, and much of this initial work has focused on determining whether results from plasticity theory can also be applied to frictional problems.

The main results of interest for frictional problems are those relating to the *elastic shakedown limit*, which is the amplitude of cyclic load above which cyclic plasticity is certain to persist in the steady state. The key result in this area is the lower-bound shakedown theorem proposed by Melan [1936]. This theorem might be stated as: if *any* self-equilibrating set of residual stresses can be found that, when added to the loading trajectory, will prevent yielding throughout the cycle, then the body will *always* shake down under the given loading trajectory, regardless of the residual stresses present at the onset of loading. Several years later, Symonds [1951] simplified Melan's proof, and Koiter [1960] subsequently proved an upper-bound elastic shakedown theorem. These results proved that if shakedown to a non-dissipative steady state is *possible*, it is actually *guaranteed* to occur, i.e. the steady-state response is independent of the initial residual stress state.

In the past decade, there has been much interest and speculation as to whether a frictional analogue of Melan's theorem could be applied to complete contacts, e.g. [Banerjee and Hills, 2006, Churchman and Hills, 2006b]. This question has only been resolved quite recently by Klarbring et al. [2007] for discrete contact problems and by Barber et al. [2008] for continuum contact problems. These authors found that Melan's theorem does not apply generally to frictional problems, though it does apply in certain special cases. Specifically, Melan's theorem only applies to *uncoupled contacts*: that is, to contacts for which relative tangential (i.e. slip) displacements do not modify the contact pressure distribution. Very recently, Barber [2011] proposed the conjecture that Melan's theorem is a special case of a more general theorem that applies to uncoupled contacts, i.e. that several other properties of the steady-state solution are also independent of initial conditions. This conjecture was proven by Barber et al. [2011] for all uncoupled continuum contact problems between bodies that can be idealized as half-planes.

1.4 Thesis summary

In the last century, fretting fatigue has gone from being an entirely unstudied phenomenon within the scientific community to a well known damage process that is taken into account when designing most critical engineering components, especially those subject to high-cycle fatigue. However, there still remain several areas of study that are not sufficiently well understood, and these result in increased economic costs due to the highly empirical approach that must be adopted to design against fretting failures. As previously noted, the fretting life of a component can be divided into a crack nucleation phase and a crack propagation phase, and it is the former whose quantification remains elusive.

A review of the literature also reveals that the majority of modern studies in fretting fatigue have focused on the Hertzian contact geometry. This work has provided invaluable fundamental insight into the nature of fretting fatigue. However, much less detailed analysis has been performed on complete contacts. The aim of this thesis is to address this issue by developing several efficient calculations for quantifying the contact behaviour of

complete contacts. The philosophy underlying this approach is that the damage that accumulates due to fretting must be controlled by the contact and material behaviour arising in the region where fretting occurs; therefore, a more accurate description of this region should provide better parameters to correlate against fretting performance.

This thesis approaches the analysis of complete contacts using two distinct methodologies. The first approach is outlined in §2-6 and focuses on developing a detailed description of the complex stress state that arises in the vicinity of sharp corners. These chapters employ the asymptotic approach that is briefly discussed in §1.3.4, which enables the stresses that arise near sharp corners to be characterized accurately and efficiently. This approach is useful because failure invariably initiates near the contact edges, but the stresses in this region are very difficult to quantify accurately using a purely numerical approach. Moreover, this analysis provides parameters that not only incorporate a measure of the severity of the stress state but also an intrinsic length scale, which enables them to account for size effects and to provide better correlations against fatigue life.

The second approach is detailed in chapters §7-9, which focus on analysing the contact interface as a whole using a purely numerical approach, though the analysis is sometimes guided by the asymptotic results obtained previously. While this numerical approach cannot describe the critical contact-edge region nearly as accurately as the previous asymptotic approach, it has two important benefits: (i) it can describe the entire contact interface, including regions where the stress state is determined by the specific geometry of the component; and (ii) it enables the behaviour of cyclically loaded complete contacts to be studied using very powerful optimization techniques that simply cannot be applied within the asymptotic framework presented in §2-6. In addition, the results of this optimization approach are particularly useful for studying the energy dissipated due to friction by cyclically loaded complete contacts, and this has very important practical consequences for engineering components.

Although the two approaches adopted in this thesis are distinct, they are complimentary and should be used together to analyse complete contacts that are subject to fretting fa-

tigue. The former approach can be used to quantify contact-edge behaviour accurately and to provide better parameters to correlate against fatigue life. The latter provides a powerful tool for studying frictional energy dissipation in cyclically loaded contacts, which has a significant influence on a contact's overall behaviour. Hence, both approaches should be applied jointly to provide a full description of the contact's behaviour and to predict its fatigue performance.

A brief summary of each chapter in this thesis is provided below:

- In §2, Williams' solution for a semi-infinite notch subjected to in-plane, remote loading is presented, and some of its important properties are discussed. This solution forms the basis of the analysis in §3-§6.

- In §3, Williams' solution is used to characterize the mode-mixity and small-scale yielding of the *process zone* arising at the tip of sharp V-notches. These calculations are then applied to mixed-mode experimental data reported in the literature.

- In §4, it is shown that Williams' solution can be used to analyse the edge of complete contacts and that it provides very detailed information on when the contact edge will be fully stuck. Fretting tests reported by Juoksukangas et al. [2013] are then analysed using this approach.

- In §5, it is shown that Williams' solution can be used to obtain an estimate of the extent of partial-slip (and also separation) arising at the contact edge, provided that the slip (or separation) zones are only moderate in extent. These calculations are also applied to the fretting tests reported by Juoksukangas et al. [2013].

- In §6, the tools developed in §2-§5 are applied to analyse three sets of fretting-fatigue tests reported in the literature by Juoksukangas et al. [2013], Mugadu [2002], and Noraphaiphipaksa et al. [2013].

- In §7, a numerical *sub-structuring* procedure is presented, which enables transient (i.e. 'marching-in-time') simulations of quasi-statically loaded, two-dimensional,

linear-elastic frictional contacts to be run over 10 times more quickly than in the commercial finite element software ABAQUS/CAE.

- In §8, the transient contact behaviour arising when a square elastic block is pressed into an elastically similar half-plane and then sheared until it begins sliding is examined.

- In §9, an efficient numerical approach is presented for calculating the frictional shakedown limit of complete contacts subjected to cyclic loading. This calculation is then applied to an example problem, and the influence of initial conditions on frictional energy dissipation in the steady state is investigated.

- In §10, the main results from each chapter are summarized, and some topics requiring further investigation are discussed.

Chapter 2

Semi-infinite notch theory

In this chapter, the elasticity solution developed by Williams [1952] for a semi-infinite notch subjected to in-plane, remote loading is presented. Since this theory is central to the analysis of sharp contacts, and hence to this thesis, this chapter begins with a derivation of Williams' solution in its standard form. Some key features of this solution are also briefly discussed. Finally, an alternative set of parameters introduced by Hills and Dini [2011] are examined, which are particularly useful for analysing mixed-mode loading conditions (i.e. those that excite both the symmetric and the anti-symmetric eigensolution).

2.1 Motivation

It is well known that sharp features result in very high stresses in their vicinity, and engineers generally try to avoid them when designing components for this reason. Still, in some instances, sharp features do arise, and it is necessary to obtain an accurate assessment of the stress state they induce. However, due to the high stresses and steep stress gradients they produce, it is notoriously difficult to do this using only numerical methods, e.g. the finite element method, unless a very fine mesh is used. Asymptotic approaches provide an accurate description of the stress state in this region, while avoiding the usual numerical issues associated with a standard numerical approach.

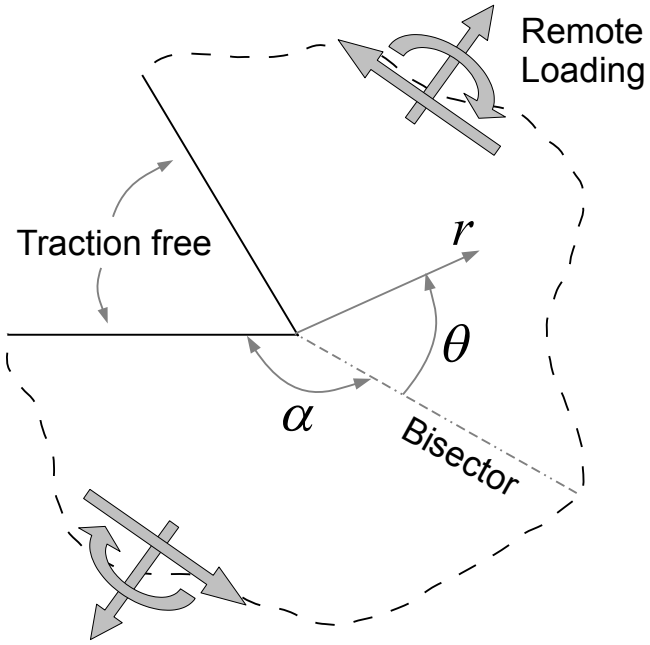


Figure 2.1: Semi-infinite notch with traction-free faces subjected to remote loading.

2.2 Williams' solution

Consider the linear-elastic, semi-infinite notch shown in Figure 2.1, which has a total angle included within the solid of 2α, where α is the angle from the bisector of the notch to the notch faces. Williams [1952] presented the solution to this problem when the notch is subjected to the following boundary conditions: (i) the faces of the notch are traction free, and (ii) some arbitrary (in-plane) surface tractions are applied remotely (i.e. from infinity).

From a practical point of view, this semi-infinite problem can be used to model the region near the tip of a notched component, such as that shown in Figure 2.2. A useful way to understand this is explained by Barber [2009] §11.2.2, where he invites the reader to imagine looking at the notch tip with a very high-powered microscope, which can zoom-in to an arbitrarily small length scale. At a high enough magnification, the nearest geometric feature will be distant from the region of focus. And since the notch causes a stress intensification, it will have an increasing influence on the stress state as the sharp notch tip is approached. Thus, very near to the notch tip, the stress state will be dominated

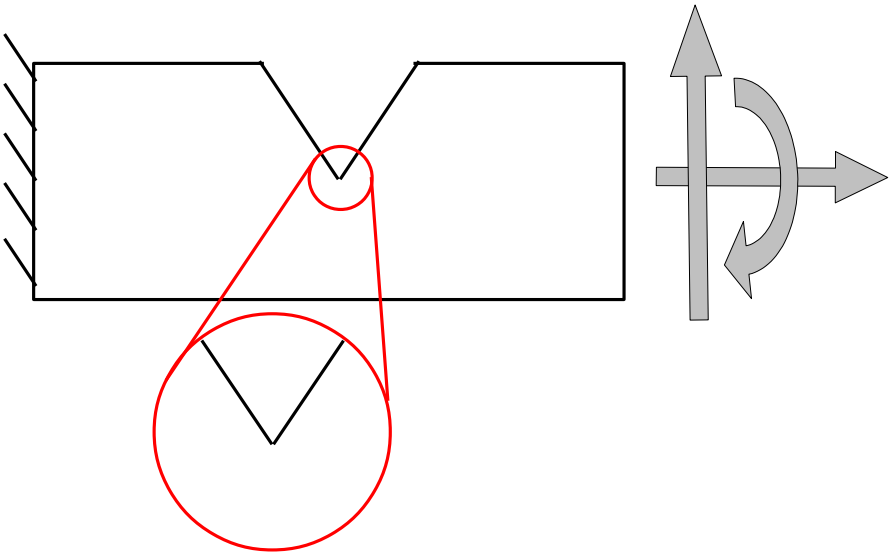


Figure 2.2: Asymptotic region in a notched component.

by the stress intensification associated with the notch, and the influence of the remote geometry will be negligible. Hence, in this region, Williams' semi-infinite solution can be used to describe the stress state accurately and efficiently.

2.2.1 Stress function

A convenient way to obtain a solution for the elastic state of stress and displacement for the notch shown in Figure 2.1 is to use the *Airy stress function* approach (see Barber [2009] §4.3). This enables the stresses to be determined by differentiating a higher-order scalar potential function, which is referred to as a *stress function*, ψ. For this analysis, it is convenient to adopt the polar coordinate set (r, θ) defined at the notch tip, which is shown in Figure 2.1. With this coordinate system, Airy's equations for the stresses are

$$\sigma_{rr}(r,\theta) = \frac{1}{r}\frac{\partial \psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \psi}{\partial \theta^2} \tag{2.1a}$$

$$\sigma_{r\theta}(r,\theta) = \frac{1}{r^2}\frac{\partial \psi}{\partial \theta} - \frac{1}{r}\frac{\partial^2 \psi}{\partial r \partial \theta} \tag{2.1b}$$

$$\sigma_{\theta\theta}(r,\theta) = \frac{\partial^2 \psi}{\partial r^2}. \tag{2.1c}$$

The first step in this analysis is to determine an appropriate stress function. In his classical solution, Williams assumes that the stresses are of a separated variable form and exhibit a power-law variation with r, which leads to the following stress function

$$\psi(r,\theta) = r^{\lambda+1} f(\theta), \tag{2.2}$$

where $f(\theta)$ describes the currently unknown θ-dependence of the stress state, and λ controls the power-law variation with r.

Since the Airy stress function is being used, the stresses resulting from substituting ψ into (2.1) will automatically satisfy equilibrium. However, the stress function, ψ, must also satisfy the compatibility condition, which is given by the *biharmonic equation*:

$$\nabla^4 \psi = 0. \tag{2.3}$$

In polar coordinates, this results in the following expression

$$\left\{ \frac{\partial}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \right\}^2 \left\{ \frac{\partial}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \right\}^2 \psi = 0. \tag{2.4}$$

Substituting (2.2) into (2.4), performing the required differentiations, and simplifying the result gives

$$r^{\lambda-3} \left\{ (\lambda+1)^2(\lambda-1)^2 f(\theta) + [(\lambda+1)^2 + (\lambda-1)^2] \frac{\partial^2 f(\theta)}{\partial \theta^2} + \frac{\partial^4 f(\theta)}{\partial \theta^4} \right\} = 0, \tag{2.5}$$

which can be further simplified to

$$\left\{ \frac{\partial^2}{\partial \theta^2} + (\lambda+1)^2 \right\} \left\{ \frac{\partial^2}{\partial \theta^2} + (\lambda-1)^2 \right\} f(\theta) = 0. \tag{2.6}$$

Thus, for $f(\theta)$ to satisfy the compatibility condition, it must satisfy the above expression. For $\lambda \neq \{-1, 1\}$, there are four solutions, which can be combined using superposition to

give

$$f(\theta) = \begin{cases} A_1 \cos[(\lambda+1)\theta] + A_2 \cos[(\lambda-1)\theta] \\ \quad +A_3 \sin[(\lambda+1)\theta] + A_4 \sin[(\lambda-1)\theta], \end{cases} \tag{2.7}$$

where A_1, A_2, A_3, A_4 are unknown constant coefficients. Therefore, the resulting stress function is

$$\psi = r^{\lambda+1} \begin{pmatrix} A_1 \cos[(\lambda+1)\theta] + A_2 \cos[(\lambda-1)\theta] \\ \quad +A_3 \sin[(\lambda+1)\theta] + A_4 \sin[(\lambda-1)\theta] \end{pmatrix}. \tag{2.8}$$

The three stress components can then be obtained from (2.1) by differentiation as

$$\sigma_{rr} = \lambda r^{\lambda-1} \begin{pmatrix} -A_1(\lambda+1)\cos[(\lambda+1)\theta] - A_2(\lambda-3)\cos[(\lambda-1)\theta] \\ \quad -A_3(\lambda+1)\sin[(\lambda+1)\theta] - A_4(\lambda-3)\sin[(\lambda-1)\theta] \end{pmatrix} \tag{2.9a}$$

$$\sigma_{r\theta} = \lambda r^{\lambda-1} \begin{pmatrix} A_1(\lambda+1)\sin[(\lambda+1)\theta] + A_2(\lambda-1)\sin[(\lambda-1)\theta] \\ \quad -A_3(\lambda+1)\cos[(\lambda+1)\theta] - A_4(\lambda-1)\cos[(\lambda-1)\theta] \end{pmatrix} \tag{2.9b}$$

$$\sigma_{\theta\theta} = \lambda r^{\lambda-1} \begin{pmatrix} A_1(\lambda+1)\cos[(\lambda+1)\theta] + A_2(\lambda+1)\cos[(\lambda-1)\theta] \\ \quad +A_3(\lambda+1)\sin[(\lambda+1)\theta] + A_4(\lambda+1)\sin[(\lambda-1)\theta] \end{pmatrix}. \tag{2.9c}$$

It can be seen from (2.9c) that A_1, A_2 result in symmetric $\sigma_{\theta\theta}$, whereas A_3, A_4 result in anti-symmetric $\sigma_{\theta\theta}$. Thus, a loading condition that is purely symmetric (i.e. mode I) in character with respect to the notch bisector (i.e. $\theta = 0$) excites only the symmetric terms, whereas a purely anti-symmetric (i.e. mode II) loading condition excites only the anti-symmetric terms. Therefore, the stresses, σ_{ij}, where $i, j \in \{r, \theta\}$, can be separated into symmetric and anti-symmetric terms as

$$\sigma_{ij} = \sigma_{ij}^{I} + \sigma_{ij}^{II}, \tag{2.10}$$

where σ_{ij}^{I} includes A_1, A_2, and σ_{ij}^{II} includes A_3, A_4.

2.2.2 Boundary conditions

Although the form of the stress function has been determined, the four coefficients and the exponent λ are still unknown. However, these can, in part, be determined by requiring

that they give a stress function that satisfies the boundary conditions. Here, these are that the faces of the notch must be traction free, i.e.

$$\sigma_{r\theta}(r, \pm\alpha) = 0 \tag{2.11a}$$

$$\sigma_{\theta\theta}(r, \pm\alpha) = 0, \tag{2.11b}$$

where α is the angle shown in Figure 2.1.

Substituting (2.9b),(2.9c) into (2.11) gives a set of four homogeneous equations, which can be conveniently written in matrix form as in Churchman [2006] §2.2.3 as

$$\boldsymbol{M}\boldsymbol{A} = \boldsymbol{0}, \tag{2.12a}$$

or

$$\lambda r^{\lambda-1} \begin{bmatrix} a & b & c & d \\ -a & -b & c & d \\ e & f & g & h \\ e & f & -g & -h \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} = \boldsymbol{0}, \tag{2.12b}$$

where

$$a = (\lambda+1)\sin[(\lambda+1)\alpha] \qquad b = (\lambda-1)\sin[(\lambda-1)\alpha] \tag{2.13a}$$

$$c = -(\lambda+1)\cos[(\lambda+1)\alpha] \qquad d = -(\lambda-1)\cos[(\lambda-1)\alpha] \tag{2.13b}$$

$$e = -c \qquad f = (\lambda+1)\cos[(\lambda-1)\alpha] \tag{2.13c}$$

$$g = a \qquad h = (\lambda+1)\sin[(\lambda-1)\alpha]. \tag{2.13d}$$

This is a standard eigenvalue problem, and for it to have a non-trivial solution, the determinant of $\boldsymbol{M}$ must be equal to zero, i.e. $\det(\boldsymbol{M}) = 0$.[1] Since λ is a common multiple of each element of $\boldsymbol{M}$, $\lambda = 0$ is an eigenvalue for all values of α. The other common

[1] If $\det(\boldsymbol{M}) \neq 0$, then $\boldsymbol{M}$ can be inverted, so the only solution is $\boldsymbol{A} = \boldsymbol{0}$.

multiple, $r^{\lambda-1}$, is only zero for particular values of r, so it can be divided out and ignored.

By performing basic operations, $\boldsymbol{M}$ can be simplified to

$$\begin{bmatrix} a & b & 0 & 0 \\ e & f & 0 & 0 \\ 0 & 0 & c & d \\ 0 & 0 & g & h \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} = \mathbf{0}, \tag{2.14}$$

which shows that $\boldsymbol{M}$ uncouples to give two independent systems of equations:

$$\begin{bmatrix} (\lambda+1)\sin[(\lambda+1)\alpha] & (\lambda-1)\sin[(\lambda-1)\alpha] \\ (\lambda+1)\cos[(\lambda+1)\alpha] & (\lambda+1)\cos[(\lambda-1)\alpha] \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \mathbf{0} \tag{2.15a}$$

$$\begin{bmatrix} (\lambda+1)\cos[(\lambda+1)\alpha] & (\lambda-1)\cos[(\lambda-1)\alpha] \\ (\lambda+1)\sin[(\lambda+1)\alpha] & (\lambda+1)\sin[(\lambda-1)\alpha] \end{bmatrix} \begin{bmatrix} A_3 \\ A_4 \end{bmatrix} = \mathbf{0}, \tag{2.15b}$$

where (2.15a) relates to the symmetric terms, and (2.15b) relates to the anti-symmetric terms. The determinant of the symmetric system leads to the following characteristic equation

$$\lambda_I \sin(2\alpha) + \sin(2\lambda_I\alpha) = 0, \tag{2.16a}$$

and the anti-symmetric system gives

$$\lambda_{II} \sin(2\alpha) - \sin(2\lambda_{II}\alpha) = 0, \tag{2.16b}$$

where λ is given the subscript I or II to denote whether it corresponds to the mode I or the mode II solution. Hence, the exponent for the σ_{ij}^{I} term in (2.10) is given by solutions of (2.16a), and the exponent for the σ_{ij}^{II} term is given by solutions of (2.16b).

2.2.3 Relevant terms

It is clear that both characteristic equations in (2.16) generally have multiple real solutions. In fact, Barber [2009] §11.2.3 notes that if complex solutions are permitted, there

are infinitely many solutions for all values of α.[2] However, Williams showed that the stress state can be expressed as a series expansion, where the value of λ for each term in the series is an eigenvalue of (2.16a) or (2.16b). As this series extends over an infinite number of terms, it is important to recall the motivation for using this solution in the first place: namely, to describe the stress intensification near the tip of sharp notches. Thus, the most relevant terms in this series expansion are those that have the largest influence on the stress state for small values of r. A truncated series containing only these dominant terms can then be used to describe the notch tip region.

It can be seen from (2.9) that the stresses are proportional to $r^{\lambda-1}$. Therefore, for each term in the series expansion, the stress state's dependence on r can take three distinct forms:

1. If $\lambda < 1$, the exponent on r is negative, so an elastic stress singularity is implied as $r \to 0$. Furthermore, a smaller value of λ results in a stronger singularity.

2. If $\lambda = 1$, the exponent on r is 0, so the stresses are invariant with r and are constant as $r \to 0$.[3]

3. If $\lambda > 1$, the exponent on r is positive, so the stresses vanish as $r \to 0$.

As the notch tip is approached (i.e. as $r \to 0$), the singular terms in the series will become arbitrarily more influential than all the other non-singular, higher-order terms. Moreover, as r becomes arbitrarily small, the term with the strongest order of singularity

[2]This is a consequence of the fundamental theorem of algebra, which can be stated as: "any complex polynomial must have a complex root" [Fine and Rosenberger, 1997] p.1. This theorem can also equivalently be stated as that every polynomial of degree n has exactly n roots (including roots of *multiplicity*, i.e. if repeated roots are counted as separate roots) [Needham, 1997] p.354. Since $sin(x)$ can be represented as a polynomial of infinite degree using a Taylor series expansion, it follows that (2.16a), (2.16b) have infinitely many roots. Note that complex roots can be found numerically, but it is difficult to find them graphically. This is because plotting a complex function requires two complex dimensions, which corresponds to four real dimensions, and it is quite difficult to visualize a four-dimensional object.

[3]As Barber [2009] §11.2.3 points out, this is a spurious eigenvalue for most notch angles since the corresponding eigenfunction takes on a null form. The exception (when this eigenvalue gives a non-trivial solution) is for the anti-symmetric solution when $\alpha = 257.4°$. A modified stress function is required to determine the corresponding eigenfunction since (2.7) does not include the solutions of (2.6) corresponding to $\lambda = 1$. However, this analysis is not pursued here since it is of minimal practical relevance.

will become arbitrarily more influential than any other less strongly singular terms. Thus, the eigenvalues resulting in the strongest stress singularities are the most relevant.

This raises the question of whether a singular state of stress is physically admissible, and if so, under what conditions. The reader is directed to Barber [2009] §11.2.1 for an in depth discussion of this question. One way to address it is to note that a singular stress state cannot actually arise since all materials possess a finite yield stress. Thus, any singularity implied in the elastic stress state will simply result in local plasticity, which will bring the stress state back within the material limits. But as long as the loading is mild enough for the size of the plastic zone to remain small in comparison to the region over which the singular solution is relevant, the singular solution can be accepted as an idealization. This is essentially a type of small-scale-yielding argument, and it is similar to that which is used to justify neglecting the influence of plasticity when using stress intensity factors to predict when brittle fracture will occur in a cracked component.

A more mathematical consideration that can be taken into account, which is described in Barber [2009] §11.2.1, is that the strain energy stored in any region of the body must vanish as the size of the region tends to zero. The strain energy, U, in a small region near the notch tip can be written as

$$U = \frac{1}{2}\int_0^{2\pi}\int_0^r \sigma\varepsilon r dr d\theta, \tag{2.17}$$

where ε denotes strain. The stresses and hence the strains are proportional to $r^{\lambda-1}$, so

$$U = C\int_0^r r^{2\lambda-1} dr = \frac{C}{2\lambda} r^{2\lambda}, \tag{2.18}$$

where C is a constant. This expression shows that U vanishes as $r \to 0$ only if $\lambda > 0$; otherwise U is unbounded. Hence, this criterion is only satisfied when λ is positive.

Therefore, the most important terms in the eigenfunction expansion are those with the smallest positive value of λ, and these are given by the lowest positive roots of (2.16a) and (2.16b).

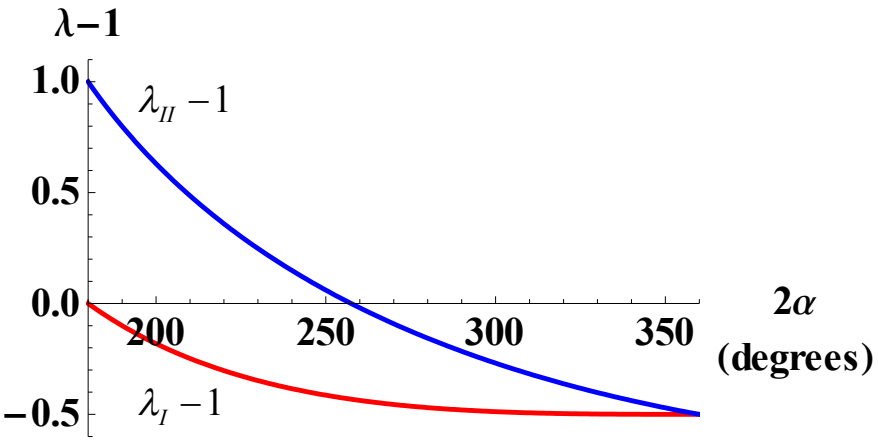


Figure 2.3: Elastic stress singularity associated with the mode I and mode II terms in Williams' solution plotted against notch angle, 2α.

2.2.4 Eigenvalues

The relevant values of λ_I, λ_{II} must of course be determined by solving (2.16a), (2.16b), respectively. However, these equations cannot be solved analytically, so a numerical (or graphical) approach must be used instead.[4] As mentioned above, it is the lowest positive eigenvalue from each of these equations that is of most interest. These were found numerically, and they are plotted against notch angle, 2α, in Figure 2.3. Note that $(\lambda - 1)$ is actually shown in this figure (as opposed to λ) to emphasize the order of the stress singularity corresponding to these eigenvalues. Recall that λ_I corresponds to the symmetric solution, while λ_{II} corresponds to the anti-symmetric solution.

This figure illustrates that the symmetric solution is singular for all notch angles within the range $180^\circ < 2\alpha \leqslant 360^\circ$. In contrast, the anti-symmetric solution is singular for $2\alpha > 257.4^\circ$ but is bounded for $2\alpha \leqslant 257.4^\circ$. Moreover, the next smallest eigenvalue for both eigensolutions gives $\lambda \geqslant 1$ for all α (although this cannot be seen in the figure). Therefore, there are a maximum of two singular terms that satisfy the strain energy criterion (2.18), and these will come to dominate the stress state as $r \to 0$. Hence, the stress state near the notch tip can be expressed as a truncated two-term series expansion comprising only the first term from each eigensolution.

[4] See [Barber, 2009] §11.2.3 for a graphical solution of these equations.

2.2.5 Eigenvectors

Once the eigenvalues λ_I, λ_{II} are known for a given notch angle, 2α, the corresponding eigenvectors can be determined by substituting $\alpha, \lambda_I, \lambda_{II}$ into (2.15a),(2.15b). The eigenvectors of (2.15a) and (2.15b) fix the ratio A_1/A_2 and A_3/A_4, respectively, as

$$\frac{A_1}{A_2} = -\frac{\cos[(\lambda_I - 1)\alpha]}{\cos[(\lambda_I + 1)\alpha]} = -\frac{(\lambda_I - 1)\sin[(\lambda_I - 1)\alpha]}{(\lambda_I + 1)\sin[(\lambda_I + 1)\alpha]} \tag{2.19a}$$

$$\frac{A_3}{A_4} = -\frac{\sin[(\lambda_{II} - 1)\alpha]}{\sin[(\lambda_{II} + 1)\alpha]} = -\frac{(\lambda_{II} - 1)\cos[(\lambda_{II} - 1)\alpha]}{(\lambda_{II} + 1)\cos[(\lambda_{II} + 1)\alpha]}. \tag{2.19b}$$

Therefore, the stress state for each mode of the solution (i.e. $\sigma_{ij}^I, \sigma_{ij}^{II}$ in (2.10)) is known to within a multiplicative constant. Hence, the stresses given in (2.9) may be written in the form

$$\sigma_{ij}(r, \theta) = K_I r^{\lambda_I - 1} f_{ij}^I(\theta) + K_{II} r^{\lambda_{II} - 1} f_{ij}^{II}(\theta), \tag{2.20}$$

where K_I, K_{II} are the mode I, II *generalized stress intensity factors*, which act as the multiplier on each mode of the solution.

It is convenient to normalize the angular functions, $f_{ij}(\theta)$, of the mode I solution by $f_{\theta\theta}^I(0)$ and those of the mode II solution by $f_{r\theta}^{II}(0)$ so that

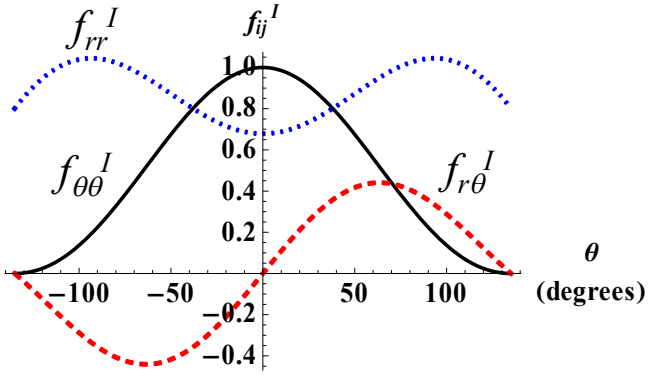
$$f_{\theta\theta}^I(0) = 1 \qquad\qquad f_{r\theta}^{II}(0) = 1. \tag{2.21}$$

As shown by Churchman [2006], these angular functions can then be written as

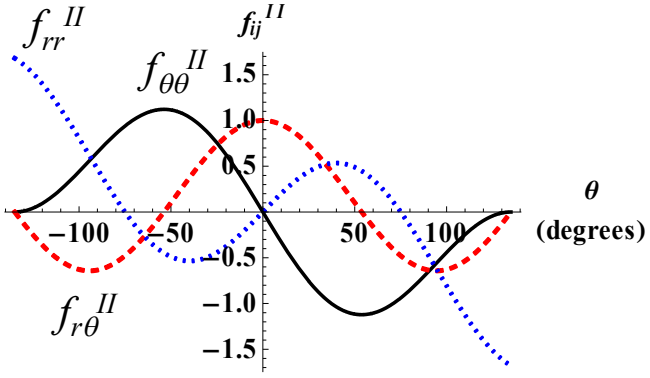
$$f_{rr}^I(\theta) = \frac{\cos[(\lambda_I - 1)\alpha]\cos[(\lambda_I + 1)\theta] - \frac{\lambda_I - 3}{\lambda_I + 1}\cos[(\lambda_I + 1)\alpha]\cos[(\lambda_I - 1)\theta]}{\cos[(\lambda_I + 1)\alpha] - \cos[(\lambda_I - 1)\alpha]} \tag{2.22a}$$

$$f_{\theta\theta}^I(\theta) = \frac{\cos[(\lambda_I - 1)\alpha]\cos[(\lambda_I + 1)\theta] - \cos[(\lambda_I + 1)\alpha]\cos[(\lambda_I - 1)\theta]}{\cos[(\lambda_I - 1)\alpha] - \cos[(\lambda_I + 1)\alpha]} \tag{2.22b}$$

$$f_{r\theta}^I(\theta) = \frac{\sin[(\lambda_I - 1)\alpha]\sin[(\lambda_I + 1)\theta] - \sin[(\lambda_I + 1)\alpha]\sin[(\lambda_I - 1)\theta]}{\sin[(\lambda_I - 1)\alpha] - \frac{\lambda_I + 1}{\lambda_I - 1}\sin[(\lambda_I + 1)\alpha]} \tag{2.22c}$$



(a) Mode I angular functions from (2.22).



(b) Mode II angular functions from (2.23).

Figure 2.4: Plots of the angular functions, f_{ij}^{I}, f_{ij}^{II}, for $2\alpha = 270°$, where $f_{rr}, f_{\theta\theta}, f_{r\theta}$ are shown as dotted, solid, dashed lines, respectively.

for the mode I solution and as

$$f_{rr}^{II}(\theta) = \frac{\sin[(\lambda_{II}-1)\alpha]\sin[(\lambda_{II}+1)\theta] - \frac{\lambda_{II}-3}{\lambda_{II}+1}\sin[(\lambda_{II}+1)\alpha]\sin[(\lambda_{II}-1)\theta]}{\sin[(\lambda_{II}-1)\alpha] - \frac{\lambda_{II}-1}{\lambda_{II}+1}\sin[(\lambda_{II}+1)\alpha]} \tag{2.23a}$$

$$f_{\theta\theta}^{II}(\theta) = \frac{\sin[(\lambda_{II}-1)\alpha]\sin[(\lambda_{II}+1)\theta] - \sin[(\lambda_{II}+1)\alpha]\sin[(\lambda_{II}-1)\theta]}{-\sin[(\lambda_{II}-1)\alpha] + \frac{\lambda_{II}-1}{\lambda_{II}+1}\sin[(\lambda_{II}+1)\alpha]} \tag{2.23b}$$

$$f_{r\theta}^{II}(\theta) = \frac{\cos[(\lambda_{II}-1)\alpha]\cos[(\lambda_{II}+1)\theta] - \cos[(\lambda_{II}+1)\alpha]\cos[(\lambda_{II}-1)\theta]}{\cos[(\lambda_{II}-1)\alpha] - \cos[(\lambda_{II}+1)\alpha]} \tag{2.23c}$$

for the mode II solution by using (2.9),(2.19),(2.21).

Figure 2.4 shows these functions for the example case of $2\alpha = 270°$. This figure

illustrates that the mode I, II solutions uncouple along the bisector in the sense that

$$f^{I}_{r\theta}(0) = 0 \qquad\qquad f^{II}_{\theta\theta}(0) = 0. \tag{2.24}$$

In other words, the mode I term does not generate any shear stress along the bisector, i.e. $\sigma^{I}_{r\theta}(r,0) = 0$; and the mode II term does not generate any normal stress along the bisector, i.e. $\sigma^{II}_{\theta\theta}(r,0) = 0$. Inspection of (2.22) and (2.23) shows that this is a general property of the solution that is not specific to this particular notch angle. Therefore, K_I, K_{II} can be defined along the notch bisector as

$$K_I = \lim_{r\to 0} \frac{\sigma_{\theta\theta}(r,0)}{r^{\lambda_I - 1} f^{I}_{\theta\theta}(0)} = \lim_{r\to 0} \sigma_{\theta\theta}(r,0) r^{1-\lambda_I} \tag{2.25a}$$

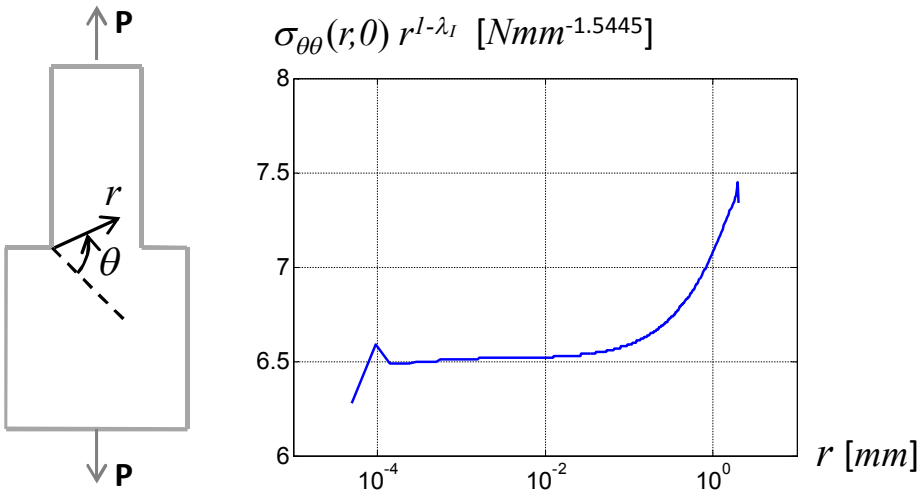
$$K_{II} = \lim_{r\to 0} \frac{\sigma_{r\theta}(r,0)}{r^{\lambda_{II} - 1} f^{II}_{r\theta}(0)} = \lim_{r\to 0} \sigma_{r\theta}(r,0) r^{1-\lambda_{II}} \tag{2.25b}$$

by using (2.20),(2.21), and (2.24).

2.2.6 Generalized stress intensity factors

Because of the semi-infinite nature of Williams' solution, the only relevant physical dimension that enters the problem is the notch angle, 2α. Once this parameter has been fixed, the stress state is determined to within two multiplicative constants, i.e. K_I, K_{II}. Hence, when Williams' solution is to be applied to a finite problem, such as that shown in Figure 2.2, it is only these multipliers, K_I, K_{II}, that remain to be determined, and these quantify the strength of the remote loads. However, K_I, K_{II} cannot be determined analytically and must instead be obtained numerically.

There are several ways to determine these multipliers, K_I, K_{II}, in practice. Perhaps the most obvious approach is simply to evaluate $\sigma_{\theta\theta}(r,0)$ and $\sigma_{r\theta}(r,0)$ near the corner region using a finely meshed finite element model. Once these stresses are obtained, the right hand side of (2.25a) or 2.25b) can be plotted against r on a semi-log plot, and the limit to which these quantities converge as $r \to 0$ gives K_I or K_{II}. An example



(a) Notched test geometry.

(b) Determination of K_I using $\sigma_{\theta\theta}$ along the notch corner bisector (dashed line in (a)).

Figure 2.5: Calculation of the mode I generalized stress intensity factor.

application of this approach is shown in Figure 2.5(b). In this figure, a calibration of K_I is calculated for the notched test geometry shown in Figure 2.5(a) when it is subjected to a load of $P = 1MPa$. In this case, the quantity $\sigma_{\theta\theta}(r,0)r^{1-\lambda_I}$ converges to the approximate value of $K_I = 6.51Nmm^{-1.5445}$.

Notice that numerical difficulties arise in the last few elements very near to the corner since the finite element model does not capture the singular stress state accurately. In addition, for large values of r, the stress state is strongly influenced by the finite geometry of the problem, so Williams' solution cannot be used to describe the stress state. Hence, care must be taken to ensure that K_I is determined at a sufficiently small value of r, while still avoiding the numerical instability in the few elements nearest to the corner.

Other numerical methods for determining the generalized stress intensity factors are discussed in Lazzarin et al. [2010]. Some of these approaches, e.g. those based on strain energy density, have the benefit of providing accurate results even when a relatively coarse mesh is used. However, all these methods converge on the same result once a sufficiently refined mesh is used.

2.3 Alternative set of parameters

In §2.2.4, the strength of the stress singularity associated with the mode I and mode II terms in Williams' solution is examined, and λ_I, λ_{II} are then plotted in Figure 2.3. This figure merits deeper examination because it illustrates the different forms that the two-term Williams solution may take depending on the notch angle. These are:

1. If $2\alpha = 360°$, the geometry becomes that of a crack, so both eigensolutions are square-root singular as in the familiar solution from fracture mechanics. Since both terms have the same order of singularity, the two-term solution is self-similar with r.

2. If $257.4° < 2\alpha < 360°$, both eigensolutions are singular, but the mode I term is more strongly singular than the mode II term. Thus, the stress state is *not* self-similar with r. Instead, the mode I term dominates the stress state as $r \to 0$ since it is more strongly singular, while the mode II term is only influential at some distance from the notch tip.

3. If $180° < 2\alpha \leqslant 257.4°$, the mode I term is singular, while the mode II term is bounded. In this case, the mode I term will dominate the stress state near the notch tip unless the loading is purely (or almost purely) mode II in character (i.e. if $K_{II} \gg K_I$).

4. If $2\alpha \leqslant 180°$, both eigensolutions are bounded. This illustrates that while re-entrant corners (i.e. $2\alpha \geqslant 180°$) sustain a singular state of stress, external corners (i.e. $2\alpha \leqslant 180°$) do not.

In the present work, the range of notch angles of interest corresponds to case 2, i.e. $257.4° < 2\alpha < 360°$, for which both mode I, II terms are singular, but where the mode I term is more strongly singular than the mode II term, i.e. $1/2 < \lambda_I < \lambda_{II} < 1$. Since the mode I, II terms are singular to different strengths, the relative influence of each solution mode, i.e. the *mode-mixity*, varies with r. Therefore, there is a length scale associated with Williams' two-term (semi-infinite) solution itself.

Hills and Dini [2011] have shown that this length scale can be brought out more clearly by abstracting an alternative set of parameters from the generalized stress intensity factors as

$$G_0 = |K_I|^{\frac{\lambda_{II}-1}{\lambda_{II}-\lambda_I}} |K_{II}|^{\frac{\lambda_I-1}{\lambda_I-\lambda_{II}}} \tag{2.26a}$$

$$d_0 = \left|\frac{K_{II}}{K_I}\right|^{\frac{1}{\lambda_I-\lambda_{II}}}. \tag{2.26b}$$

Williams' solution (2.20) can then be re-written as

$$\frac{\sigma_{ij}(r,\theta)}{G_0} = \left(\frac{r}{d_0}\right)^{\lambda_I-1} f_{ij}^{I}(\theta) + \left(\frac{r}{d_0}\right)^{\lambda_{II}-1} f_{ij}^{II}(\theta). \tag{2.27}$$

When written in this form, it is clear that d_0 has units of length and represents, in some sense, the boundary between mode I dominant behaviour (if $r \ll d_0$) and mode II dominant behaviour (if $r \gg d_0$). In contrast, G_0 has units of stress and provides a measure of the strength of the applied loads.

Note that if the remote loads result in pure mode I or pure mode II conditions, these alternative parameters cannot be used since d_0 and G_0 will tend either to zero or to infinity. However, so long as there is some contribution from both terms in the solution (even if one term dominates the stress field), these parameters remain valid. Also note that numerical difficulties may arise when $\lambda_I \sim \lambda_{II}$, i.e. for $2\alpha \gtrsim 330^\circ$ (see Figure 2.3), due to the definition of d_0, G_0.

Some further insight into the nature of these alternative parameters can be gained by defining them in terms of dimensionless generalized stress intensity factors. From dimensional considerations, it is clear that the generalized stress intensity factors can be written in the form

$$K_n = mka^{1-\lambda_n} S_n, \tag{2.28}$$

where $n \in \{I, II\}$, a is a characteristic length scale of the finite problem (e.g. the distance to the nearest geometric feature), k is the yield stress of the material in pure shear, m is

a dimensionless scalar that scales the applied load, and S_n is the dimensionless mode n generalized stress intensity factor. The parameters G_0 and d_0 may then be written in terms of dimensionless generalized stress intensity factors as

$$\frac{G_0}{k} = |mS_I|^{\frac{\lambda_{II}-1}{\lambda_{II}-\lambda_I}} \, |mS_{II}|^{\frac{\lambda_I-1}{\lambda_I-\lambda_{II}}} \tag{2.29a}$$

$$\frac{d_0}{a} = \left|\frac{S_{II}}{S_I}\right|^{\frac{1}{\lambda_I-\lambda_{II}}} . \tag{2.29b}$$

Once defined in this way, it is clear that d_0 is independent of the magnitude of the applied load, m, whereas G_0 is a function of m. Furthermore, as d_0 depends on the ratio of the generalized stress intensity factors, it provides a measure of the mode-mixity of the applied load, and this is examined in greater detail in subsequent chapters.

2.4 Summary

In this chapter, Williams' solution for a semi-infinite notch is derived, and several of its important features are discussed. A two-term expansion of Williams' solution is then examined with emphasis on the variation of the order of the singularity with notch angle. An alternative set of parameters introduced by Hills and Dini [2011] are also presented, which are most useful for notches within the range $257.4^\circ < 2\alpha < 360^\circ$, where both terms are singular but to different strengths. These parameters are particularly useful for analysing mixed-mode loading conditions and are employed extensively in §3-§5 for this purpose.

Chapter 3

Characterization of the process zone of sharp V-notches

In this chapter, Williams' solution is used to characterize the zone of local plasticity that develops at the root of sharp V-notches due to the implied singularity in the elastic state of stress at the notch tip. First, a method is presented for quantifying the mode-mixity of the plastic zone in terms of whether it is 'mode I like', 'mode II like', or mixed-mode in character. Second, an approach for calculating the small-scale-yielding limit is presented. Finally, these two techniques are applied to monotonic, mixed-mode experimental test data reported in the literature. The results indicate that: (i) most tests were carried out within 5% small-scale yielding, and (ii) the plastic zone of practical engineering components is likely to be either mainly mode I or mixed-mode in character.

Note that parts of this chapter have been published in [Flicek et al., 2014].

3.1 Introduction

When sharp features are present in engineering structures, e.g. at the edge of bonded or bolted interfaces, they are a common site of crack nucleation and initiation, especially when these components are subject to fatigue conditions. This is due to the implied singularity in the elastic stress state that results at the tip of sharp V-notches, which is examined in §2. Of course, a singular stress state cannot occur in practice, and these high stresses are instead truncated by local plasticity. This plastic zone near the notch

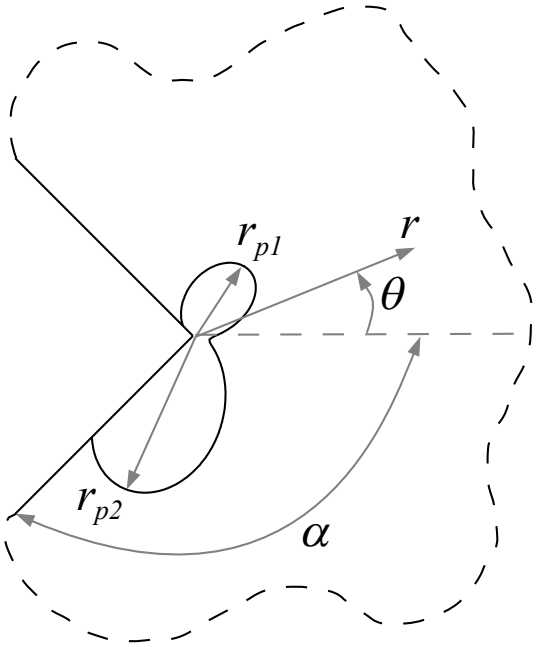


Figure 3.1: Idealized diagram of a semi-infinite notch of interior angle 2α showing the polar coordinate set (r, θ) and both lobes of the plastic zone, $r_{p1}(\theta), r_{p2}(\theta)$.

tip, $r_p(\theta)$, which is shown in Figure 3.1, can be thought of as a *process zone*, where the damage processes that lead to crack nucleation and initiation occur. The aim of this chapter is to provide an accurate description of this critical region using techniques that can easily be applied to any notch angle, 2α.

This analysis builds on that of Hills and Dini [2011] and makes use of the alternative parameters, d_0, G_0, introduced in §2.3, which allow the stress state near the notch tip to be written as in (2.27). Whereas the generalized stress intensity factors, K_I, K_{II}, emphasize the individual strength of each mode of loading, d_0 and G_0 place emphasis on the mode-mixity and overall magnitude of the applied loads, respectively. In addition, d_0 has units of length and provides a length scale within the semi-infinite solution.

As is discussed in §2.3, for notch angles within the range $180^\circ < 2\alpha \leqslant 257.4^\circ$, the exponents on the mode I, II terms are not only unequal but also have the property that $1/2 < \lambda_I < \lambda_{II} < 1$ (see Figure 2.3). An important implication of this property is that the stress state is *not* self similar with r; hence, the more strongly singular mode I term dominates the stress state close to the notch tip (i.e. when $r \ll d_0$), whereas the

less strongly singular mode II term is influential further from the notch tip (i.e. when $r \gg d_0$). This length scale effect is an inherent property of a two-term expansion of Williams' solution, and it can be used to characterize the mode-mixity of the process zone.

3.2 Mode-mixity at the process zone

Because of the implied singularity in the elastic stress state, there is always some small amount of plasticity at the root of a sharp V-notch that is subjected to any remote loading. The size of this plastic zone is controlled by the ratio of the magnitude of the applied load, G_0, to the strength of the material, k. Thus, when the loading is light (i.e. $G_0 \ll k$), the process zone, r_p, is small in comparison to d_0, i.e. $r_p \ll d_0$, and is mainly mode I in character because it lies within a region where the stress state is mode I dominated. As G_0 is increased, the process zone increases in size, but it also rotates as it moves through the changing mode-mixity of the surrounding stress field. Thus, as G_0 is increased, the process zone initially becomes mixed-mode in character when $r_p \sim d_0$ and finally becomes mode II in character when $r_p \gg d_0$. Therefore, this ratio, G_0/k, determines the mode-mixity *at the scale of the process zone*.

Hills and Dini [2011] used this property of the solution to characterize the process zone of a 270° notch as being mainly mode I or mainly mode II in character based on the angle corresponding to the maximum plastic radius, $\theta_{p1}^{\max}, \theta_{p2}^{\max}$, for each lobe of the plastic zone, r_{p1}, r_{p2}. The aim of this section is to extend this work by using a different approach: namely, by performing a full field collocation of the mixed-mode process zone with reference pure mode I and pure mode II process zones, $r_p^I(\theta), r_p^{II}(\theta)$.

To do this, an estimate of the process zone is obtained based on violations of von Mises yield criterion. This approach is similar to the way that the size of the plastic zone is estimated when calculating the small-scale-yielding limit in fracture mechanics. First,

it is noted that the second invariant of deviatoric stress can be written as

$$J_2 = \sigma_{rr}^2 + \sigma_{\theta\theta}^2 + \sigma_{zz}^2 - (\sigma_{rr}\sigma_{\theta\theta} + \sigma_{\theta\theta}\sigma_{zz} + \sigma_{zz}\sigma_{rr}) + 3\sigma_{r\theta}^2, \tag{3.1}$$

and, with this scaling, von Mises yield criterion is

$$J_2 = 3k^2. \tag{3.2}$$

Plane strain conditions are assumed here, so

$$\sigma_{zz} = \nu(\sigma_{rr} + \sigma_{\theta\theta}), \tag{3.3}$$

where ν is Poisson's ratio. All other stress components in (3.1) can be obtained from (2.20) or (2.27).

For the case of mixed-mode loading (i.e. when $K_I, K_{II} \neq 0$), both the *size and shape* of the mixed-mode process zone, $r_p(\theta)$, can be estimated *within the asymptote* by equating (3.1) and (3.2) and substituting in stresses from (2.27) and (3.3). Both the size and shape of the process zone can be estimated because the two-term Williams expansion contains an intrinsic length scale, d_0. In contrast, the stress state under pure mode n loading, where $n \in \{I, II\}$, can be expressed with a one-term expansion as

$$\sigma_{ij}^n(r,\theta) = K_n r^{\lambda_n - 1} f_{ij}^n(\theta) = (mka^{1-\lambda_n} S_n) r^{\lambda_n - 1} f_{ij}^n(\theta), \tag{3.4}$$

where $i, j \in \{r, \theta\}$. In this case, the stress state is self similar with r, so there is no length dimension that the size of the process zone can be compared against. Consequently, only the *shape* of a pure mode n process zone, $r_p^n(\theta)$, can be determined within the asymptotic solution.[1] Nevertheless, the *shape* of r_p^n can be estimated as before by equating (3.1) and (3.2) but this time using stresses from (3.4) and (3.3).

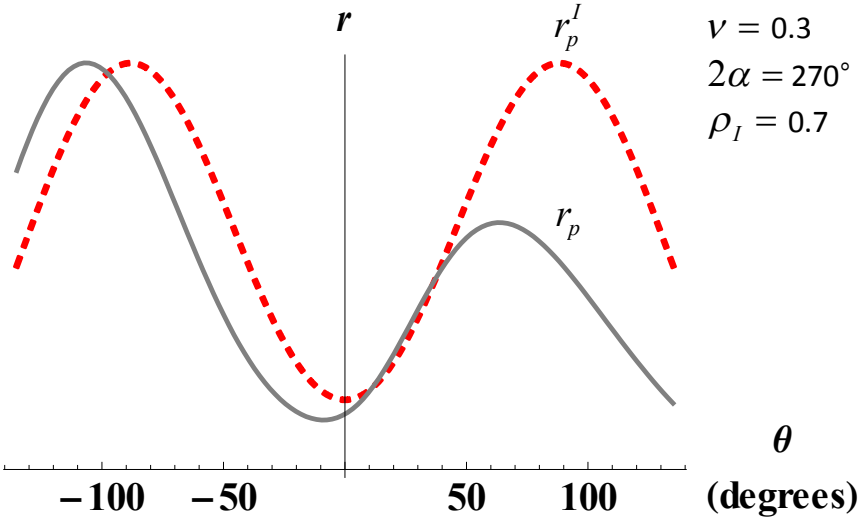
[1] However, the size of the pure mode I, II process zones can be estimated once this solution is applied to a finite problem as a corner asymptote because the solution can then be normalized with respect to a characteristic length from the finite problem, a.

3.2.1 Comparison of pure and mixed-mode process zones

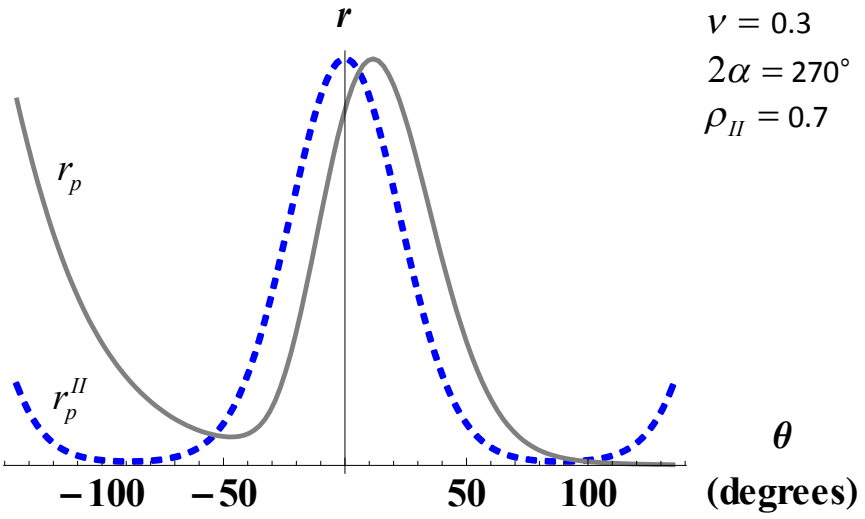
The aim is now to compare the mixed-mode process zone to the pure mode I, II process zones in order to assess its similarity to the reference states as the strength of the applied load, G_0/k, is varied. However, since the size of a pure mode n process zone cannot be determined within the semi-infinite solution, only the *shape* of these process zones can be compared. Thus, one way to perform this comparison is to use the *correlation coefficient*, ρ, (also referred to as *Pearson's product-moment correlation coefficient*) because it is invariant with both scale and origin, and it quantifies the degree of linear dependence between two curves. The correlation coefficient of the mixed-mode process zone, r_p, with respect to a pure mode n process zone, r_p^n, is denoted ρ_n. Note that ρ is calculated by dividing the covariance of two curves by the product of their standard deviations. In addition, ρ is always within the range $-1 \leq \rho \leq 1$, where $\rho = 1$ implies a perfect positive linear relationship, $\rho = -1$ implies a perfect negative linear relationship, and $\rho = 0$ implies no linear relationship.

It is probably easiest to visualize the calculation of the correlation coefficients, ρ_I, ρ_{II}, on a Cartesian plot of $r_p(\theta)$ vs. θ. An example of this is shown in Figure 3.2 for a 270° notch with $\nu = 0.3$, which shows the *shape* of: (a) a pure mode I process zone and a mixed-mode process zone of $G_0/k = 0.2495$, and (b) a pure mode II process zone and a mixed-mode process zone of $G_0/k = 1.292$. Each process zone in this figure is normalized by its maximum value because the relative sizes of these process zones has no influence on the correlation coefficient. Also note that the correlation between the process zones in Figure 3.2(a),(b) are $\rho_I = 0.7, \rho_{II} = 0.7$, respectively. See §A.1.1 for some more details on how this calculation is performed numerically.

In order to illustrate more clearly the relationship between the correlation coefficients and the strength of the applied load, ρ_I and ρ_{II} are plotted (when positive) over a wide range of G_0/k in Figure 3.3 for the example case of a 270° notch with $\nu = 0.3$. This figure shows that for very small G_0/k (when $r_p \ll d_0$) the correlation with a pure mode I



(a) Pure mode I process zone, r_p^I, and a mixed-mode process zone, r_p, of $G_0/k = 0.2495$.



(b) Pure mode II process zone, r_p^{II}, and a mixed-mode process zone, r_p, of $G_0/k = 1.292$.

Figure 3.2: Cartesian plot of r vs. θ displaying the relative *shape* of pure and mixed-mode process zones for a $270°$ notch with $\nu = 0.3$ resulting in $\rho = 0.7$.

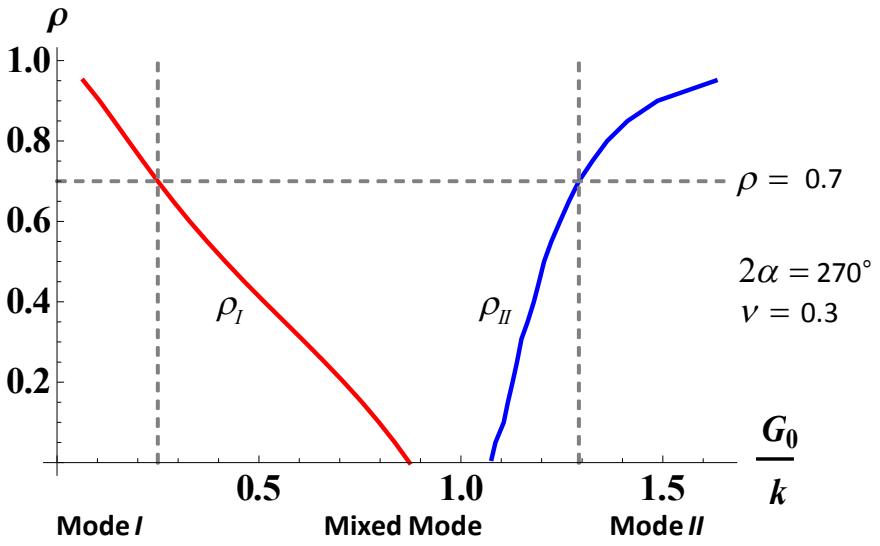


Figure 3.3: Plot of the correlation coefficients, ρ_I, ρ_{II}, (when positive), vs. G_0/k for a notch of $2\alpha = 270°$ and $\nu = 0.3$, which illustrates the variation of the mode-mixity of the process zone with the applied load. The range of applied loads for which the process zone is mode I dominated, mode II dominated, or mixed-mode are shown (assuming $\rho_I, \rho_{II} = 0.7$ mark these boundaries).

process zone is high (and the correlation with a pure mode II process zone is negative and is not shown). However, as G_0/k is increased, ρ_I decreases until it reaches zero (when $r_p \sim d_0$) and the correlation with the mode II process zone becomes positive. As G_0/k is increased further, ρ_{II} increases and eventually approaches unity (when $r_p \gg d_0$).

3.2.2 Mode-mixity results

This approach for characterizing the mode-mixity of the process zone can now be used to obtain general results, which can be used to characterize experimental data. To do this, values of ρ_I, ρ_{II} must be chosen that are assumed to mark the boundary between a mixed-mode process zone and a process zone that is 'mainly mode I' or 'mainly mode II' in character. An example of these boundaries is shown in Figure 3.3 assuming that $\rho_I, \rho_{II} > 0.7$ corresponds to a mode I, II dominated process zone, respectively.

Notice from Figure 3.3 that any value of G_0/k corresponds to particular values of ρ_I, ρ_{II} and thus to a particular level of mode-mixity at the process zone. This is exploited in Figure 3.4, which shows the values of G_0/k for which the process zone is mode I

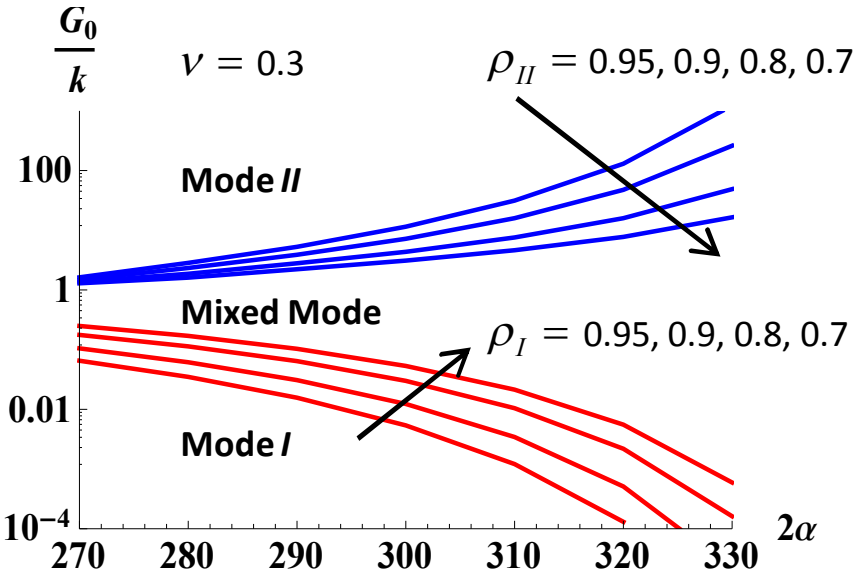
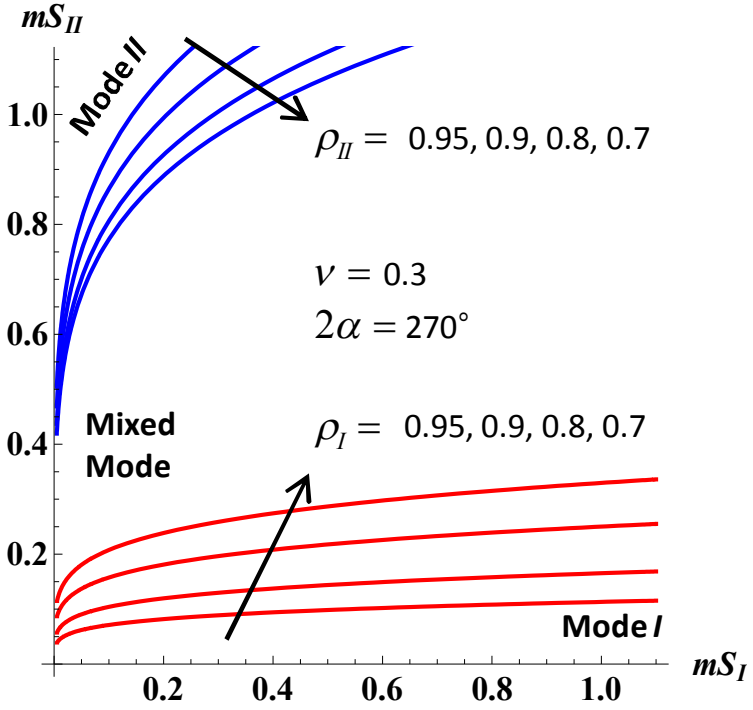


Figure 3.4: Plot of G_0/k vs. notch angle, 2α, showing the mode-mixity of the process zone for four choices of ρ_I, ρ_{II}, viz. $0.95, 0.9, 0.8, 0.7$, when $\nu = 0.3$.

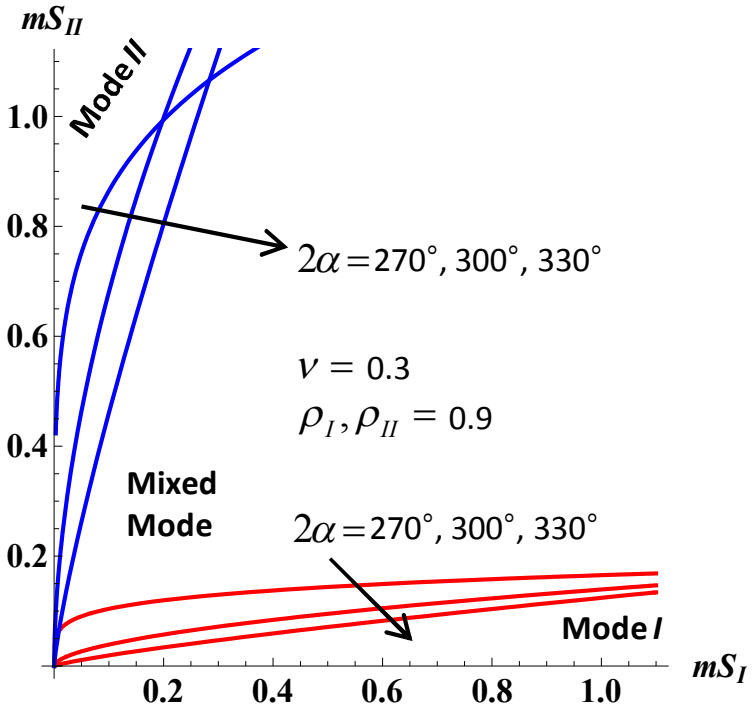
dominated, mode II dominated, or mixed-mode in character for notch angles within the range $270^\circ < 2\alpha < 330^\circ$ for $\nu = 0.3$. Because the value of ρ_I, ρ_{II} that is assumed to constitute a process zone that is mode I or mode II dominated is arbitrary, results are shown for four choices of ρ_I, ρ_{II}, viz. $0.95, 0.9, 0.8, 0.7$.

Although Figure 3.4 is useful and comprehensive, it employs an unfamiliar characterization of the magnitude of the load, G_0/k, which must be calculated from (2.26a) or (2.29a). Alternatively, the regions of mode I, II domination of the process zone can be plotted in generalized-stress-intensity-factor space, i.e. on a plot of mS_{II} vs. mS_I (or equivalently $K_{II}a^{\lambda_{II}-1}/k$ vs. $K_I a^{\lambda_I - 1}/k$), and this is done in Figure 3.5. In Figure 3.5(a), results are shown for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ but only for a particular notch angle, $2\alpha = 270^\circ$, with $\nu = 0.3$. Conversely, in Figure 3.5(b), results are only shown for $\rho_I, \rho_{II} = 0.9$ but for three notch angles, viz. $2\alpha = 270^\circ, 300^\circ, 330^\circ$, again with $\nu = 0.3$. Note that to generate Figure 3.5, the relevant value of G_0/k corresponding to ρ_I, ρ_{II} must simply be taken from Figure 3.4 and substituted into (2.29a).

From Figures 3.4 and 3.5(a), it is clear that the selection of ρ_I, ρ_{II} has a very significant effect on the size of the regions in which the process zone may be said to be mode I



(a) Mode-mixity boundaries corresponding to $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ for a $2\alpha = 270^\circ$ with $\nu = 0.3$



(b) Mode-mixity boundaries corresponding to $\rho_I, \rho_{II} = 0.9$ for $2\alpha = 270^\circ, 300^\circ, 330^\circ$ with $\nu = 0.3$.

Figure 3.5: Plots in dimensionless-generalized-stress-intensity-factor space, i.e. mS_{II} vs. mS_I, showing the mode-mixity of the process zone.

or mode II dominated. This is, of course, because ρ_I, ρ_{II} control the definition of what constitutes a process zone that is mode I or mode II dominated. Figure 3.5(a) also shows that the effect of varying ρ_I, ρ_{II} is to apply a simple dilation to these boundaries.

In contrast, Figure 3.5(b) shows that varying the notch angle, 2α, has marked effects not only on the size but also on the shape of the mode I and mode II dominated regions. In fact, it is clear that varying the notch angle significantly alters the relative sizes of the mode I and mode II dominated regions. This is not surprising and is explained by the variation of the strength of the mode I and mode II singularities with notch angle, which is shown in Figure 2.3. Thus, for 'small' notch angles, the mode I dominated region is much larger than the mode II dominated region because the mode I singularity is much stronger than the mode II singularity, e.g. for a 270° notch $(\lambda_I - 1) = -0.4555$ and $(\lambda_{II} - 1) = -0.0915$. Conversely, for larger notch angles, the mode I and mode II dominated regions are of a similar size because the singularities are of similar strengths, e.g. for a 330° notch $(\lambda_I - 1) = -0.4985$ and $(\lambda_{II} - 1) = -0.4018$.

3.3 Small-scale yielding

In this section, the limit on the applied load required to satisfy the small-scale-yielding assumption is calculated and displayed on plots of mS_{II} vs. mS_I. This calculation is very similar to the calculation of the small-scale-yielding limit for a crack tip in fracture mechanics. The motivation for this calculation is that Williams' solution can only be applied to finite problems in an asymptotic sense: that is, it is only valid very near to the sharp corner where the singular solution dominates the state of stress, i.e. for $r \ll a$, where a is the distance to the nearest geometric feature in the finite problem (see Figure 3.8 for examples). Thus, the analysis described in §3.2 can only be used to describe the process zone in finite problems when $r_p^{\max}/a \ll 1$, where $r_p^{\max}$ is the maximum extent of the plastic zone.

To perform this calculation, the size of the process zone is estimated in a very similar way to that described in §3.2 except that information about the distance to the nearest

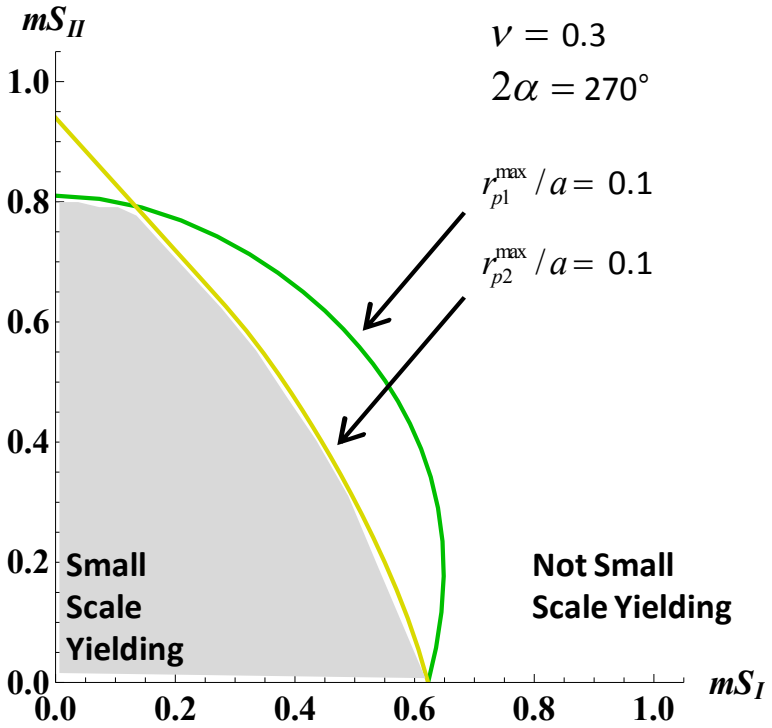


Figure 3.6: Plot of mS_{II} vs. mS_I showing the range of loads that result in $r_{p1}^{\max}, r_{p2}^{\max} = 0.1a$ for $2\alpha = 270°$ with $\nu = 0.3$. The shaded area shows the region in which both plastic lobes, $r_{p1}^{\max}, r_{p2}^{\max}$, are within small-scale yielding.

geometric feature, a, is included by normalizing the stresses in (2.20) by this parameter. For this calculation, it is convenient to use stresses with the standard parameters, K_I, K_{II}, from (2.20) instead of those with d_0, G_0 from (2.27). In this case, the solution can be normalized by a simply by substituting the right hand side of (2.28) for K_I, K_{II} and working in terms of dimensionless generalized stress intensity factors, S_I, S_{II}. An estimate of the size of the process zone can be obtained in this way by substituting stresses from (2.20),(3.3) into (3.1) and using (2.28),(3.2). The combinations of mS_I, mS_{II} that result in $r_p^{\max}/a$ being equal to the desired small-scale-yielding limit can then be calculated numerically using the resulting expression. See §A.1.2 for some more details on how this calculation is performed.

One detail of this calculation that is worth mentioning is that because the process zone has two lobes, $r_{p1}(\theta), r_{p2}(\theta)$, its maximum extent, $r_p^{\max}$, will either correspond to the maximum extent of r_{p1} or r_{p2}, i.e. to $r_{p1}^{\max}$ or $r_{p2}^{\max}$. For example, Figure 3.6 displays

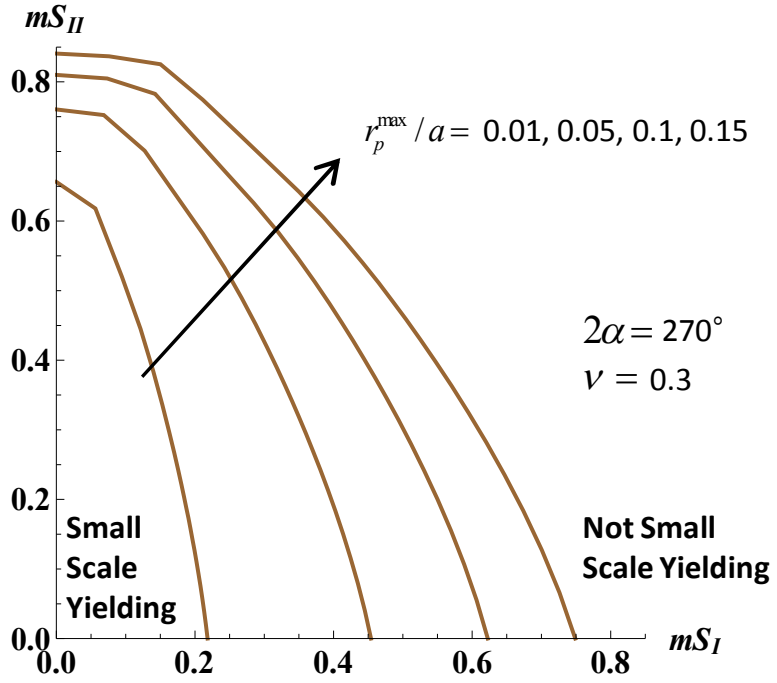
the values of mS_I, mS_{II} corresponding to $r_{p1}^{\max}/a, r_{p2}^{\max}/a = 0.1$ when $2\alpha = 270°$ and $\nu = 0.3$. Thus, the shaded area in this figure corresponds to the range of mS_I, mS_{II} for which both plastic lobes, $r_{p1}^{\max}, r_{p2}^{\max}$, are within small-scale yielding, so it is this shaded region that gives the range of loads that are within small-scale yielding. Hence, only the boundary corresponding to the edge of this shaded region is shown in all subsequent plots.

3.3.1 Small-scale-yielding results

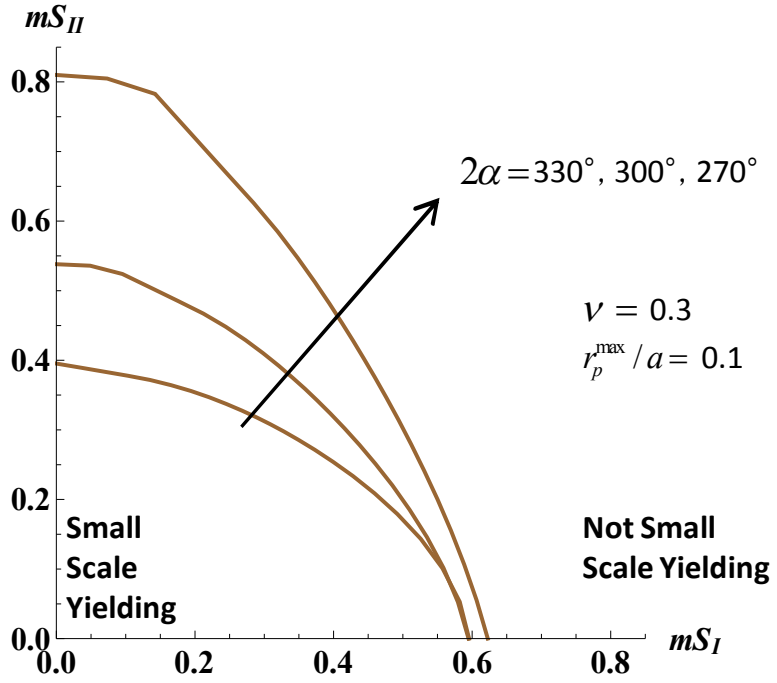
This small-scale-yielding calculation can be applied to a wide range of notch angles and used to create general plots of mS_{II} vs. mS_I, which can be superposed with the mode-mixity results shown in Figure 3.5(a) and used to classify experimental data. Of course, the selection of the small-scale-yielding limit, $r_p^{\max}/a$, is arbitrary and depends on how much influence of the neglected higher-order (bounded) terms in the series expansion may be accepted. Thus, results are shown for several small-scale-yielding limits, viz. $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$, in Figure 3.7(a) for the example case of $2\alpha = 270°$ with $\nu = 0.3$. Conversely, only one small-scale-yielding limit is shown in Figure 3.7(b), $r_p^{\max}/a = 0.1$, but for three notch angles, viz. $2\alpha = 270°, 300°, 330°$, again with $\nu = 0.3$.

As is no surprise, Figure 3.7(a) illustrates that the selection of the small-scale-yielding limit, $r_p^{\max}/a$, has a significant effect on the range of loads, i.e. mS_I, mS_{II}, that fall within small-scale yielding. It is also apparent that the effect of increasing $r_p^{\max}/a$ is to apply a fairly uniform dilation to the small-scale-yielding curve.

In contrast, and notably, Figure 3.7(b) shows that varying the notch angle has only the most minuscule effect on the area within small-scale yielding for loading conditions that are mainly mode I in character (i.e. that are very near to the mS_I axis) but a marked effect for loading conditions that are mainly mode II in character (i.e. very near to the mS_{II} axis). This can be attributed to the fact that the order of the mode II singularity changes much more rapidly with notch angle than does the mode I singularity (see Figure 2.3).



(a) Small-scale-yielding limits of $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ for $2\alpha = 270°$ with $\nu = 0.3$.



(b) Small-scale-yielding limits of $r_p^{\max}/a = 0.1$ for $2\alpha = 270°, 300°, 330°$ with $\nu = 0.3$.

Figure 3.7: Plots of mS_{II} vs. mS_I showing the small-scale-yielding limits.

3.4 Application to sharp V-notches

These techniques for characterizing the mode-mixity and small-scale yielding of the process zone can now be used to analyse experimental data reported in the literature. A review of the literature reveals that a great deal of experiments have been carried out to validate the various brittle fracture criteria that have been proposed for sharp and blunt notches, e.g. [Ayatollahi et al., 2011, Dunn et al., 1997a, Elices et al., 2002, Lazzarin and Zambardi, 2001, Lazzarin et al., 2009, Leguillon, 2002, Priel et al., 2007, Seweryn, 1994, Seweryn and Łukaszewicz, 2002]. Of the various experiments reported, five groups of tests have been selected, which were performed by [Ayatollahi et al., 2011, Dunn et al., 1997b, Kim and Cho, 2008, Priel et al., 2007, Seweryn et al., 1997] on polymethyl metacrylate (PMMA) notched specimens under mixed-mode, monotonic loading conditions.

Idealized diagrams (not to scale) of the selected test geometries are shown in Figure 3.8. In this figure, the applied loads denoted P, T, M correspond to an applied normal force, shear force, and moment, respectively, and $R1, R2$ are reaction forces that arise at supports in the test rig. The dimension that has been selected as the characteristic length, a, which controls the definition of the small-scale-yielding limit and also affects the values computed for the dimensionless generalized stress intensity factors, S_n, is also shown in this figure. Note that a was selected as the smaller of: (i) half the distance to the nearest sharp re-entrant corner ($2\alpha > 180°$), or (ii) the full distance to any other geometric feature. The characteristic length has been selected in this way because sharp re-entrant corners sustain a very large stress intensification and thus have an early influence on the stress state, whereas external corners ($2\alpha < 180°$) and smooth surfaces do not.

To analyse these experiments, the generalized stress intensity factors at failure must be known, and these have been obtained from [Ayatollahi et al., 2011, Dunn et al., 1997b, Kim and Cho, 2008, Priel et al., 2007, Seweryn and Łukaszewicz, 2002, Seweryn et al., 1997]. More details on how these data have been processed are given in §A.1.3, and tabulated values of mS_I, mS_{II} at failure are reported in §A.1.4. In addition to these data,

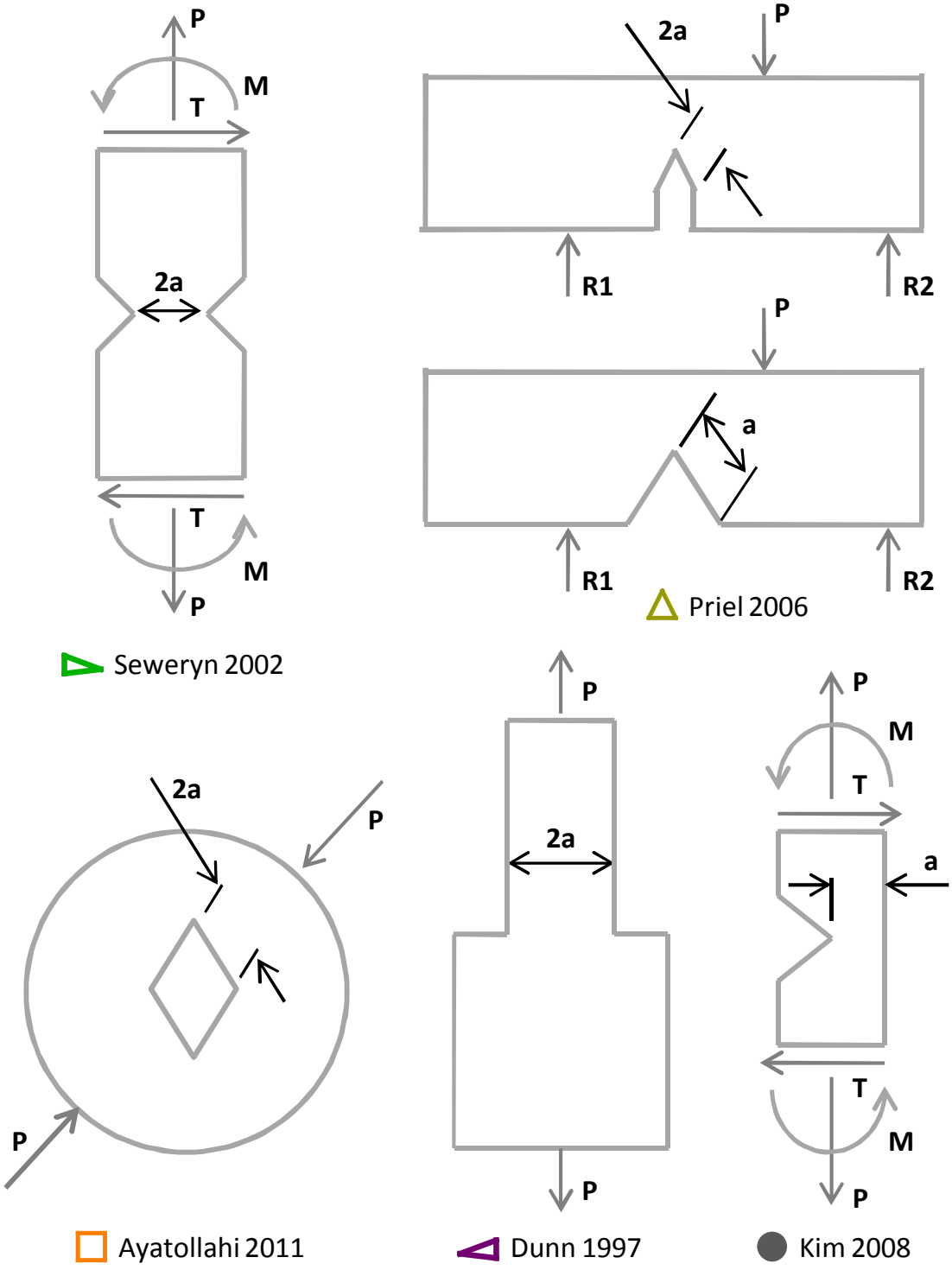
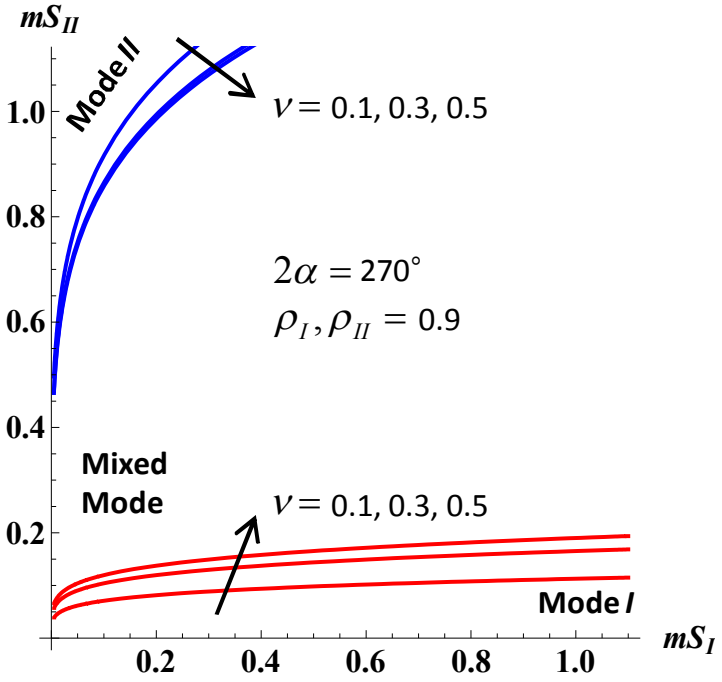
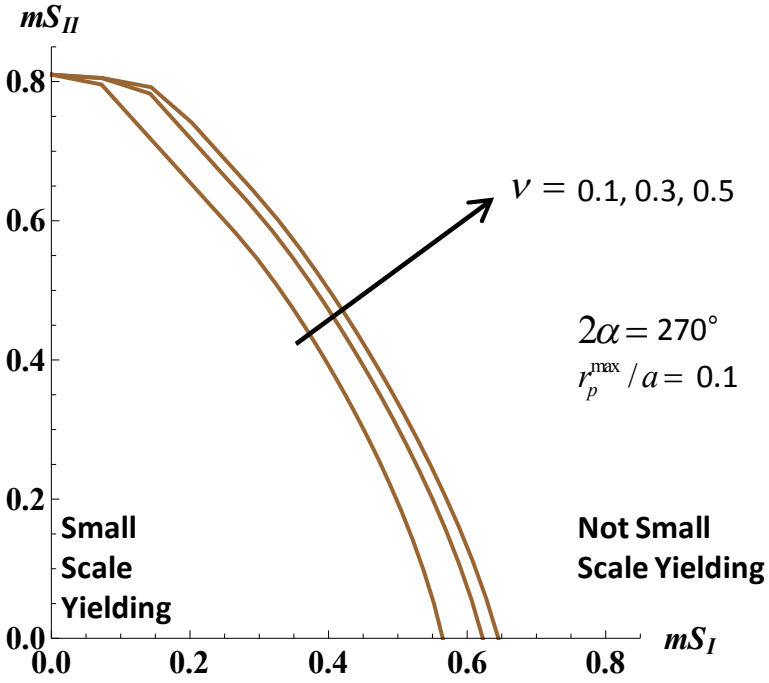


Figure 3.8: Idealized diagrams (not to scale) of the selected test geometries. The applied loads and the dimension selected to be the characteristic length, a, are also shown.



(a) Mode-mixity results for $\rho_I, \rho_{II} = 0.9$.



(b) Small-scale-yielding results for $r_p^{\max}/a = 0.1$

Figure 3.9: Plots showing the influence of Poisson's ratio for $2\alpha = 270^\circ$.

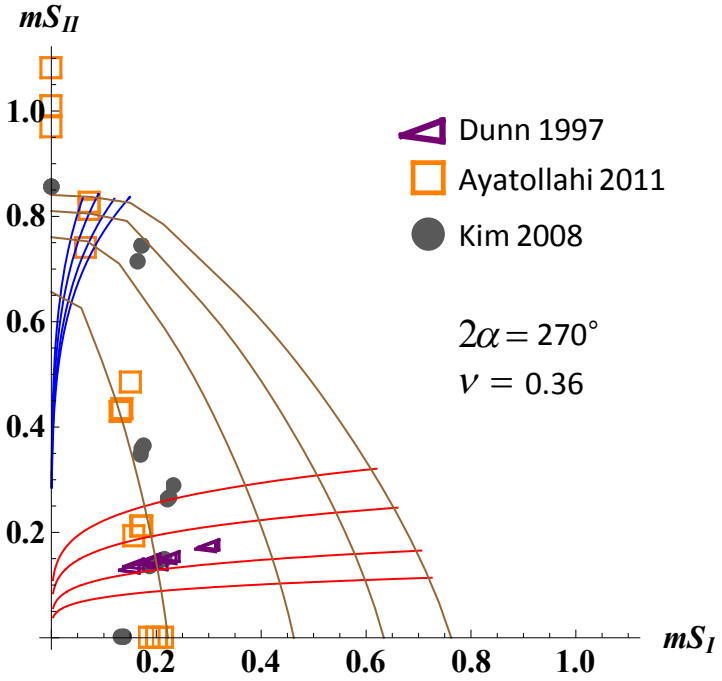
the value of Poisson's ratio, ν, must also be known, and this is quoted as 0.35 in [Seweryn and Łukaszewicz, 2002], 0.36 in [Dunn et al., 1997b, Kim and Cho, 2008, Priel et al., 2007], and 0.38 in [Ayatollahi et al., 2011]. To examine the influence of this level of variation in ν, Figure 3.9 shows: (a) the mode-mixity results for $\rho_I, \rho_{II} = 0.9$ and (b) the small-scale-yielding results for $r_p^{\max}/a = 0.1$ over a wide range of values of Poisson's ratio, viz. $\nu = 0.1, 0.3, 0.5$, for the example case of a 270° notch. This figure makes clear that the effect of varying ν from 0.35 to 0.38 is negligible, so all subsequent results are computed for $\nu = 0.36$.

Figures 3.10-3.13 show the dimensionless generalized stress intensity factors at failure on plots of mS_{II} vs. mS_I. These plots show the regions in which the process zone is mode I dominated, mode II dominated, or is mixed-mode in character according to four definitions, viz. $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$.[2] Small-scale-yielding limits of $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown simultaneously. Alternatively, all of the mixed-mode data shown in Figures 3.10-3.13 can be viewed on a single plot of G_0/k vs. 2α, and this is done in Figure 3.14, which includes boundaries corresponding to $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$. However, this plot excludes data obtained in pure I, II loading conditions because d_0, G_0 are not defined under these conditions. The benefit of this method of display is that data from many different notch angles can be plotted simultaneously, but this comes at the cost of being unable to view small-scale-yielding information.

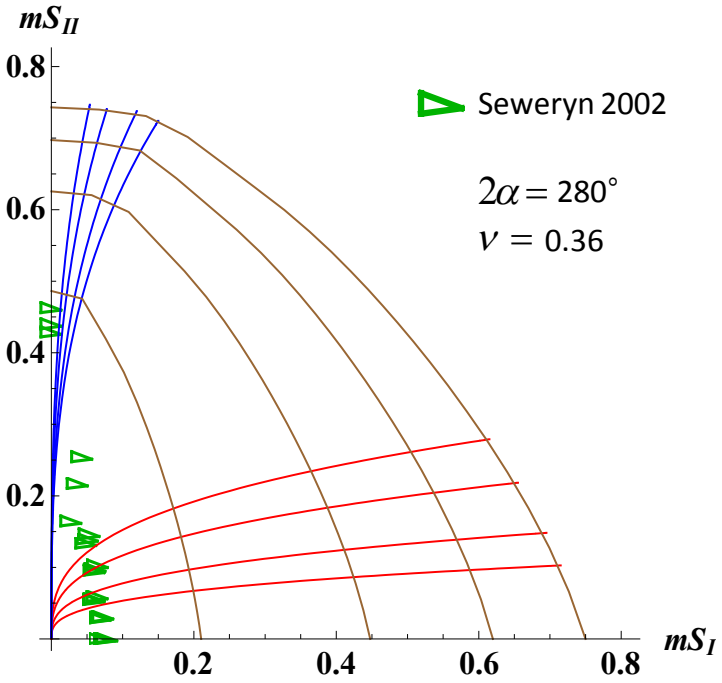
3.4.1 Small-scale yielding of experimental data

Figures 3.10-3.13 show that much of the data are within 1% small-scale yielding ($r_p^{\max}/a = 0.01$), including all of the tests by Seweryn et al. [1997] and Priel et al. [2007] and also most of those by Dunn et al. [1997b], Kim and Cho [2008], and Ayatollahi et al. [2011]. The remainder of the tests by Dunn et al. [1997b] are within 5% small-scale yielding. This is also true of most of the tests by Kim and Cho [2008] and Ayatollahi

[2]The only exception to this is in Figure 3.13 for $2\alpha = 330^\circ$ for which only the results for $\rho_{II} = 0.95, 0.9, 0.8, 0.7$ and $\rho_I = 0.9, 0.8, 0.7$ are shown because the $\rho_I = 0.95$ case failed to converge.

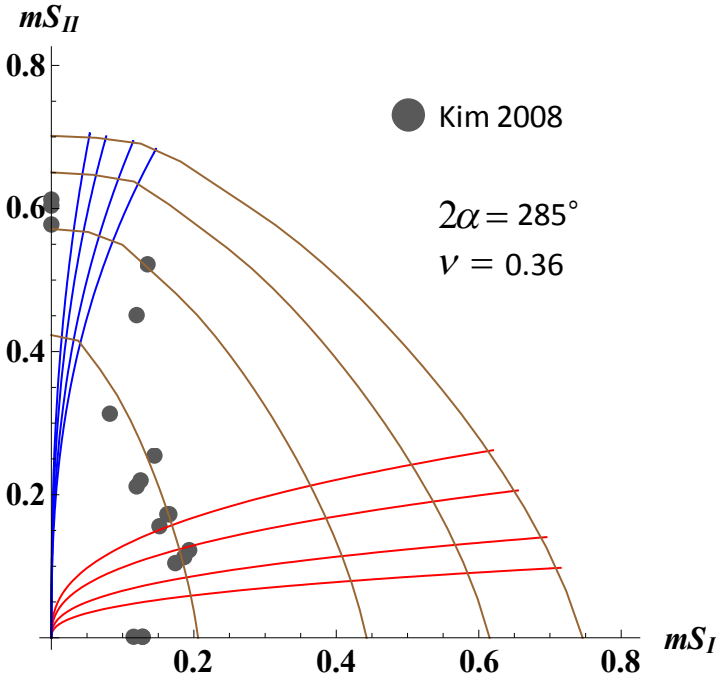


(a) $2\alpha = 270°$

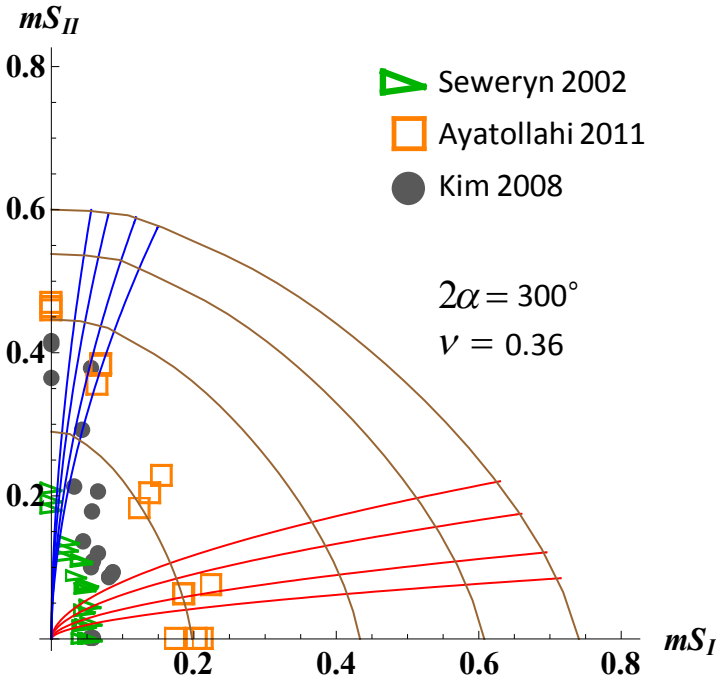


(b) $2\alpha = 280°$

Figure 3.10: Plots showing mS_I, mS_{II} at failure for the fracture test data for $2\alpha = 270°, 280°$. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown (all calculated assuming $\nu = 0.36$).

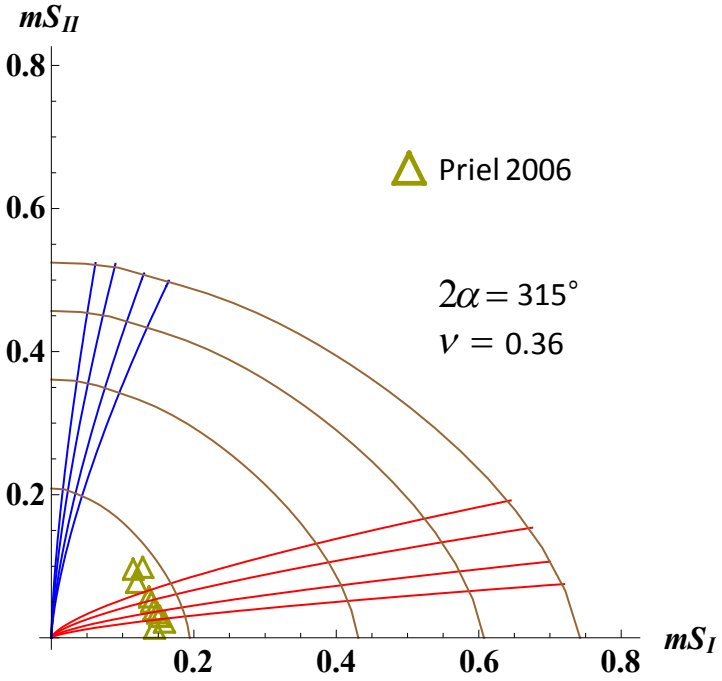


(a) $2\alpha = 285°$

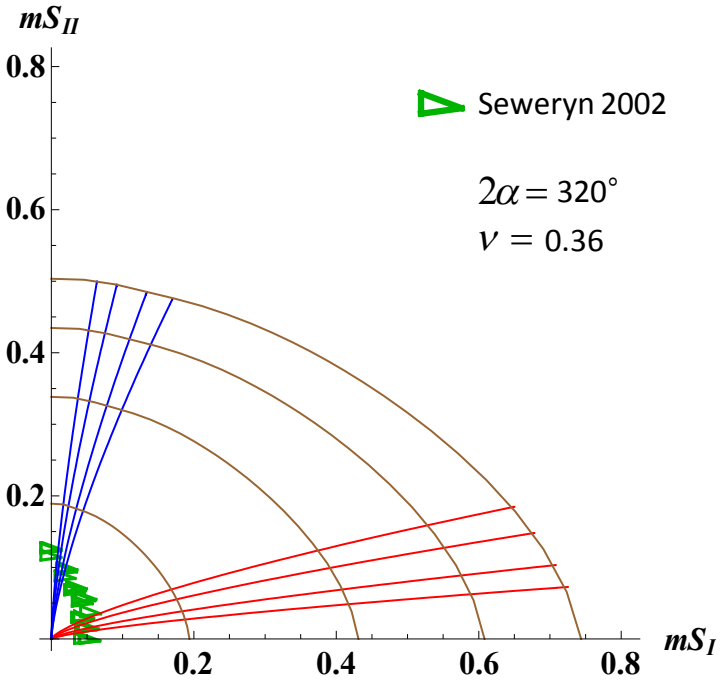


(b) $2\alpha = 300°$

Figure 3.11: Plots showing mS_I, mS_{II} at failure for the fracture test data for $2\alpha = 285°, 300°$. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown (all calculated assuming $\nu = 0.36$).



(a) $2\alpha = 315°$



(b) $2\alpha = 320°$

Figure 3.12: Plots showing mS_I, mS_{II} at failure for the fracture test data for $2\alpha = 315°, 320°$. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown (all calculated assuming $\nu = 0.36$).

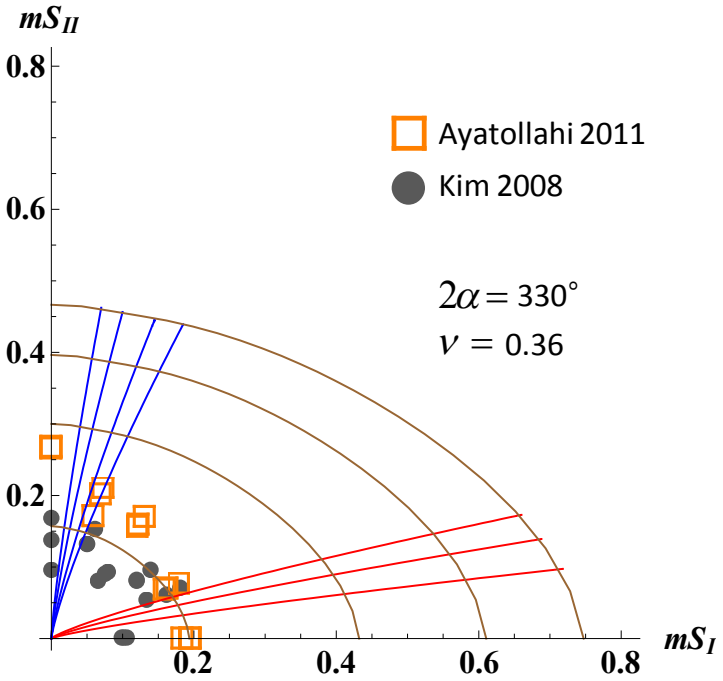


Figure 3.13: Plot showing mS_I, mS_{II} at failure for the fracture test data for $2\alpha = 330°$. Mode-mixity results for $\rho_I = 0.9, 0.8, 0.7$ and $\rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown (all calculated assuming $\nu = 0.36$).

et al. [2011] aside from those conducted under mode II dominated loading on notches of $2\alpha \lesssim 300°$ (see Figures 3.10(a) and 3.11). Figure 3.10(a) also shows that for mode II dominated loading, the tests by Kim and Cho [2008] are between 5% and 15% small-scale yielding, whereas those by Ayatollahi et al. [2011] are outside 15% small-scale yielding. Here, it is clear that the small-scale-yielding assumption is being stretched, and this is probably because the stress field is only very weakly singular, i.e. $(\lambda_{II} - 1) = -0.0915$.

That some of the tests stretch the small-scale-yielding assumption may be evidenced in [Kim and Cho, 2008] by the observed critical loads exceeding the predictions of the brittle fracture criterion used by the authors, which is based on an asymptotic analysis. However, Kim and Cho [2008] only remark on the general underestimation of failure loads (for all notch angles) and attribute this to the effect of a small but finite notch root radius. In contrast, the analysis by Ayatollahi et al. [2011] predicts failure accurately even in pure mode II conditions for $2\alpha = 270°$ when the analysis presented here suggests

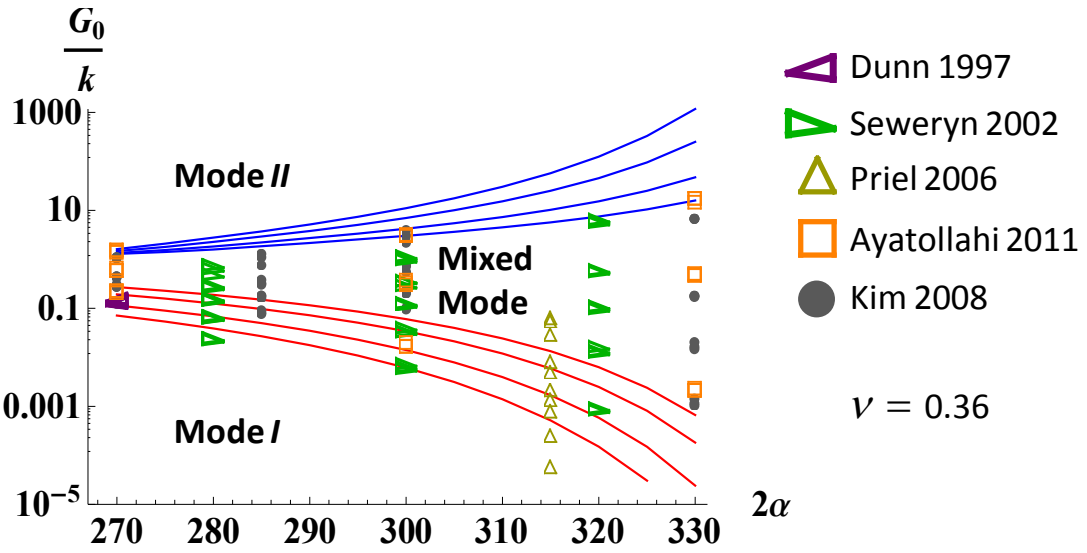


Figure 3.14: All the mixed-mode data from Figures 3.10-3.13 shown together with the mode-mixity information for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ (calculated assuming $\nu = 0.36$).

that the process zone is outside 15% small-scale yielding. Although this is somewhat surprising, similar findings are also reported by Dunn et al. [1997b]. This seems to imply the useful result that Williams' solution can be used to predict crack initiation even as higher-order terms begin to exert an influence on the process zone.

3.4.2 Mode-mixity of experimental data

Figures 3.10(a), 3.12(a), and 3.14 show that all the data reported by Dunn et al. [1997b] and the majority of those by Priel et al. [2007] resulted in a process zone that was mainly mode I in character. In fact, Figures 3.10(a) and 3.14 show that the tests by Dunn et al. [1997b] resulted in a process zone with a correlation coefficient of between $0.8 < \rho_I < 0.9$ with a pure mode I process zone. This is in agreement with the findings of Dunn et al. [1997b] as these authors were able to predict failure accurately based on a critical mode I generalized stress intensity factor alone. As can be seen in Figure 3.14, the tests carried out by Priel et al. [2007] extend over a wider range of mode-mixities than the tests by Dunn et al. [1997b] but still are only mildly mixed-mode in character at the level of the process zone. This finding is in agreement with the analysis of Priel et al. [2007] as

well as with remarks made by Ayatollahi et al. [2011] in their analysis of a subset of the data in [Priel et al., 2007].

The data reported by Ayatollahi et al. [2011], Seweryn et al. [1997] and Kim and Cho [2008] explore the full range of mode-mixities. However, for small notch angles ($2\alpha \lesssim 300^\circ$), this analysis illustrates that nominally mixed-mode loading conditions usually result in a process zone that is mainly mode I or only mildly mixed-mode in character. In fact, for these notch angles, the process zone is almost never found to be mainly mode II in character aside from pure mode II loading conditions. Thus, for practical engineering components, the process zone will often be mainly mode I in character. This is of course due to the difference in the order of the mode I, II singularities (see Figure 2.3). However, as $2\alpha \to 360^\circ$, this difference decreases; hence, for larger notch angles, mixed-mode conditions at the process zone become increasingly likely, and mainly mode II conditions become plausible.

3.5 Summary

In this chapter, further steps have been taken beyond those described by Hills and Dini [2011] to derive useful information from Williams' asymptotic solution for analysing notched test experiments. This is done by quantifying the mode-mixity of the process zone near the notch tip, which is the region that is relevant to crack nucleation and initiation. In addition, small-scale-yielding limits are calculated and shown simultaneously with the mode-mixity information on plots in generalized-stress-intensity-factor space, which provide information on the range of validity of this asymptotic analysis.

Application of this analysis to mixed-mode fracture data reported in the literature shows that a majority of these tests were conducted within 5% small-scale yielding. The only exception is for oblique notches subject to mode II dominated loading conditions for which the singularity of the stress field is relatively weak. The results also indicate that mode II domination of the stress field is very unlikely to occur for oblique notch angles and hence in practice; indeed, in many cases, the process zone will be mainly mode I in

character – even under nominally mixed-mode remote loading. The exception is for very acute notch angles for which mixed-mode conditions are most likely and mainly mode II conditions are also possible.

The techniques presented in this section can also be used in combination with the various brittle fracture criteria described in [Elices et al., 2002, Lazzarin and Zambardi, 2001, Lazzarin et al., 2009, Priel et al., 2007] to assess the likelihood of brittle failures in notched components. In addition, higher-order terms in Williams' solution could also be considered as is done by Kim and Cho [2008] and Cheng et al. [2012].

Chapter 4

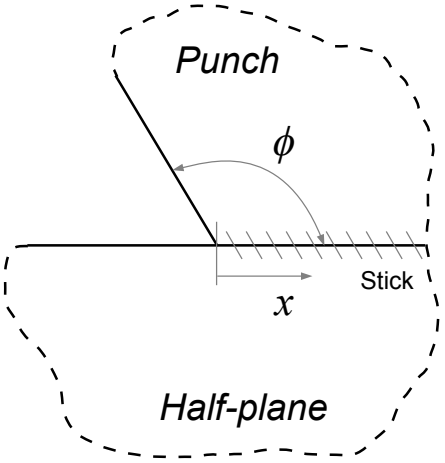
Asymptotic analysis of complete contacts

In this chapter, it is shown that Williams' solution can be applied to analyse sharp-edged contacts (i.e. *complete* contacts) under a range of loading conditions. This analysis enables the minimum friction coefficient required to prevent edge slip to be calculated, which depends only on the contact-edge angle. In addition, the combination of applied loads at which edge separation first occurs is determined. These calculations are then applied to the clamped-cantilever test rig that Juoksukangas et al. [2013] used to perform fretting-fatigue tests on complete contacts.

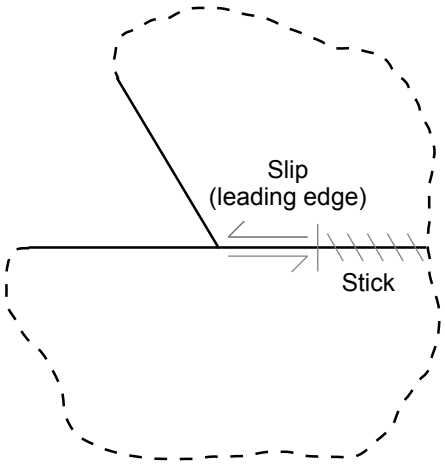
Note that parts of this chapter have been published in [Hills et al., 2014] and [Flicek et al., 2015b].

4.1 Introduction

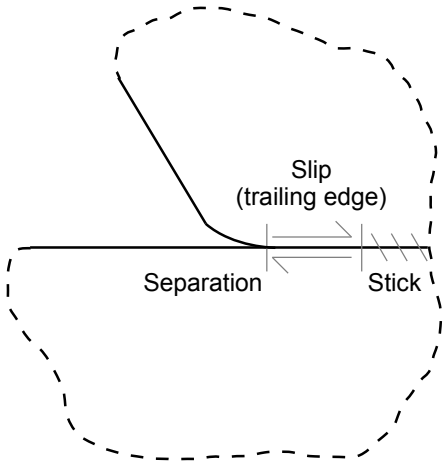
In general, the first steps towards predicting a contact's performance are: (i) to identify the region of the contact interface from which failure typically initiates and (ii) to determine the contact behaviour and stress state in the vicinity of this region. For complete contacts, failure commonly initiates near the sharp corner at the contact edge (see e.g. [Flicek et al., 2013, Juoksukangas et al., 2013]). However, this region is generally subject to a very complex multi-axial stress state even if it is assumed that the interface is completely stuck, such that no slip or separation occurs. Indeed, if the interface is completely



(a) Fully stuck.



(b) Leading-edge slip.



(c) Separation and trailing-edge slip.

Figure 4.1: Idealized diagrams of the three most likely types of contact-edge behaviour. The contact angle, ϕ, and the distance along the contact interface, x, are shown in (a).

stuck, the contact can be modelled as a notch using Williams' solution (see §2), which implies a singularity in the elastic stress state at the contact edge. This suggests that there will be increasingly high stresses and stress gradients as the contact edge is approached. As a consequence, it is difficult to obtain an accurate description of the stress state using numerical methods, e.g. the finite element method, and this is even more so when slip and/or separation also occur in this region. However, since failure tends to initiate near the contact edge, inaccuracies in the description of this region could lead to significant errors in predictions of component performance. Hence, it is very important to obtain an accurate description of the corner region.

One approach for modelling the edge of complete contacts, which is described in Hills et al. [2012a], is to use semi-infinite elasticity solutions to describe the contact edge in an asymptotic sense. To perform such an analysis, the behaviour at the contact edge must first be known since this determines which asymptotic form must be applied. The three most likely conditions to arise[1] are shown in Figure 4.1, which are: (a) closed and stuck, (b) closed and slipping in the leading-edge sense, and (c) open and slipping in the trailing-edge sense. Asymptotic solutions have already been presented in the literature that describe each of these conditions with a high degree of accuracy.

The simplest case is of course when the contact edge is closed and stuck as in Figure 4.1(a). Under these conditions, the displacement field across the contact interface is continuous, so a *bilateral* model can be used.[2] When the bodies are elastically similar, Williams' solution for a (semi-infinite) monolithic notch is the relevant bilateral solution. To apply this solution, the notch angle, 2α, is taken to be the full angle formed by the contact. For example, for the clamped cantilever shown in Figure 4.2, the angle is taken to be 270° (i.e. 180° for the specimen plus 90° for the clamping structure). If the bodies are not elastically similar, the Bogy solution [Bogy, 1968, 1971] must be used instead. However, the analysis in this chapter focuses on the elastically similar case, so only Williams' solution is required here. The other two common types of contact behaviour that are shown in Figure 4.1(b-c) are not considered in detail until §5.

In this chapter, it is shown that Williams' solution can be used to describe the edge of complete contacts. The primary output of this analysis is information on the range of loading conditions for which the bilateral contact condition is valid and also detailed information on when transitions in contact behaviour will occur (e.g. when the applied loads will result in a change in contact state from Figure 4.1(a) to Figure 4.1(c)). This asymptotic analysis is then used to re-examine fretting-fatigue tests that Juoksukangas

[1] An example of another contact-edge state that may occur under pure mode II loading conditions is trailing-edge slip but without edge separation, but this is quite unlikely to arise in practice.

[2] Note that *bilateral* is meant in the sense that the interface can transmit both tension and compression, so the solution varies linearly with the applied loads. In contrast, with a *unilateral* model, the interface can only support compression, and it separates when subjected to tension.

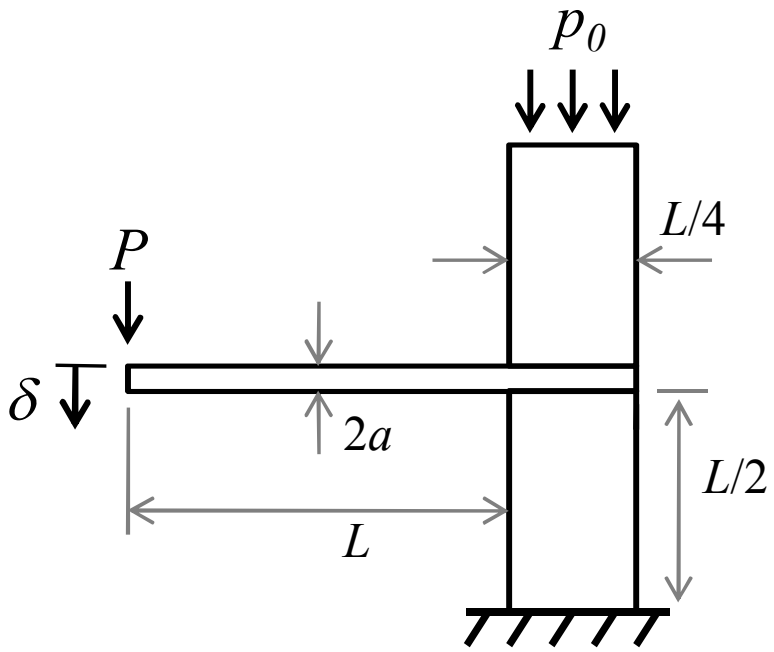


Figure 4.2: Idealized diagram of the clamped-cantilever test rig used by Juoksukangas et al. [2013].

et al. [2013] carried out using the clamped-cantilever test rig shown in Figure 4.2.

In the original analysis by Juoksukangas et al. [2013], the results of the fretting-fatigue tests are compared with some simple plain-fatigue tests on a standard S-N plot by assuming the stress state near the contact edge is given by elementary bending theory. However, since the stress state near the contact edge is of course much more severe than is suggested by beam theory, inferences cannot easily be drawn about other similar fretting configurations when the results are shown in this form. By re-examining the data using an asymptotic approach, it is shown that these results can potentially be used to infer the life of other fretting tests that use the same contact angle.

It is emphasized that the calculations in this chapter can be applied to contacts of arbitrary edge angle. However, the results are frequently applied to the particular case when $2\alpha = 270^\circ$ to facilitate their interpretation.

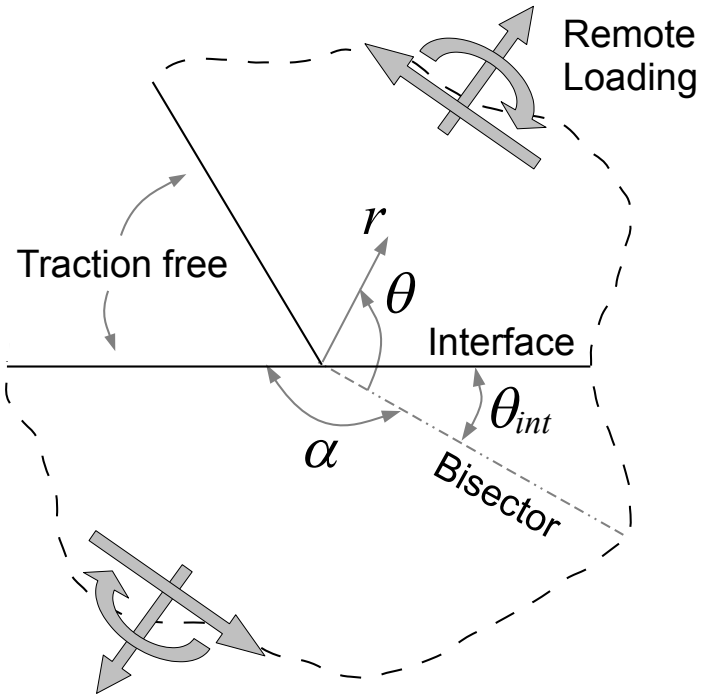


Figure 4.3: Idealized diagram of the edge of a complete contact, where (r, θ) is a polar coordinate set, θ_{int} is the angle corresponding to the contact interface, and 2α is the total angle included within the notch.

4.2 Implications of Williams' solution for adhered complete contacts

As noted in the previous section, the sharp corner formed at the edge of a complete contact can be modelled using Williams' solution provided that the bodies are made of elastically similar materials and that no slip or separation occurs along the interface. For these conditions to be met, the friction coefficient must be sufficiently high to prevent edge slip, and the applied loads must not cause the contact edges to separate. In §4.3, the range of applied loads for which these conditions hold is quantified. However, to begin this analysis, it is convenient simply to assume that both these conditions hold so that the bilateral assumption is valid. Following from this assumption, several properties of Williams' solution can be brought out clearly, and their relevance for complete contacts can be examined. In addition, when the bilateral assumption is valid, the calculations described

in §3 for assessing the mode-mixity and small-scale yielding of the process zone at the tip of sharp V-notches can also be applied to complete contacts without modification.

Note that although the formulation described in §2 can be used directly, it is convenient to modify the notation slightly. First, it is usually more intuitive to work in terms of the punch angle, ϕ, shown in Figure 4.1(a) instead of the full notch angle formed by the contact, 2α, shown in Figure 4.3. To do this, the full notch angle, 2α, must simply be replaced by

$$2\alpha = 180^\circ + \phi. \tag{4.1}$$

In this way, the contact is modelled as being between a half-plane (i.e. a notch of $2\alpha = 180^\circ$), and a semi-infinite wedge of interior angle ϕ. In addition, it is helpful to define θ_{int} as the angle corresponding to the contact interface, which is shown in Figure 4.3 and is given by

$$\theta_{int} = \frac{1}{2}(180^\circ - \phi). \tag{4.2}$$

It is also convenient to define a new distance coordinate along the contact interface, x, and this is shown in Figure 4.1(a). Thus, the normal, shear contact tractions, $p(x), q(x)$, respectively, can be defined as

$$p(x) = -\sigma_{\theta\theta}(r, \theta_{int}) \tag{4.3a}$$

$$q(x) = \sigma_{r\theta}(r, \theta_{int}), \tag{4.3b}$$

where the minus sign in (4.3a) arises because the contact pressure is taken to be positive in compression. It is noted that depending on the aim of the analysis, $\sigma_{\theta\theta}$ and $\sigma_{r\theta}$ in (4.3) can be obtained from (2.20), (2.27), or (3.4).

Finally, it is emphasized that this semi-infinite solution is only meant to model the region near the contact edge where the stress state is dominated by the singular terms in Williams' solution; it is not intended to model the overall finite geometry under consideration because higher-order terms will become increasingly important as the distance from

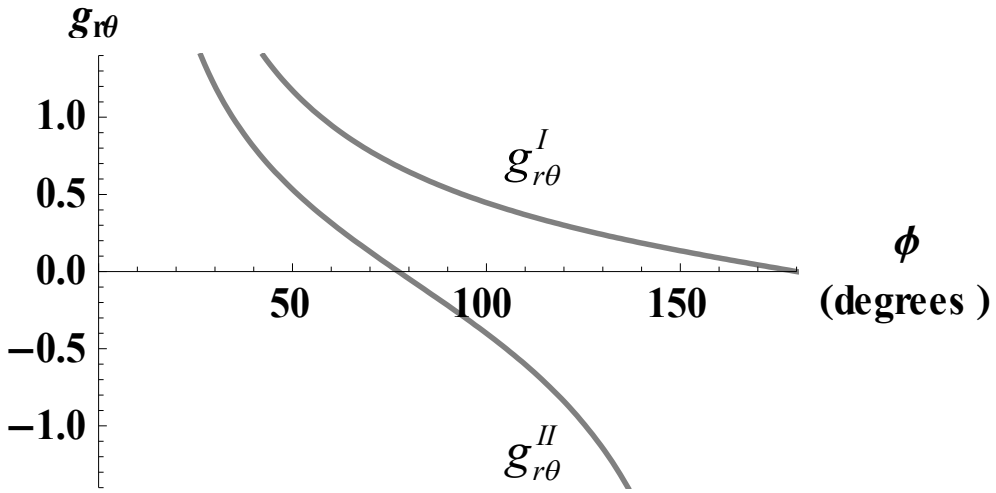


Figure 4.4: Traction ratio implied by Williams' solution near to ($g^{I}_{r\theta}$) and remote from ($g^{II}_{r\theta}$) the contact edge plotted against contact angle, ϕ.

the contact edge increases.

4.2.1 Properties of the mode I eigensolution

Consider a complete contact subjected to some arbitrary mixed-mode loading, i.e. loading that excites both mode I and mode II eigensolutions, but which results in the entire interface being completely stuck throughout the loading history (as in Figure 4.1(a)). In addition, recall the arguments presented in §2.3 showing that the mode-mixity of a two-term Williams expansion (2.20) is *not* self-similar with r and that $\lambda_I < \lambda_{II}$ (aside from the limiting case when $\phi = 180°$, which is not considered here). Thus, for $0° \leqslant \phi < 180°$, the mode I term in (2.20) is more strongly singular than the mode II term. An immediate implication of this property of the solution is that if the observation point is allowed to move arbitrarily close to the contact edge, it is always possible to find a region where the stress state is dominated by the mode I term.

Recall from §2.3 that d_0 can be viewed roughly as the distance from the contact edge at which each mode of the solution is equally influential. Since d_0 provides a length scale within the two-term Williams expansion, it is clear that the values of r for which the mode I term will dominate the stress state can be determined in relation to it: namely, the

stress state will be mode I dominated when $r \ll d_0$. Thus, when the observation point is *near* the contact edge (in the sense that $r \ll d_0$), the mode II term can be neglected, and the stress state can be represented with a one-term expansion of Williams' solution. In this way, the ratio of the normal and shear tractions near the contact edge, $g^I_{r\theta}$, can be expressed as

$$\frac{q(x)}{-p(x)} = \frac{\sigma^I_{r\theta}(r,\theta_{int})}{\sigma^I_{\theta\theta}(r,\theta_{int})} = \frac{f^I_{r\theta}(\theta_{int})}{f^I_{\theta\theta}(\theta_{int})} \equiv g^I_{r\theta} \tag{4.4}$$

by substituting stresses from (3.4) with $n = I$ into (4.3). Churchman and Hills [2006b] show that this ratio can be simplified to

$$g^I_{r\theta} = \frac{\sin\theta^+_i - \dfrac{\sin\theta^+_e}{\sin\theta^-_e}\sin\theta^-_i}{\cos\theta^+_i - \dfrac{\cos\theta^+_e}{\cos\theta^-_e}\cos\theta^-_i}, \tag{4.5}$$

where

$$\theta^-_i = (\lambda_n - 1)\theta_{int} \qquad \theta^-_e = (\lambda_n - 1)\theta_{ext} \tag{4.6a}$$

$$\theta^+_i = (\lambda_n + 1)\theta_{int} \qquad \theta^+_e = (\lambda_n + 1)\theta_{ext}, \tag{4.6b}$$

where $n = I$ in (4.6), and

$$\theta_{ext} = \alpha = \frac{1}{2}(180^\circ + \phi). \tag{4.7}$$

Notice from (4.4) that the traction ratio near the contact edge, $g^I_{r\theta}$, is independent of the finite geometry of the problem and depends solely on the angular functions given in (2.22) and (2.23), which in turn depend only on the notch angle (see §2.2.5). Moreover, perhaps surprisingly, this ratio is independent of the applied loads (provided only that $K_I \neq 0$, i.e. that the loading is not purely mode II in character) since these only enter the asymptotic solution through the multipliers K_I, K_{II}. This result that the traction ratio at the edge of a complete contact depends only on the contact angle is due to Hills and Dini [2004].

This quantity, $g^I_{r\theta}$, is plotted against contact angle, ϕ, in Figure 4.4, which shows that it

is a monotonically decreasing function of ϕ and that it is always positive. In addition, $g^I_{r\theta}$ ranges from being unbounded when $\phi = 0°$ to being zero when $\phi = 180°$. Although these limiting cases do not correspond to realistic contact geometries, it is instructive to examine them. This reveals that when $\phi = 180°$, the interface coincides with the bisector, where the mode I term results in zero shear stress (as is shown in Figure 2.4(a)). Conversely, when $\phi = 0°$, the interface is along a traction-free surface, so $p(x) = 0$.

Implications of the mode I solution

Recall that the analysis thus far assumes a bilateral contact condition, which, in part, implies that separation does not occur at any point along the interface. For this condition to be met, the contact pressure must be compressive (i.e. $p(x) > 0$). Writing out $p(x)$ and $q(x)$ near the contact edge (with a one-term expansion using (3.4) with $n = I$) gives

$$p(x) = -K_I x^{\lambda_I - 1} f^I_{\theta\theta}(\theta_{int}) \tag{4.8a}$$

$$q(x) = K_I x^{\lambda_I - 1} f^I_{r\theta}(\theta_{int}). \tag{4.8b}$$

It is clear from Figure 2.4(a) that $f^I_{\theta\theta}$ is always positive, so the condition $p(x) > 0$ is satisfied when $K_I < 0$. Hence, the results obtained in this section are only valid when the applied loads satisfy this condition, such that the contact edge is closed.

In addition, solely from consideration of the angular functions, f^I_{ij}, in (2.22), it can be seen that $K_I < 0$ is consistent with *leading-edge* shear tractions alone (i.e. with $q(x) < 0$), such that the punch slips outwards (as in Figure 4.1(b)). This implies the strong conclusion that if closure is maintained, only leading-edge slip is possible at the contact edge.[3]

If separation is implied near the contact edge, the concept of 'contact tractions' in this region is meaningless. Nevertheless, it is noted that when $K_I > 0$, $f^I_{\theta\theta}$ implies tensile normal tractions (and hence separation), and $f^I_{r\theta}$ implies *trailing-edge* shear tractions,

[3]This assumes that the applied loads do, in fact, excite the mode I eigensolution; indeed, this remark does not apply to a contact subjected to pure mode II loading. For example, for a contact of $\phi > 77.4°$ subjected to pure $K_{II} > 0$ loading, closure and *trailing-edge* slip are implied (see §4.2.2), but it is emphasized that such conditions are very unlikely to occur in practice.

such that the contact defining body separates and slips inwards (as in Figure 4.1(c)).

4.2.2 Properties of the mode II eigensolution

Recall that the mode II solution is less strongly singular than the mode I solution. Hence, whereas the mode I solution controls behaviour at the contact edge, the mode II solution exerts its strongest influence at a distance somewhat further away from the contact edge. However, it is important to emphasize that this entire asymptotic analysis is only valid in the vicinity of the contact edge where the singular terms in Williams' solution dominate the stress state (i.e. when $r \ll a$, where a is the distance to the nearest feature); indeed, as the observation point moves further from the contact edge, the neglected higher-order (bounded) terms will have an increasing influence on contact behaviour.

Nevertheless, consider a contact subjected to some arbitrary mixed-mode loading, and imagine that there exists a region somewhat *remote* from the contact edge where the mode II solution entirely dominates the state of stress (i.e. $r \gg d_0$ but also $r \ll a$). Such a region will not always exist in practice (e.g. if $d_0 \sim a$) because higher-order terms in the series expansion are likely to be important as the distance from the contact edge increases. Yet while quantitative results obtained using this remote (mode II) solution will be nowhere near as reliable as those obtained near the contact edge (with the mode I solution), qualitative information can still be obtained using this approach. Hence, in a region remote from the contact edge where the mode II term dominates the stress state, the traction ratio can be written as

$$\frac{q(x)}{-p(x)} = \frac{\sigma_{r\theta}^{II}(x,\theta_{int})}{\sigma_{\theta\theta}^{II}(x,\theta_{int})} = \frac{f_{r\theta}^{II}(\theta_{int})}{f_{\theta\theta}^{II}(\theta_{int})} \equiv g_{r\theta}^{II} \tag{4.9}$$

by using (3.4) and setting $n = II$. Churchman and Hills [2006b] show this can be simplified to give

$$g_{r\theta}^{II} = \frac{-\cos\theta_i^+ + \dfrac{\cos\theta_e^+}{\cos\theta_e^-}\cos\theta_i^-}{\sin\theta_i^+ - \dfrac{\sin\theta_e^+}{\sin\theta_e^-}\sin\theta_i^-}, \tag{4.10}$$

where the θ terms are given by (4.6) with $n = II$.

This traction ratio, $g^{II}_{r\theta}$, is plotted against ϕ in Figure 4.4. Unlike $g^{I}_{r\theta}$, which is positive for all contact angles, the mode II traction ratio, $g^{II}_{r\theta}$, changes sign when $\phi = 77.4°$. It is noted that this change in sign of $g^{II}_{r\theta}$ is associated with a change in sign of $f^{II}_{r\theta}(\theta_{int})$ (see Figure 2.4(b)). In addition, it can be seen from Figure 2.3 that this coincides with the mode II solution changing from being singular to being bounded, which occurs when $2\alpha = 257.4°$, i.e. when $(\lambda_{II} - 1) = 0$. Finally, examining the limiting cases as before reveals that $g^{II}_{r\theta}$ is unbounded for $\phi = 0°, 180°$. This is because the interface corresponds to a free surface when $\phi = 0°$, so $p(x) = 0$. Conversely, when $\phi = 180°$, the interface coincides with the bisector where $f^{II}_{\theta\theta}$ is zero (see Figure 2.4(b)), so $p(x) = 0$.

Implications of the mode II solution

The contact tractions remote from the contact edge can be written as

$$p(x) = -K_{II} r^{\lambda_{II}-1} f^{II}_{\theta\theta}(\theta_{int}) \tag{4.11a}$$

$$q(x) = K_{II} r^{\lambda_{II}-1} f^{II}_{r\theta}(\theta_{int}) \tag{4.11b}$$

by using (3.4) with $n = II$. It can be seen from Figure 2.4(b) that $f^{II}_{\theta\theta}$ is negative for $\theta > 0$. Since θ_{int} always corresponds to $\theta > 0$ as it is defined in Figure 4.3, the condition $K_{II} > 0$ must be satisfied for the contact pressure to remain compressive in the mode II region of the contact (where $r \gg d_0$ and $r \ll a$). However, note that a one-term expansion of the mode II solution should not be used to predict the onset of interior separation because higher-order terms are likely to become influential (and this is examined in more detail in §5). Instead, when $K_{II} < 0$, the only inference that should be drawn is that the solution is likely to be unreliable.

Again, solely from consideration of the angular functions, it can be seen that $K_{II} > 0$ is consistent with leading-edge shear tractions (i.e. $q(x) < 0$) for $\phi < 77.4°$ and trailing-edge shear tractions (i.e. $q(x) > 0$) for $\phi > 77.4°$. If instead $K_{II} < 0$, separation is implied in this remote region (although this may not be the case once this solution is

applied to a practical problem as is demonstrated in §5.5). In this situation, the angular functions imply tensile normal contact tractions, and the sense of the shear tractions switches from leading-edge to trailing-edge in character and *vice versa*. However, note that these shear tractions occur away from the contact edge, and the terms 'leading-edge' and 'trailing-edge' refer only to the sign of the shear traction, not to the region of the contact in which they arise.

One other noteworthy feature of the solution relates to the relative magnitude of the near and remote traction ratios: namely, Figure 4.4 shows that $g^I_{r\theta} > |g^{II}_{r\theta}|$ for contact angles less than $101.7°$. Thus, for these contact angles, the traction ratio is always greater at the contact edge than it is at the interior of the contact. This suggests that a coefficient of friction, f, sufficient to inhibit slip near the contact edge is also sufficient to inhibit slip remote from the contact edge for these contact angles (but only within the asymptotic region, of course). Conversely, for $\phi > 101.7°$, slip is likely to initiate from the interior of the contact interface.

4.3 Application to clamped-cantilever test rig

The results derived in the previous sections are now applied to the example complete contact problem shown in Figure 4.2 to illustrate their practical implications. For this problem, $\phi = 90°$, so $(g^I_{r\theta}, g^{II}_{r\theta}) = (0.543, -0.219)$, and $(\lambda_I, \lambda_{II}) = (0.5445, 0.9085)$. Thus, given knowledge of the contact angle alone, the traction ratio at the contact edge, $g^I_{r\theta}$, can be determined. Hills and Dini [2004] have shown that this also determines the minimum coefficient of friction required to prevent slip at the contact edge, which is $f > g^I_{r\theta} = 0.543$ for this geometry. Thus, if the friction coefficient is lower than this, there will certainly be edge slip. Notice that no details of the finite geometry are required for this result.

If $f > g^I_{r\theta}$, it is still possible for slip to occur if the applied loads result in edge separation. However, the region where separation occurs can be very small, so it may be difficult to detect using a standard finite element approach. For this reason, it is more reliable to

look at the multipliers on the eigensolutions and to observe that for the normal traction to remain compressive at the corner (i.e. $p(0) > 0$), the condition $K_I < 0$ must hold. Thus, to determine if the contact edge will separate, calibrations of the generalized stress intensity factors must be calculated. Calibrations for the notional clamping pressure, p_0, (which has units of $[FL^{-2}]$) and the tip load P (which has units of force per unit width $[FL^{-1}]$) were obtained from a bilateral finite element model and are given by[4]

$$\left\{ \begin{array}{c} K_I a^{\lambda_I - 1} \\ K_{II} a^{\lambda_{II} - 1} \end{array} \right\} = \left[\begin{array}{cc} -0.238 & 16.6 \\ 0.414 & 14.1 \end{array} \right] \left\{ \begin{array}{c} p_0 \\ P/a \end{array} \right\}. \tag{4.12}$$

Note that these calibrations relate to the *upper* contact edge and were computed for the case when $L/a = 40$. An image of the finite element model that was used to obtain these calibrations is shown in Figure 4.5.

Examination of the calibrations of K_I shows that, as intuitively expected, the clamping pressure, p_0, induces closure (i.e. $K_I < 0$). On the other hand, a positive (downward) cantilever tip load, i.e. $P > 0$, excites edge-initiated separation at the top corner (i.e. $K_I > 0$), whereas a negative (upwards) tip load, $P < 0$, encourages further closure at that point. If these calibrations are used with the condition $K_I < 0$, the requirement for maintaining closure at both contact edges is found to be

$$\left| \frac{P}{p_0 a} \right| < 0.0143. \tag{4.13}$$

Therefore, an asymptotic analysis of the contact edge reveals that for the bilateral assumption to be valid near both contact edges, the following conditions must hold: (i) $|P/(p_0 a)| < 0.0143$ to prevent edge separation, and (ii) $f > g^I_{r\theta} = 0.543$ to prevent edge slip. In addition, since $|g^I_{r\theta}| > |g^{II}_{r\theta}|$, a friction coefficient sufficient to prevent edge slip

[4]These calibrations were calculated by constructing a two-dimensional, plane-strain, bilateral finite element model of the cantilever geometry in ABAQUS/CAE version 6.11, which incorporated a combination of `CPE4R` and `CPE3` elements. With this model, the stress state near the corner resulting from a unit application of each applied load was determined. These stresses were then used to plot the right hand side of (2.25a) and (2.25b) against r to view the value to which this quantity converged as $r \to 0$ (as in Figure 2.5). A mesh refinement study was also performed in which the fineness of the mesh was increased until the resulting calibrations did not change.

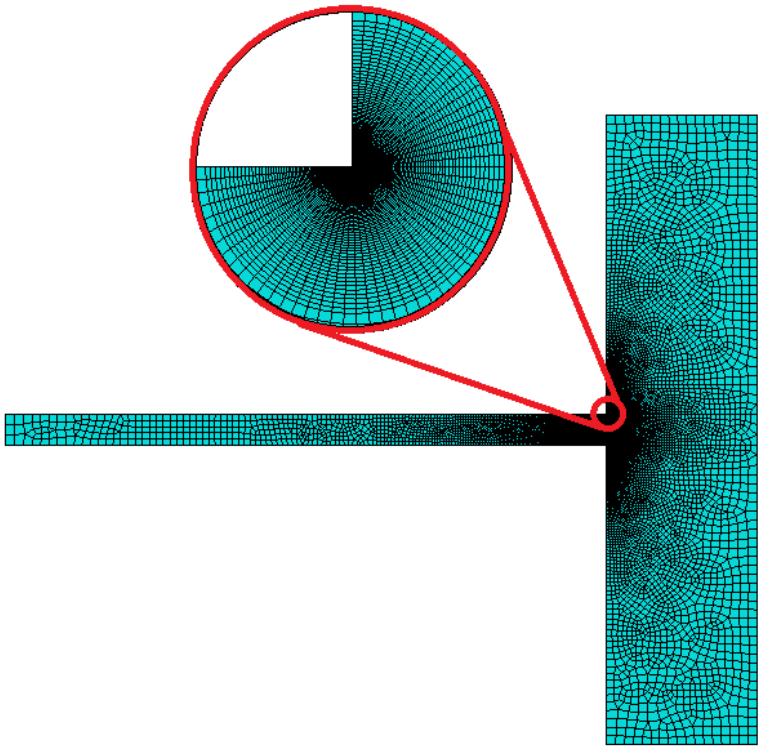


Figure 4.5: Image of the finite element model of the clamped-cantilever test rig.

will also prevent slip further from the edge (where the mode II term becomes influential). Moreover, these results have been obtained purely from a bilateral model, so none of the numerical difficulties or computational expense associated with using a unilateral contact model are encountered.

4.3.1 Application to fretting-fatigue tests

The dimensionless results presented above can easily be specified to the actual test results obtained by Juoksukangas et al. [2013]. It is noted that the dimensions of the clamped-cantilever test rig used by the authors are $a = 5mm$ and $L = 200mm$ (with a depth of $40mm$ into the page in Figure 4.2), and all the fretting-fatigue tests reported by the authors used a nominal clamping pressure of $p_0 = 100Nmm^{-2}$. Hence, at this level of clamping pressure, separation is implied to occur at the contact edge when the magnitude of the cantilever tip load exceeds $|P| = 7.2N$, which corresponds to an applied displacement of $\delta = \pm 1.14mm$. The relationship between the cantilever tip load, P, and

the bending stress at the cantilever root given by elementary bending theory, σ_a, is

$$P = \frac{2a^2}{3L}\sigma_a, \tag{4.14}$$

so separation will occur when the apparent bending stress reaches $\sigma_a = \pm 86 N mm^{-2}$.

Juoksukangas et al. [2013] note that the fretting tests were conducted by first applying a clamping pressure of $p_0 = 100 N mm^{-2}$ and then applying a fully reversing cantilever tip load, i.e. cycling between $-P$ and $+P$. The tests were conducted such that the amplitude (i.e. half the range) of the apparent bending stress at the contact edge was within the range $116 N mm^{-2} < \sigma_a < 303 N mm^{-2}$. This corresponds to $0.02 < |P/(p_0 a)| < 0.05$, so separation occurred in all the tests that were carried out. Thus, even if the friction coefficient in these tests was greater than $g^I_{r\theta} = 0.543$, frictional slip and hence fretting conditions certainly were present in all these tests (as the authors intended) because a slip zone must arise adjacent to the separated edge region.

4.4 Re-examination of fretting-fatigue data

In their paper, Juoksukangas et al. [2013] report the results of two sets of fatigue tests on a steel alloy, which they refer to as 'quenched and tempered 34CrNiMo6 steel': one set in fretting fatigue and the other in plain fatigue. Both sets of tests were carried out using the clamped-cantilever test rig shown in Figure 4.2 and resulted in brittle fracture of the test specimen. However, the crack initiation site differed in these sets of tests. For the fretting-fatigue tests, failure initiated at the root of the cantilever, whereas for the plain-fatigue tests, the initiation site was located at some distance from the cantilever root at a free edge of the specimen. To cause cracks to nucleate away from the cantilever root, the authors tapered the width of the specimen from $40mm$ for the fretting specimens to $12mm$ for the plain-fatigue specimens with a radius of $50mm$ near the cantilever root.

Based on a finite element analysis, the authors conclude that tapering the plain-fatigue specimens in this way produces a mild stress concentration factor of 1.05 in the region

where the plain-fatigue cracks initiated. However, the stress state in these tests is still almost uni-axial throughout the crack initiation region. Based on these tests, the authors find the fatigue limit in plain fatigue to be $517Nmm^{-2}$. For the fretting-fatigue tests, Juoksukangas et al. [2013] perform a similar analysis in which they assume the stress state near the root of the cantilever, σ_a, to be given by beam theory. Based on this analysis, the authors find the fatigue limit for the fretting configuration to be $\sigma_a^{fl} = 172Nmm^{-2}$. The authors then plot both sets of test results on an S-N plot and note that the fatigue life of the fretting specimens is a factor of 3 lower than that of the plain-fatigue specimens.

However, the stress state near the cantilever root incorporates an implied singularity in the elastic state of stress, which is inherently multi-axial in character. Although this stress singularity will be truncated by plasticity and by the local root radius (which Juoksukangas et al. [2013] estimate to be $20\mu m$), it is clear that the stress state in this region is very different from the almost uni-axial stress state present in the plain-fatigue tests. Thus, comparison of these two very different stress states on an S-N plot is not the optimal approach for obtaining data that can be used to predict the performance of similar fretting-fatigue test configurations.

A better approach is first to assume that any slip or separation that occurs is small compared to the region over which Williams' solution dominates behaviour (i.e. that it is smaller than d_0). In addition, it is assumed that the loading is light so that the plastic zone, r_p, at the cantilever root is small: that is, (i) that $r_p \ll a$, so the process zone is within small-scale yielding; and (ii) that $r_p \ll d_0$, so the characteristic of the process zone is dominated by the mode I eigensolution (see §3).

The validity of these assumptions is examined in §6, but assuming they hold, the fretting-fatigue life is likely to be determined by the amplitude of the mode I generalized stress intensity factor, ΔK_I, resulting from the cantilever tip load, P. Moreover, ΔK_I may still control fatigue life even if the bilateral assumption is not entirely valid. This is because the characteristic of the slip and separation zones will still be controlled by Williams' solution if the slip/separation zone is small (in comparison to d_0). Hence,

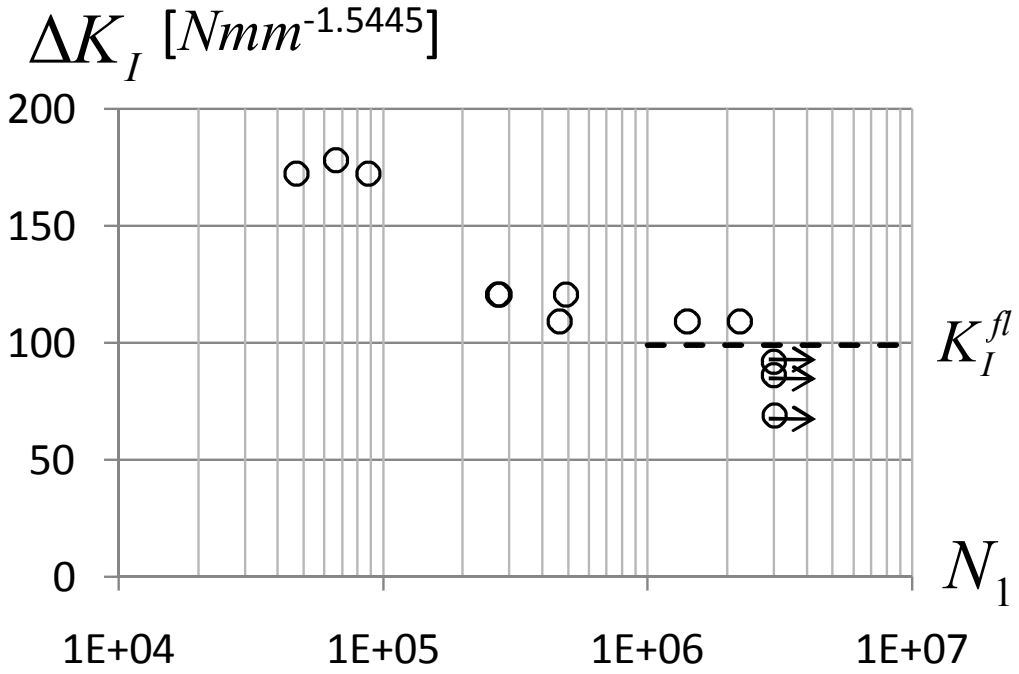


Figure 4.6: Semi-log plot of ΔK_I $(Nmm^{-1.5445})$ vs. cycles to macroscopic crack nucleation N_1. The fatigue limit of $K_I^{fl} = 99.1Nmm^{-1.5445}$ is also shown.

a more general plot can be obtained by plotting ΔK_I against the number of cycles to macroscopic crack initiation, N_1.

It can be seen from (4.12) that the clamping pressure of $p_0 = 100Nmm^{-2}$ produces a static contribution of $K_I = -49.6Nmm^{-1.5445}$ for the duration of the fretting-fatigue test. In addition, the amplitude (i.e. half the total range) of the variation in the mode I generalized stress intensity factor, ΔK_I, resulting from the cantilever tip load, P, can be calculated as

$$\Delta K_I = \pm 16.6 a^{-\lambda_I} P = \frac{\pm 11.1 a^{2-\lambda_I}}{L} \sigma_a \tag{4.15}$$

using (4.12) and (4.14). It is emphasized that the value calculated for ΔK_I accounts for the amplitude of the *variation* of mode I generalized stress intensity factor at the cantilever root and *neglects* the static contribution from the clamping load, p_0. The S-N plot can now be converted into a plot of ΔK_I vs. N_1 using (4.15), and this is shown in Figure 4.6. Note that the fatigue limit of $\sigma_a = 172Nmm^{-2}$ that was found by Juoksukangas et al. [2013] corresponds to $\Delta K_I^{fl} = 99.1Nmm^{-1.5445}$, and this is also shown in Figure 4.6.

Note that the asymptotic stress field represented by ΔK_I only controls the corner stress state during crack nucleation. Therefore, while Juoksukangas et al. [2013] provide both the number of cycles to macroscopic crack initiation, N_1, and also the total number of cycles to failure, only the former of these is examined here.

4.5 Discussion

The main reason for advocating the use of asymptotic forms for the quantification of crack nucleation is that tests then reveal true material properties, which can be used for design purposes with other geometries. Note that the fatigue limit found in this set of tests (in terms of K_I^{fl}) is specific to the particular (static) clamping load that was used (i.e. $K_I = -49.6Nmm^{-1.5445}$), and it is not known how strongly the fatigue limit, K_I^{fl}, depends on the static clamping pressure, p_0.

Nevertheless, suppose K_I^{fl} is taken to be the fretting-fatigue limit since slip and separation certainly occurred at the contact edges during these tests. Now suppose that the same cantilever geometry is subjected to a remote tension at the cantilever tip of magnitude, σ_0. Calibrations of the generalized stress intensity factors for this load case are[5]

$$\left\{ \begin{array}{c} K_I a^{\lambda_I - 1} \\ K_{II} a^{\lambda_{II} - 1} \end{array} \right\} = \left[\begin{array}{c} 0.698 \\ 0.168 \end{array} \right] \sigma_0, \tag{4.16}$$

which also relate to the *upper* contact corner.

If the mode II generalized stress intensity factor range, ΔK_{II}, does not influence crack nucleation, the fatigue limit in terms of ΔK_I, which was derived from the cantilever bending test data, can be applied to this new load case as well. Whether ΔK_I alone controls crack nucleation can be investigated using the tools presented in §3, and this type of analysis is considered in §6. In any case, suppose that the same clamping load (i.e.

[5]The calibrations for the bulk load, σ_0, are very sensitive to the seemingly remote geometry and boundary conditions to an extent that the calibrations for the clamping and bending loads are not. Therefore, it is noted that the boundary conditions for the model from which the calibrations for the bulk load were derived were: zero horizontal displacement along the top and bottom surfaces of the two contact pads, and zero vertical displacement at a single point at the rightmost edge of the model along its horizontal centreline.

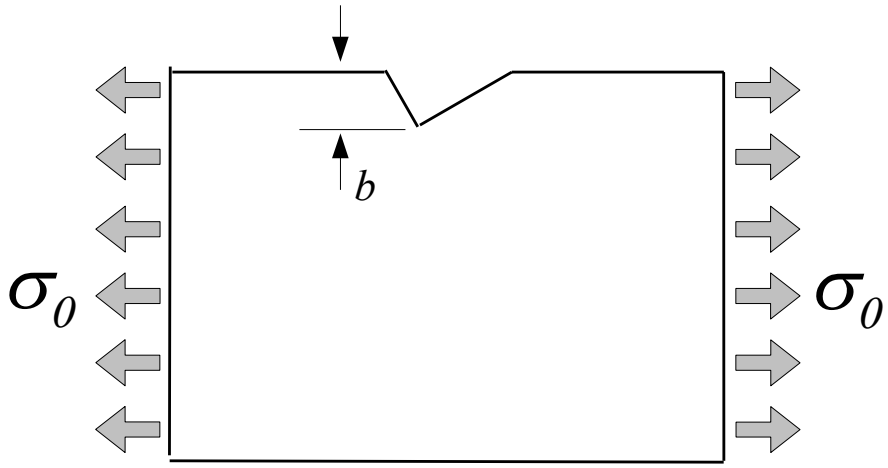


Figure 4.7: Notched plate studied by Hills and Dini [2011].

$p_0 = 100Nmm^{-2}$) is applied and that the bulk load, σ_0, is subsequently cycled in the fully reversing sense. Assuming that ΔK_I controls crack nucleation, the fretting-fatigue limit derived from the bending tests (i.e. $K_I^{fl} = 99.1Nmm^{-1.5445}$) can be applied to this new load case, which gives a fatigue limit of $\Delta\sigma_0 = 68.2Nmm^{-2}$, where $\Delta\sigma_0$ represents the amplitude (i.e. half the range) of the variation of the bulk load.

If instead the tests run by the authors resulted in the contact edges being fully stuck, such that they represented *notch* conditions (although this is not the case for these particular tests), then K_I^{fl} could be taken as the threshold for notch fatigue. Hence, consider the large plate shown in Figure 4.7, which incorporates an oblique edge notch of depth b with a total included angle in the solid of $270°$ oriented at $30°$ to the edge normal. Calibrations of the generalized stress intensity factors when this plate is subjected to a remote tension, σ_0, are given in Hills and Dini [2011] as

$$\left\{ \begin{array}{c} K_I b^{\lambda_I - 1} \\ K_{II} b^{\lambda_{II} - 1} \end{array} \right\} = \left[\begin{array}{c} 0.924 \\ 0.327 \end{array} \right] \sigma_0. \tag{4.17}$$

Now suppose an experiment were carried out with $b = 10mm$, that a (static) mean load of $\sigma_0 = -18.8Nmm^{-2}$ were applied (producing a mean value of $K_I = -49.6Nmm^{-1.5445}$ as in the bending tests), and that σ_0 were cycled about this mean value. Assuming, again,

that the mode I stress field controls crack nucleation, the fatigue limit in terms of K_I^{fl} would remain unchanged. Applying the fatigue limit from the bending tests to the load case just mentioned implies a fatigue limit of $\sigma_0 = 37.6Nmm^{-2}$. Thus, if fatigue data were collected for either of the load cases described above, they could be plotted on Figure 4.6 along with the bending data that are presented in the original paper.

4.6 Summary

In this chapter, Williams' solution is applied to analyse the edge of complete contacts. This is done by assuming a bilateral contact condition, which is valid when no slip or separation occurs near the contact edges. Under these conditions, the ratio of the normal and shear tractions at the contact edge depends only on the contact angle, ϕ, and this in turn determines the minimum coefficient of friction required to prevent slip at the contact edges. In addition, the combination of applied loads that results in edge separation is accurately determined by computing calibrations of the mode I generalized stress intensity factors from a bilateral finite element model. These calculations are then applied to a set of fretting-fatigue tests performed on complete contacts by Juoksukangas et al. [2013]. Finally, the benefit of applying an asymptotic analysis is demonstrated by showing how these data could be used to infer the fretting performance of other test configurations of the same contact angle.

Chapter 5

Slip and separation at the contact edge

This chapter builds on the analysis presented in the previous chapter by providing an estimate of the extent of slip and/or separation at the edge of a complete contact. First, the types of contact-edge responses that are likely to arise under fretting conditions are discussed. Each of these contact-edge responses is then examined, and estimates of the extent of slip and/or separation are obtained in closed form. When the extent of slip and/or separation is implied to remain near the contact edge, an order of magnitude estimate of its extent is obtained. Conversely, only qualitative implications can be derived when larger slip or separation lengths are implied. As in the previous chapter, these calculations are then applied to the fretting-fatigue tests carried out by Juoksukangas et al. [2013].

Note that parts of this chapter have been published in [Hills et al., 2013] and [Flicek et al., 2015b].

5.1 Introduction

Recall from §4 that the three most likely conditions to arise at the edge of a complete contact are: (i) closed and stuck, (ii) closed and slipping in the leading-edge sense, and (iii) open and slipping in the trailing-edge sense, and these are shown in Figure 4.1. The previous chapter focuses mainly on the closed-and-stuck case and shows how to determine the conditions under which this contact state will arise. The focus of this chapter is the other two of these contact states, which are associated with slip and/or separation at the contact edge.

Under these conditions, the contact edge can be modelled using the distributed dislocation technique, which is described in detail by Hills et al. [1996]. This approach enforces the *unilateral* contact conditions precisely, and enables the true extent of slip and/or separation at the contact edge to be calculated (in the sense that it accounts for the redistribution of contact tractions resulting from slip and/or separation). This procedure involves performing a numerical inversion of Cauchy singular integral equations, which can be a rather involved process and must be carried out separately for each contact angle, ϕ. Nevertheless, this has been done for the closed-and-slipping case for contact between a quarter-plane and an elastically similar half-plane by Churchman and Hills [2006c]. More recently, the open-and-slipping case has also been solved for the same problem by Paynter et al. [2010].

While the true extent of slip and/or separation can be calculated as described above, the level of effort required to perform such an analysis may be a barrier to applying it in practice if results have not already been published for the relevant contact angle. Hence, the intention of this chapter is to present a more modest approach for describing the behaviour at the edge of a complete contact, but which is very easy to apply to any contact angle, ϕ. This is done by obtaining a closed-form estimate of the extent of slip and/or separation near the contact edge, which is based on violations of the Signorini conditions (i.e. it does not account for the redistribution of contact tractions due to slip or separation).

Although this approach does not quantify the extent of slip and/or separation with great precision, the main application of these calculations is to analyse fretting-fatigue experiments, which of course involve cyclic loading. Due to the process of *frictional shakedown* during cyclic loading (see e.g. [Barber et al., 2008]), the slip extent can be expected to be smaller after several load cycles than it is in the first cycle. Moreover, the slip/stick behaviour that arises after several load cycles is likely to have a greater influence on fatigue performance than the slip extent during the initial load application. This is simply because contacts spend a much greater proportion of their service life subject to the conditions that result after several load cycles than to the conditions present during

the first load cycle.

However, unlike incomplete contacts, which almost always have a slip zone at the edge of contact, it can be difficult to determine whether the edge of a complete contact will stick, slip, or separate – even under monotonic loading conditions. In addition, obtaining a sufficiently accurate model of the contact interface using the finite element method is very computationally expensive. Thus, the aim of these calculations is to provide a quick and efficient means of estimating the extent of slip during the initial load application, which provides a qualitative indication of the conditions that will result at the contact edge due to a given loading regime. This enables fretting experiments to be designed that excite a specific type of contact-edge response in order to examine its effect on fatigue performance.

5.2 Types of contact-edge response

In §4, it is shown that the stress state sufficiently near the edge of an adhered complete contact can be represented using a two-term expansion of Williams' solution. Furthermore, recall from §2.2.6 that the influence of the remote loads is fully quantified by the multipliers on each eigensolution: K_I, K_{II}. There are clearly four possible sign combinations of K_I, K_{II} that may arise, and each of these has qualitatively different implications for the contact-edge response. These are:

1. Case (a): $K_I < 0 < K_{II}$: Both eigensolutions imply closure.

2. Case (b): $0 < K_I, K_{II}$: The mode I term implies edge separation, and the mode II term implies interior closure.

3. Case (c): $K_I, K_{II} < 0$: The mode I term implies edge closure, and the mode II term implies interior separation.

4. $K_{II} < 0 < K_I$: Both eigensolutions imply separation.

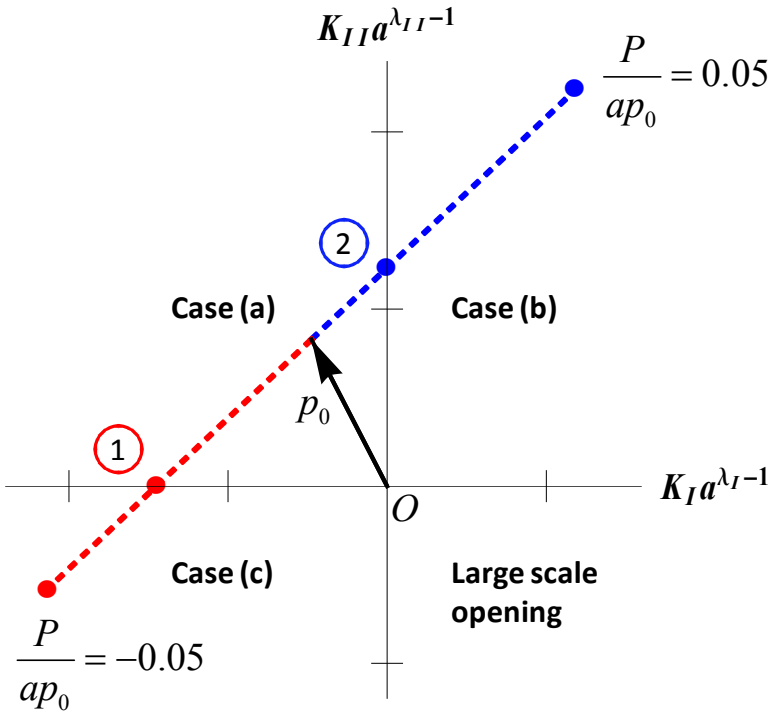
Notice that the implications of $K_{II} < 0 < K_I$ completely violate the monolithic notch geometry that is assumed when modelling the contact using Williams' solution, so this case is not considered any further. Indeed, the implication of significant separation is all that can be drawn from an asymptotic analysis in this case.

The implications of the remaining three sign combinations are considered in the following section. However, prior to examining these in detail, it is helpful to investigate which sign combinations are likely to arise in practice. To do this, recall the fretting-fatigue tests that Juoksukangas et al. [2013] performed with the clamped-cantilever test rig, which is shown in Figure 4.2. In these tests, the clamping pressure, p_0, was first applied and held constant, and then a cantilever tip load, P, was applied cyclically in a fully reversing sense. It is noted that the largest load range used in these fretting tests was $-0.05 < P/(p_0 a) < 0.05$.

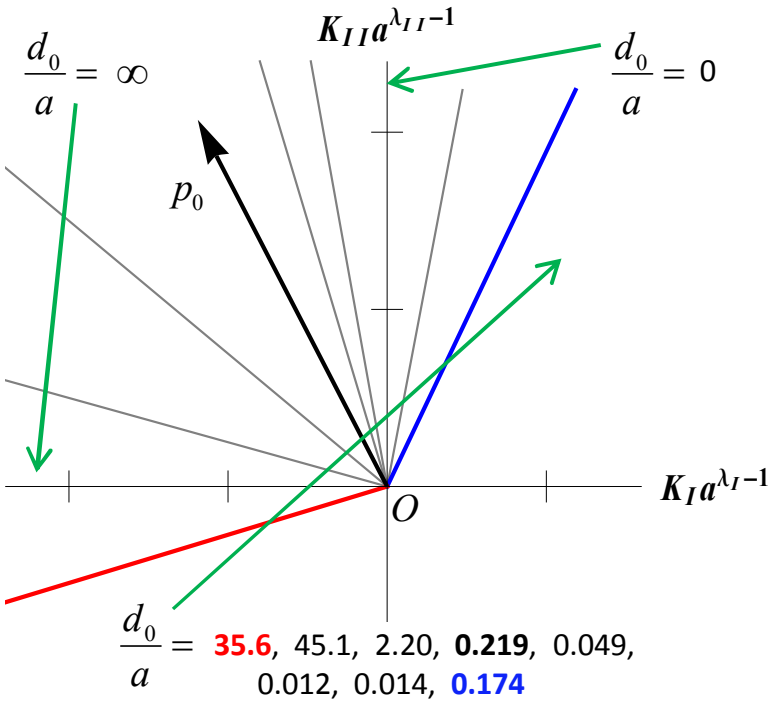
If a bilateral contact interface is assumed, this loading trajectory (for the *upper* contact corner) can be plotted in generalized-stress-intensity-factor space, i.e. on a plot of $K_{II} a^{\lambda_{II}-1}$ vs. $K_I a^{\lambda_I - 1}$, using the calibrations given in (4.12). This is done in Figure 5.1(a), which also includes other information that is explained in §5.4. This figure illustrates that application of p_0 drives the loading trajectory into case (a), which implies closure throughout the contact-edge region. Thus, for case (a), there is either full stick (i.e. Figure 4.1(a)) or leading-edge slip with interior stick (i.e. Figure 4.1(b)) depending on the level of friction.

Subsequent application of the cantilever tip load, P, produces a trajectory with a positive slope. A positive (downwards) application of P drives the *upper* corner towards case (b), which results in the open-and-slipping behaviour shown in Figure 4.1(c). Conversely, a negative (upwards) application of P drives the *upper* corner towards case (c), where K_I implies closure, but K_{II} implies opening. However, in practice, case (c) will usually result only in a reduction in contact pressure and not in interior opening (see §5.5).

Notice that the large-scale separation case (i.e. $K_{II} < 0 < K_I$) never results from any combination of clamping pressure and tip load. Furthermore, it is clear from the



(a) Loading trajectory used by Juoksukangas et al. [2013] to conduct complete contact fretting-fatigue tests.



(b) Values of d_0/a during the loading trajectory shown in (a).

Figure 5.1: Plots in generalized-stress-intensity-factor space.

calibrations given in (4.16) that the application of a bulk load to the tip of the specimen, σ_0, also results in a trajectory with a positive slope, although this is not shown in Figure 5.1. Hence, large-scale separation cannot result from any combination of clamping pressure and bulk load. In addition, several other loading trajectories for different geometries and loading conditions are examined by Hills et al. [2012a], and these authors also find large-scale separation to be unlikely to occur.

5.3 Calculation of the extent of slip and/or separation

In this section, closed-form estimates of the extent of slip and/or separation are obtained for cases (a-c). For this analysis, it is assumed that the loads are applied proportionally or, more generally, that the load path is such that, as time progresses, any zones of slip or separation remain either of fixed length or increase in size. Note that in order to bring out the implications of the results more clearly, the calculations in this section are performed for the particular contact angle of $\phi = 90°$. The results are then extended to the full range of contact angles in §5.6.

5.3.1 Case (a): $K_I < 0 < K_{II}$

In this case, both the mode I and the mode II terms in Williams' solution imply a compressive contact pressure, $p(x) > 0$, and hence closure throughout the contact-edge region (as in Figure 4.1(a-b)). Under these conditions, edge slip will be inhibited provided only that $f > g^I_{r\theta}$, which corresponds to $f > 0.543$ when $\phi = 90°$ (see §4.3). Conversely, if $f < g^I_{r\theta}$, there will be *leading-edge* slip at the contact edge (as defined in Figure 4.1(b)).

An order of magnitude estimate of the slip extent can be obtained based on violations of the Coulomb friction law, which states that slip will occur if the following inequality is violated

$$|q(x)| \leqslant fp(x). \tag{5.1}$$

For this calculation, it is important to make use of the alternative set of parameters, d_0, G_0,

(see §2.3) for reasons that will become clear below. However, a slight modification must first be made to their definition.

In (2.26), modulus signs are included in the definition of d_0, G_0 to eliminate the sign of the K_{II}/K_I ratio and to ensure that the argument that is raised to a fractional power is always positive. However, for these calculations, it is important to preserve the information regarding the sign of K_I, K_{II}, while still making sure a negative number is not raised to a fractional power. For case (a), K_I is negative, so a new parameter can simply be defined as

$$\tilde{K}_I = -K_I, \tag{5.2}$$

the modulus signs in the definition of d_0, G_0 can then be removed, and these parameters can be redefined as

$$d_0 = \left(\frac{K_{II}}{\tilde{K}_I}\right)^{\frac{1}{\lambda_I - \lambda_{II}}} \tag{5.3a}$$

$$G_0 = \tilde{K}_I^{\frac{\lambda_{II}-1}{\lambda_{II}-\lambda_I}} K_{II}^{\frac{\lambda_I - 1}{\lambda_I - \lambda_{II}}}. \tag{5.3b}$$

Thus, (2.27) can be rewritten as

$$\frac{\sigma_{ij}(r,\theta)}{G_0} = -\left(\frac{r}{d_0}\right)^{\lambda_I - 1} f_{ij}^{I}(\theta) + \left(\frac{r}{d_0}\right)^{\lambda_{II}-1} f_{ij}^{II}(\theta), \tag{5.4}$$

where d_0, G_0 are given by (5.3).

Using (5.4) and (4.3), the ratio of the contact tractions can be written as

$$\frac{q(x)}{p(x)} = -\left(\frac{-\left(\frac{x}{d_0}\right)^{\lambda_I - 1} f_{r\theta}^{I}(\theta_{int}) + \left(\frac{x}{d_0}\right)^{\lambda_{II}-1} f_{r\theta}^{II}(\theta_{int})}{-\left(\frac{x}{d_0}\right)^{\lambda_I - 1} f_{\theta\theta}^{I}(\theta_{int}) + \left(\frac{x}{d_0}\right)^{\lambda_{II}-1} f_{\theta\theta}^{II}(\theta_{int})}\right). \tag{5.5}$$

This expression demonstrates the benefit of using d_0, G_0 instead of K_I, K_{II} when writing the traction ratio: namely, G_0 cancels out, which shows that the traction ratio within the asymptote is independent of the *magnitude* of the remote loads. Indeed, it only depends on the mode-mixity, which is accounted for by d_0.

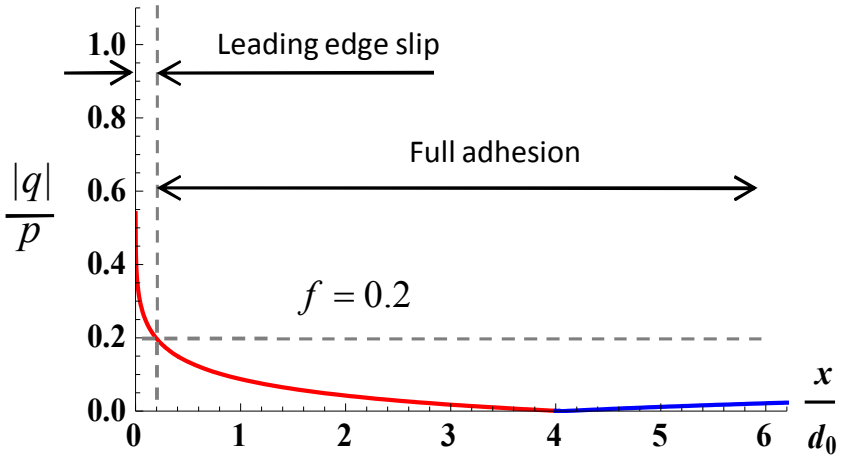
The magnitude of the traction ratio, $|q(x)|/p(x)$, is plotted against position within the asymptote, x/d_0, in Figure 5.2(a). This figure shows that the traction ratio passes through zero when the shear traction changes from being leading-edge ($q(x) < 0$) in character near the contact edge to being trailing-edge ($q(x) > 0$) in character when $x/d_0 > 4$. A horizontal dashed line is shown in this figure, which corresponds to an example coefficient of friction of $f = 0.2$. A vertical dashed line is also included that illustrates the extent of slip within the asymptote, x_s/d_0, that is implied by violations of (5.1) when $f = 0.2$.

Since $(g^I_{r\theta}, g^{II}_{r\theta}) = (0.543, -0.219)$ when $\phi = 90°$, for friction coefficients within the range $0.543 > f > 0.219$, there will be *leading-edge* slip with interior stick, and this contact response is shown in Figure 4.1(b). An estimate of the extent of slip, x_s, can be obtained in closed form based on violations of the Coulomb friction condition as

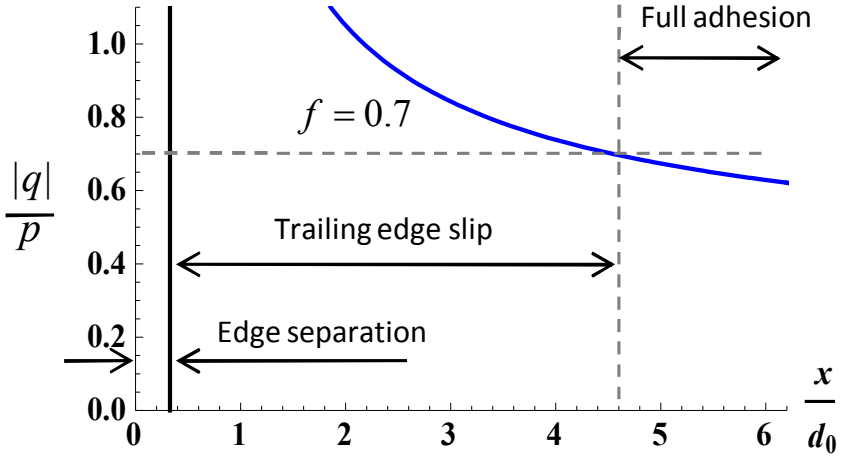
$$\frac{x_s}{d_0} = \left(\frac{f^{II}_{r\theta}(\theta_{int}) - f \cdot f^{II}_{\theta\theta}(\theta_{int})}{f^{I}_{r\theta}(\theta_{int}) - f \cdot f^{I}_{\theta\theta}(\theta_{int})} \right)^{\frac{1}{\lambda_I - \lambda_{II}}} \tag{5.6}$$

by using (5.5) and (5.1), where slip is implied to occur for $x < x_s$. Notice that the slip extent in (5.6) is normalized by d_0, so this result is given entirely within the two-term Williams asymptote. Moreover, it is independent of the finite geometry of the problem to which it is applied.

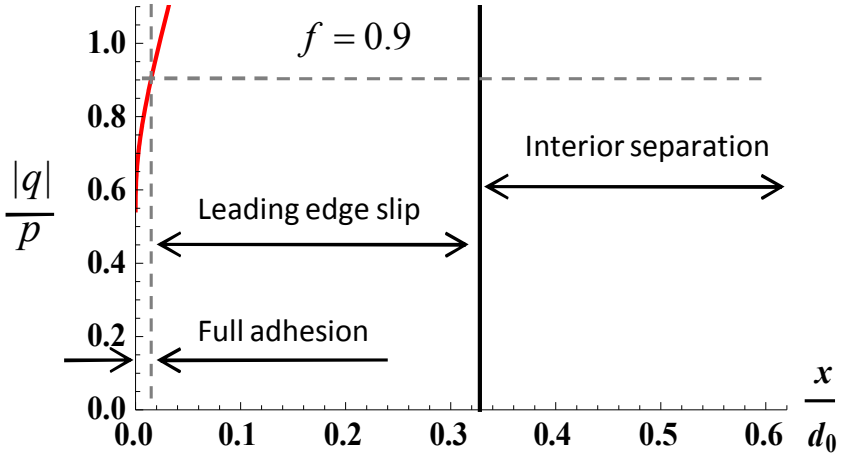
Note that if $f < 0.219$, there is implied to be leading-edge slip, stick, and then trailing-edge slip as the observation point moves in from the contact edge. The latter stick/slip transition is also given by (5.6), but with f replaced by $-f$. In practice, it is very unlikely that this slip-stick transition will actually occur because it is not implied until $r \gg d_0$ (as can be seen in Figure 5.2(a)) at which point the asymptote is certainly not relevant. However, when the contact angle, ϕ, is reduced, this more complex pattern of behaviour may become important, and this is briefly examined in §5.6.



(a) Case (a): $K_I < 0 < K_{II}$



(b) Case (b): $0 < K_I, K_{II}$



(c) Case (c): $K_I, K_{II} < 0$

Figure 5.2: Plots of the magnitude of the traction ratio, $|q(x)|/p(x)$, shown against position within the asymptotic solution, x/d_0, for $\phi = 90°$.

5.3.2 Case (b): $0 < K_I, K_{II}$

In this case, the near edge (mode I) eigensolution implies a tensile normal traction, while the remote (mode II) eigensolution implies closure. Thus, case (b) implies that there will be edge separation, which will be followed by an adjacent slip zone and possibly a stick zone further towards the interior of the contact. Since both generalized stress intensity factors are positive in this case, the standard definition of d_0, G_0 given in (2.26) can be used here.

The first step in restoring contact conditions is to determine the extent of separation at the contact edge, x_o. For the interface to be in contact, the contact pressure must be compressive, so the condition

$$p(x) > 0 \tag{5.7}$$

must be satisfied. In this case, the contact pressure is given by

$$p(x) = \left(\frac{x}{d_0}\right)^{\lambda_I - 1} f^I_{\theta\theta}(\theta_{int}) + \left(\frac{x}{d_0}\right)^{\lambda_{II} - 1} f^{II}_{\theta\theta}(\theta_{int}), \tag{5.8}$$

so an estimate of the extent of separation can be obtained in closed form based on violations of (5.7), which gives

$$\frac{x_o}{d_0} = \left(-\frac{f^{II}_{\theta\theta}(\theta_{int})}{f^I_{\theta\theta}(\theta_{int})}\right)^{\frac{1}{\lambda_I - \lambda_{II}}}, \tag{5.9}$$

where separation occurs for $x < x_o$.

Thus, closure is implied for $x > x_o$, and the traction ratio in this region is implied to be

$$\frac{q(x)}{p(x)} = -\left(\frac{\left(\frac{x}{d_0}\right)^{\lambda_I - 1} f^I_{r\theta}(\theta_{int}) + \left(\frac{x}{d_0}\right)^{\lambda_{II} - 1} f^{II}_{r\theta}(\theta_{int})}{\left(\frac{x}{d_0}\right)^{\lambda_I - 1} f^I_{\theta\theta}(\theta_{int}) + \left(\frac{x}{d_0}\right)^{\lambda_{II} - 1} f^{II}_{\theta\theta}(\theta_{int})}\right). \tag{5.10}$$

This traction ratio is plotted against position within the asymptote in Figure 5.2(b). In this case, the shear traction is *trailing-edge* ($q(x) > 0$) in character for both the mode I and the mode II eigensolution. Dashed lines are also shown in this figure to illustrate the

implied extent of violations of the Coulomb friction condition but this time for $f = 0.7$.

As before, an estimate of the extent of slip can be obtained in closed form and is given by

$$\frac{x_s}{d_0} = \left(-\frac{f_{r\theta}^{II}(\theta_{int}) + f \cdot f_{\theta\theta}^{II}(\theta_{int})}{f_{r\theta}^{I}(\theta_{int}) + f \cdot f_{\theta\theta}^{I}(\theta_{int})} \right)^{\frac{1}{\lambda_I - \lambda_{II}}}, \tag{5.11}$$

where slip occurs in the range $x_o < x < x_s$. This expression is valid provided only that $f > |g_{r\theta}^{II}| = 0.219$. If the coefficient of friction is lower than this, slip extends throughout the region, i.e. $x_s/d_0 \to \infty$.

5.3.3 Case (c): $K_I, K_{II} < 0$

This case is more difficult to interpret, and it is less informative. In contrast to the previous cases in which violations of the Signorini conditions originated from the *edge* of contact, and contact and stick conditions were gradually restored as the observation point moved *away* from the contact edge, here precisely the reverse occurs. Violations of the Signorini conditions originate *remote* from the contact edge, and contact and stick conditions are gradually restored as the observation point *approaches* the contact edge. Of course, deductions made about violations originating from the interior of the contact must be viewed with caution since the influence of the asymptotic solution decreases as the distance from the contact edge increases.

Nevertheless, the near edge (mode I) solution now implies closure, while the remote (mode II) solution implies separation. Although interior separation is implied *within* Williams' solution itself, separation will not be implied once the asymptote is applied to a finite problem if $d_0 \gg a$, and this is examined further in §5.4 and §5.5. Instead, the near edge (mode I) solution will dominate the state of stress and will imply closure. However, when d_0/a is not large, this solution cannot be used because higher-order terms that have been neglected are likely to be important.

To estimate the extent of slip and separation for this case, d_0, G_0 must be redefined

since K_{II} is negative. As before, a new parameter can be defined as

$$\tilde{K}_{II} = -K_{II}, \tag{5.12}$$

so d_0, G_0 can be redefined as

$$d_0 = \left(\frac{\tilde{K}_{II}}{K_I}\right)^{\frac{1}{\lambda_I - \lambda_{II}}} \tag{5.13a}$$

$$G_0 = K_I^{\frac{\lambda_{II}-1}{\lambda_{II}-\lambda_I}} \tilde{K}_{II}^{\frac{\lambda_I - 1}{\lambda_I - \lambda_{II}}}, \tag{5.13b}$$

and (2.27) can be rewritten as

$$\frac{\sigma_{ij}(r,\theta)}{G_0} = \left(\frac{r}{d_0}\right)^{\lambda_I - 1} f_{ij}^{I}(\theta) - \left(\frac{r}{d_0}\right)^{\lambda_{II}-1} f_{ij}^{II}(\theta). \tag{5.14}$$

In this case, separation is implied to initiate from the interior of the contact, so the first step towards restoring contact conditions is to estimate the point of closure, x_o. This is done as in the previous section, and the separation extent is again given by (5.9) but with d_0 defined as in (5.13a). Separation now occurs in the region $x > x_o$. Thus, closure is implied for $x < x_o$, and the traction ratio in this region is given by (5.10) (but again with d_0 defined as in (5.13a)).

This traction ratio is plotted in Figure 5.2(c), which uses a considerably magnified horizontal axis because separation is implied for $x/d_0 > 0.335$. This figure also includes dashed lines to illustrate the slip extent implied by violations of the Coulomb friction condition for $f = 0.9$. Note that there may be a region of interior slip (depending on the value of d_0, see §5.5), which will be *leading-edge* in character. However, this slip zone will not extend to the contact edge provided that $f > g_{r\theta}^{I} = 0.543$. Here, an estimate of the stick-slip transition point, x_s, is given by

$$\frac{x_s}{d_0} = \left(-\frac{f_{r\theta}^{II}(\theta_{int}) - f \cdot f_{\theta\theta}^{II}(\theta_{int})}{f_{r\theta}^{I}(\theta_{int}) - f \cdot f_{\theta\theta}^{I}(\theta_{int})}\right)^{\frac{1}{\lambda_I - \lambda_{II}}}, \tag{5.15}$$

where slip occurs for $x_s < x < x_o$, and stick is implied for $x < x_s$ (i.e. at the contact edge).

5.4 Implications of the internal length dimension

A key feature of all the results obtained in the previous section is that the *sizes* of the zones of stick, slip, and separation are normalized by d_0. Moreover, these results are completely general in the sense that they apply to any complete contact with an edge angle of $\phi = 90°$ that is subjected to mixed-mode loading. Furthermore, the magnitude of the remote loads, G_0, does not enter into any of the expressions for the extent of stick, slip, or separation. Thus, the pattern of stick, slip, or separation that is implied *within the semi-infinite solution* is solely determined by the signs of K_I, K_{II}, while the sizes of these zones *within the finite problem* to which it is applied are determined by the mode-mixity, which is quantified by d_0.

For this reason, it is instructive to consider the deductions that can be made about contact behaviour simply by comparing the size of d_0 to a characteristic length dimension, a, of the finite problem to which this asymptote is applied. For this purpose, several lines of constant d_0/a are shown in Figure 5.1(b), which illustrates the variation of d_0 during a loading trajectory in generalized-stress-intensity-factor space, such as that shown in Figure 5.1(a). Note that the standard definition of d_0 given in (2.26b) is used in this figure, which includes modulus signs.

Figure 5.1(b) shows that when p_0 is first applied, d_0/a is fixed at 0.219. This is because the mode-mixity associated with any single applied load is fixed, which is clear from the calibrations given in (4.12). In contrast, once P is also applied and varied over the range $-0.05 < P/ap_0 < 0.05$, d_0 varies significantly, and this is due to the resulting variation in mode-mixity that is associated with this combination of loads.

Now consider the point in Figure 5.1(a) when $P/ap_0 = -0.05$, and the cantilever tip is pushed as far upwards as occurs in the loading trajectory. Figure 5.1(b) shows that $d_0/a = 35.6$ at this point; thus, d_0 is much larger than the contact half-width, a. An

immediate implication of this is that the contact edge is effectively under pure mode I loading. This is because the remote (mode II) solution is not relevant until $r \gg d_0$, which (if $d_0 \gg a$) does not occur until $r \gg a$, and this is well outside the region where Williams' solution is relevant.

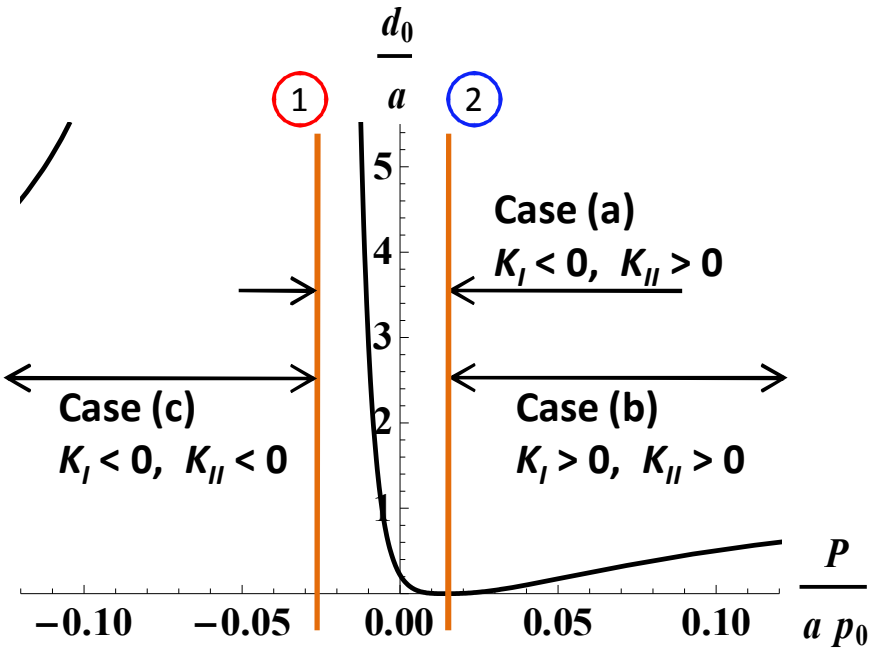
As the magnitude of P is reduced, d_0 continues to increase until $P/ap_0 = -0.0294$, which corresponds to point (1) of Figure 5.1(a). Here, $d_0 \to \infty$, the loading is purely mode I in character, and thus only the near edge (mode I) solution applies. As P is varied further in this direction, d_0 progressively decreases until point (2) is reached when $P/ap_0 = 0.0143$. At this point, $d_0 \to 0$, the loading is purely mode II in character, and thus only the remote (mode II) solution applies. A further increase in P causes d_0 to increase again, and at the end of this interval, $d_0/a = 0.174$.

5.5 Application to clamped-cantilever test rig

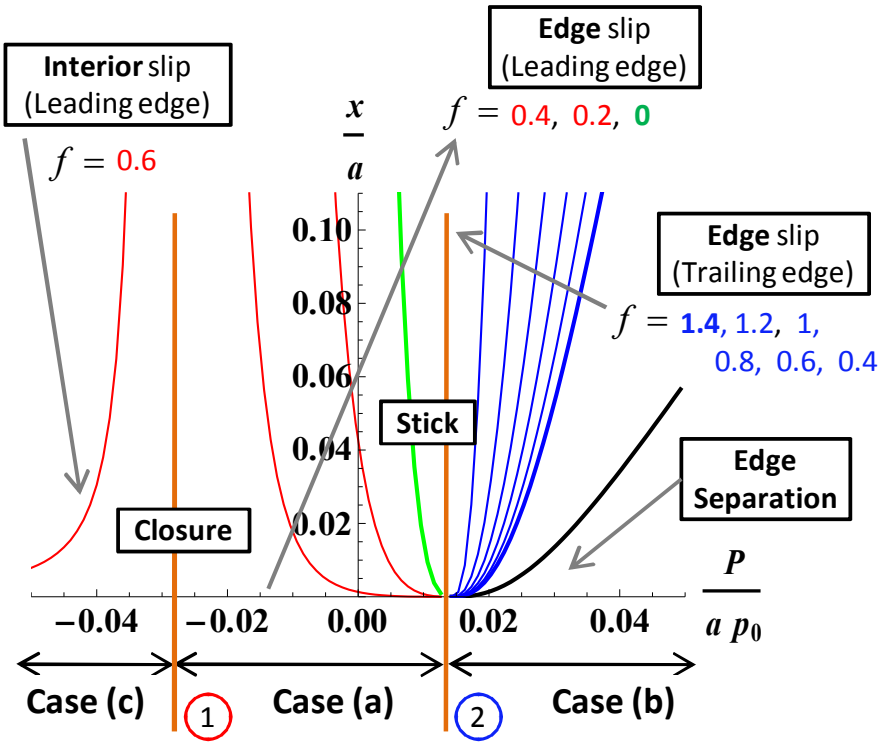
The results presented up to this point, which are given in terms of x/d_0, are very general and apply to any contact with an edge angle of $\phi = 90°$. However, the practical implications of these general results can be difficult to interpret, so it is helpful to examine their implications for the clamped-cantilever test rig shown in Figure 4.2. To do this, the results in Figure 5.2 must simply be scaled by d_0/a. As noted above, this quantity varies with the load combination, and it is plotted over the range $-0.12 < P/(ap_0) < 0.12$ in Figure 5.3(a). The implied extent of slip and (where applicable) separation is then plotted in terms of x/a within the range $-0.05 < P/(ap_0) < 0.05$ in Figure 5.3(b). Note that all analysis is with respect to the upper corner.

Figure 5.3(b) illustrates that case (a) results from the application of p_0 (i.e. when $P/(ap_0) = 0$), so the contact edge is closed. If $f > g^I_{r\theta} = 0.543$, the contact will be stuck throughout (as in Figure 4.1(a)); otherwise, *leading-edge* slip is implied at the contact edge (as in Figure 4.1(b)). The extent of slip will nevertheless be very small unless the coefficient of friction is quite low, i.e. $f \lesssim 0.4$.

If the cantilever tip is pushed downwards (i.e. $P > 0$), initially the extent of slip (if



(a) Variation of d_0/a with the ratio of the applied loads, $P/(ap_0)$.



(b) Implications for contact behaviour in terms of the zones of stick, slip, and separation at the contact edge.

Figure 5.3: Application of Williams' solution to the clamped-cantilever test rig.

present) is implied to decrease monotonically. It follows that because the slip increment now opposes the sign of the shear traction, stick will occur everywhere. Notice that this is always the case, regardless of the coefficient of friction.

The behaviour described above applies until point (2) of Figure 5.3 is reached when $P/(ap_0) = 0.0143$. At this point, the loading is purely mode II in character, $d_0 \to 0$, and the edge response changes from case (a) to case (b). Therefore, increasing P past this point produces a qualitative change in behaviour because the cantilever root opens. The response then becomes that shown in Figure 4.1(c), such that an observation point moving in from the edge sees: opening, *trailing-edge* slip, and then stick, provided only that $f > |g_{r\theta}^{II}| = 0.219$. Figure 5.3(b) shows that increasing P further causes these regions of slip and separation to extend deeper into the contact until eventually they extend into a region where the asymptote has no further relevance.

Now consider the case when P is applied upwards (i.e. $P < 0$) after the initial application of p_0. In this case, if $f < g_{r\theta}^{I}$, then as the tip load is increased in magnitude, the extent of *leading-edge* slip will increase monotonically until it fills the whole asymptotic region. Conversely, if $f > g_{r\theta}^{I}$, there will be full stick throughout the contact-edge region. As P is applied further in this direction, the edge moves towards point (1), which is reached when $P/(ap_0) = -0.0294$. At this point, $d_0 \to \infty$, and the cantilever root is subjected to pure mode I loading. As point (1) is passed, the response changes from case (a) to case (c). However, note that d_0/a is still very large, which is illustrated in the top left hand corner of Figure 5.3(a). Thus, although the cantilever root response is case (c), the interior separation and slip that are implied by this case do not actually occur. Instead, the edge remains firmly closed and dominated by a mode I field, and there is no immediate qualitative change in contact behaviour.

5.5.1 Other considerations

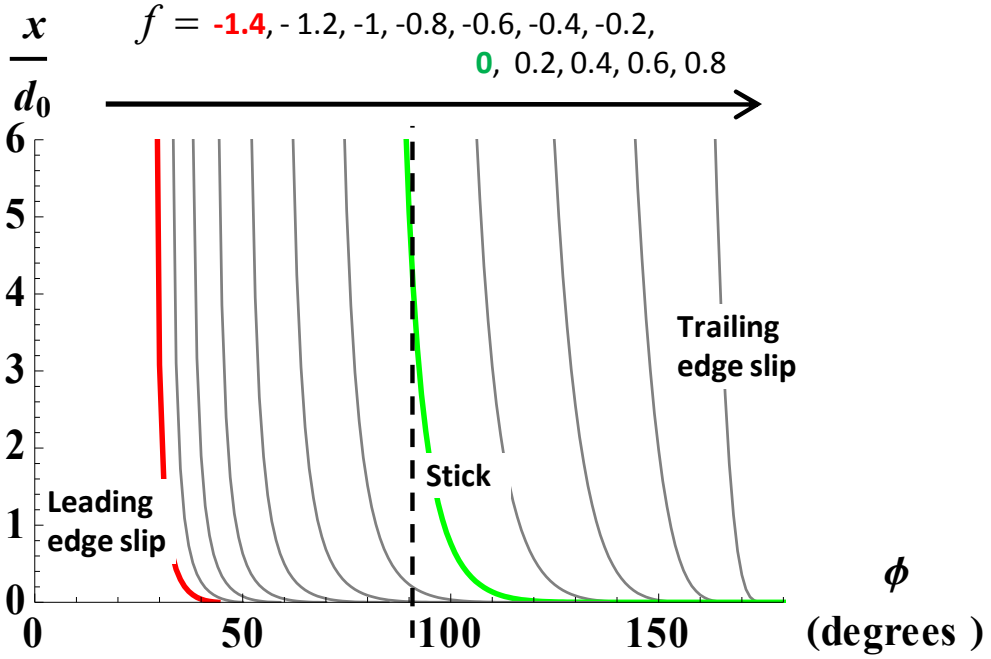
It is worth making a few remarks on the accuracy of the results presented above. First, although the clamped-cantilever test rig that is examined above is representative of

common conditions in fatigue experiments, many general inferences should not be made based on the results shown in Figure 5.3. These results are based on calibrations for the particular case when $L/a = 40$ and are specific to the test rig that was used by Juoksukangas et al. [2013]. In addition, the calibrations can sometimes be very sensitive to minor variations in the boundary conditions, and varying L/a would also have a highly non-linear effect. Therefore, when applying this method in practice, calibrations should be calculated for the specific problem under consideration to make sure they represent the complex stress state at the contact edge as accurately as possible. Of course, changes to the calibrations will not modify the solution within the asymptote (i.e. Figure 5.2 and all of the results in §5.3 will remain identical for any contact of $\phi = 90°$), but they will change the implications for the finite problem (i.e. Figure 5.3).

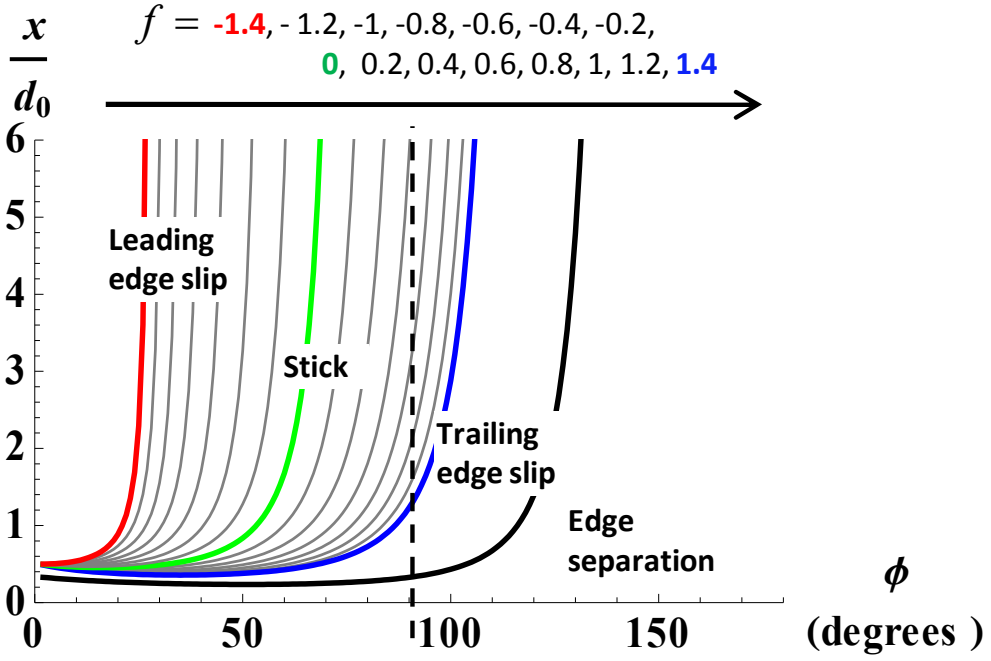
Another thing to consider is: how accurate are the calibrations at, say, the upper corner when slip or separation is implied to occur at the lower corner? The most rigorous answer is that if the stress state at the contact edge changes, so too do the calibrations. However, recall that this approach is only useful when any slip and/or separation that occurs remains near to the contact edge and within the region where the asymptote dominates behaviour, i.e. $r \ll a$; otherwise, it should not be applied. Thus, as this contact-edge region is necessarily small, small amounts of slip or separation at, say, the upper contact edge should have minimal influence on the results at the lower contact edge.

5.6 Extension to other contact angles

The results presented in §5.3 for $\phi = 90°$ can now easily be extended to the full range of punch angles, i.e. $0° < \phi < 180°$. The only modification that must be made when generalizing these results is that the $+$ or $-$ sign in (5.6) and (5.11) must be replaced with a $\pm$ sign since the slip direction does not necessarily remain the same as the contact angle is varied. Note that $+f$ corresponds to trailing-edge slip, and $-f$ corresponds to leading-edge slip. Once (5.6) and (5.11) are modified as described above, they should be understood as giving the distances from the contact edge at which a transition from slip to



(a) Case (a): $K_I < 0 < K_{II}$



(b) Case (b): $0 < K_I, K_{II}$

Figure 5.4: Implied extent of slip and/or separation plotted against contact angle, ϕ. Slip-stick transitions are shown for $-1.4 < f < 1.4$ at intervals of 0.2 along with the transition from edge separation to trailing-edge slip for case (b).

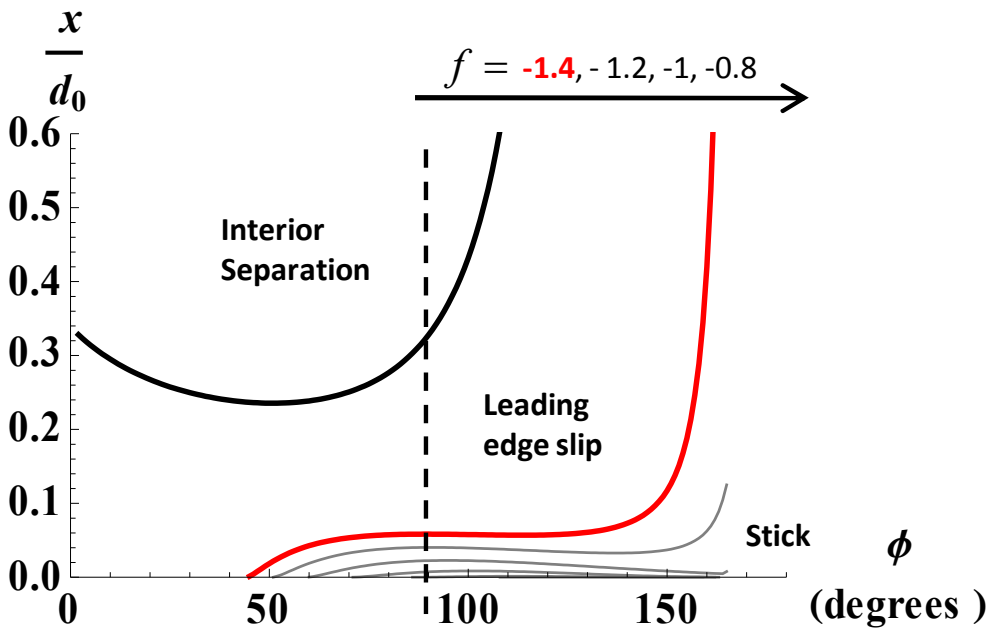


Figure 5.5: Implied extent of slip and/or separation plotted against contact angle, ϕ for case (c) when $K_I, K_{II} < 0$. Slip-stick transitions are shown for $-1.4 < f < -0.8$ at intervals of 0.2 along with the transition from leading-edge slip to interior separation.

stick is implied to occur. Thus, care must be taken to ensure that the correct implications are inferred. In addition, when the results suggest complex patterns of behaviour, e.g. case (b) when $\phi \lesssim 50°$, the results should be viewed with caution since higher-order terms are likely to be important in these cases.

The transitions in contact behaviour implied by (5.6), (5.9), and (5.11) are plotted against contact angle, ϕ, for cases (a) and (b) in Figure 5.4 and for case (c) in Figure 5.5. Note that, as before, the behaviour for case (c) is shown on a considerably magnified scale since the solution is not reliable once significant interior separation is implied. To interpret these figures, it is helpful to begin by considering the implications for a particular angle. For instance, the case when $\phi = 90°$, which is examined in the previous sections, corresponds to the vertical dashed lines in Figures 5.4-5.5. The intersection of these vertical lines with the curves in Figures 5.4-5.5 correspond to the locations at which the horizontal dashed lines intersect the traction ratios that are plotted in Figure 5.2.

The main purpose of Figures 5.4-5.5 is simply to illustrate the ease with which these

results can be generalized to other contact angles. It is emphasized that when using these calculations in practice, it is best to create plots such as Figure 5.2 for the relevant contact angle, to examine the implications of these plots, and then to apply these results to the finite problem under consideration as is done in Figure 5.3. However, although Figures 5.4-5.5 are difficult to interpret, a few remarks can be made. First, these figures illustrate that while the extent of slip and/or separation generally varies smoothly with the contact angle, the rate of change of the slip extent can be very steep. In addition, an exception to this smooth variation occurs when the mode II eigensolution changes from being singular to being bounded when $\phi \approx 77.4°$. At this angle, there is a qualitative change in the slip-stick pattern that is implied within the Williams asymptote because the shear traction associated with the mode II eigensolution changes sign.

5.7 Discussion

The purpose of the analysis presented in this chapter is to provide engineers with a useful technique for designing and analysing complete contacts subject to fretting conditions for which the load amplitude is relatively low (i.e. not fretting-wear or macro-slip conditions). Unlike tests on incomplete contacts subjected to similar loading regimes for which edge slip invariably occurs, it can be unclear whether a complete contact test configuration will actually result in partial-slip conditions or if the contact will be fully stuck throughout the load cycle. Although it is not yet clear which of these circumstances, i.e. partial-slip vs. full stick, is better for the fatigue performance of complete contacts, the details of the slip/stick behaviour at the contact edges are likely to play an important role.

The primary, and most reliable, output from the analysis presented in §4-§5 is information about the behaviour at the contact edges; specifically, which of the three states shown in Figure 4.1 is implied to occur, viz. closed and stuck, closed and slipping, or open and slipping. Of course, in practical loading regimes (e.g. Figure 5.1(a)), it is common for the contact edges to alternate between states, and the calculations presented here provide accurate information on if (and when) these transitions will occur. Indeed, this information

is obtained much more accurately than would be obtained using conventional numerical methods, which commonly suffer from convergence issues due to the very steep stress gradients near the contact edges.

It is noted that all of the calculations presented here are based on the assumption of a bilateral contact interface, and the extent of slip or separation is estimated based on violations of the Signorini conditions. This approach is useful not only because of its mathematical simplicity and ease of application in practice but also because it is usually more important to know with confidence what qualitative behaviour is implied at the contact edges than it is to quantify the extent of slip or separation with great precision. This is because this information enables tests to be designed that excite specific types of contact-edge behaviour, so the isolated influence of each of these on fatigue performance can be compared. In addition, in the context of fatigue experiments, it is very likely that some *frictional shakedown* will occur, which will reduce the extent of slip during the course of the loading history (see e.g. [Ahn and Barber, 2008, Ahn et al., 2008, Barber et al., 2008, Churchman and Hills, 2006b, Klarbring et al., 2007]).

As the results for the true extent of slip and/or separation have already been presented by Churchman and Hills [2006c] and Paynter et al. [2010] for the $\phi = 90^\circ$ case, the accuracy of the closed-form estimate used here can be assessed. For the closed-and-slipping case, Churchman and Hills [2006c] showed that the true extent of slip is simply 2.4 times greater than that implied by a calculation based on violations. Interestingly, these authors found that this result is independent of the coefficient of friction. However, this result will vary with ϕ and must be re-computed for each new contact angle.

The true extent of slip and separation for the open-and-slipping case was obtained by Paynter et al. [2010]. However, it is more complex as the relative size of the zones of slip and separation depend on the coefficient of friction, f. Specifically, these authors found that as f is increased, the slip zone decreases in extent, while the opening region increases in extent and vice versa. Thus, a direct comparison of these results is not pursued here, but it may be examined in future work.

5.8 Summary

In this chapter, an estimate of the extent of slip and/or separation at the edge of a contact complete is obtained in closed form. When slip and/or separation is implied to occur *near* the contact edge, these expressions provide an order of magnitude estimate of its extent. Conversely, only qualitative implications can be derived when larger amounts of slip and/or separation are implied. These results are then applied to the clamped-cantilever test rig studied by Juoksukangas et al. [2013], and their practical implications are discussed. It is then shown that these results can easily be extended to any contact angle of interest. Finally, the applications of these calculations are discussed, including designing fretting experiments that excite specific contact-edge responses in order to compare their effect on fatigue performance.

Chapter 6

Asymptotic analysis of fretting-fatigue experiments

In this chapter, it is shown that the techniques presented in §3 can be applied to complete contacts under some conditions. The tools developed in §2-§5 are then applied to analyse several groups of fretting-fatigue tests that were performed on complete contacts by Juoksukangas et al. [2013], Mugadu [2002], and Noraphaiphipaksa et al. [2013]. The results suggest that the tests performed by Juoksukangas et al. [2013] are amenable to an asymptotic analysis because only a moderate amount of slip and separation is likely to have occurred in these tests. In contrast, the tests performed by Mugadu [2002] and Noraphaiphipaksa et al. [2013] are not as easily analysed in this way because a significant amount of slip and separation occurred in these tests.

Note that parts of this chapter have been published in [Flicek et al., 2013] and [Hills and Flicek, 2015].

6.1 Mode-mixity and small-scale yielding of fretting tests

In §3, several calculations are presented for assessing the mode-mixity and small-scale yielding of the process zone that arises at the tip of sharp V-notches. In this section, the conditions under which these calculations can be applied to complete contacts is examined. To do this, recall the implications of the four possible sign combinations of the generalized stress intensity factors:

1. Case (a): $K_I < 0 < K_{II}$: Both eigensolutions imply closure.

2. Case (b): $0 < K_I, K_{II}$: The mode I term implies edge separation, and the mode II term implies interior closure.

3. Case (c): $K_I, K_{II} < 0$: The mode I term implies edge closure, and the mode II term implies interior separation.

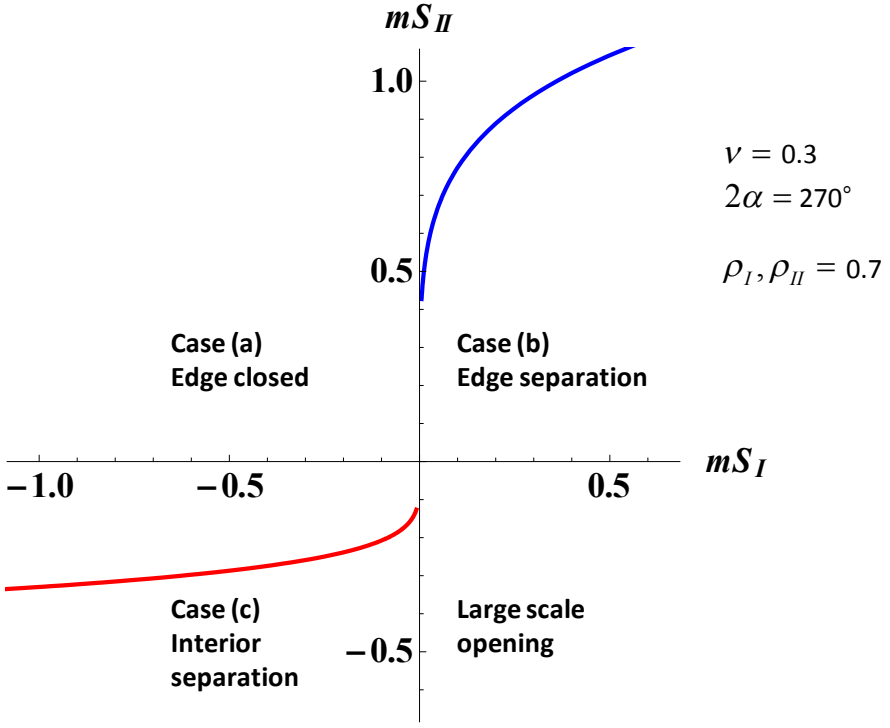
4. $K_{II} < 0 < K_I$: Both eigensolutions imply separation.

The region of generalized-stress-intensity-factor space corresponding to each of these four possible contact-edge responses is shown in Figure 6.1(a). Note that since both eigensolutions imply opening when $K_{II} < 0 < K_I$, Williams' solution cannot be used, so this case is not considered any further in the following analysis.

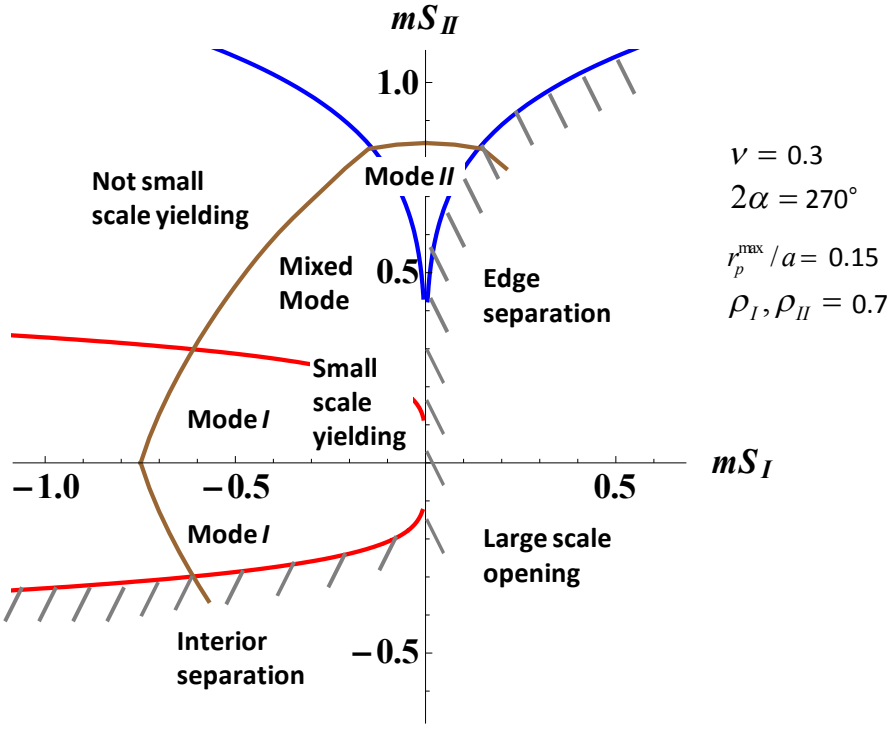
6.1.1 Range of applicability

To examine when it is appropriate to employ the calculations from §3 to complete contacts, consider case (a), which is shown in Figure 6.1(a). For this case, both K_I and K_{II} imply closure. Hence, if the friction coefficient is sufficient to prevent edge slip (i.e. $f > g^I_{r\theta}$), the contact edge will be fully stuck (see §4.3). Since the displacement state across the contact interface is continuous under these conditions, the contact edge will behave like a sharp V-notch. Thus, the calculations from §3 can be applied to the case (a) region of Figure 6.1(a) without modification provided only that $f > g^I_{r\theta}$.

If the friction coefficient is not sufficient to prevent edge slip, a slip zone will form at the contact edge, and its extent can be estimated using the techniques described in §5.3.1. If the extent of edge slip is implied to be minuscule based on this calculation, such that the process zone is likely to be larger than the implied slip extent, the analysis from §3 can still be used. This is because Williams' solution will still dominate the stress state in the vicinity of the contact edge. However, when larger slip lengths are implied, this analysis should not be applied because Karuppanan and Hills [2008b] have shown that the stress state at the contact edge is significantly modified by the presence of frictional slip.



(a) Load range in which the calculations from §3 can be applied to complete contacts.



(b) Mode-mixity and small-scale-yielding information of the process zone at the edge of a complete contact.

Figure 6.1: Application of the calculations from §3 to complete contacts.

For case (b), K_I implies opening, and K_{II} implies closure, so separation and trailing-edge slip are implied to occur at the contact edge. Clearly, if the extent of slip and separation is implied to be moderate, the process zone will be of an entirely different character from that arising at the tip of a sharp V-notch under similar loading conditions, so the calculations from §3 are not appropriate. Conversely, if the extent of slip and separation is implied to be very small based on the calculations in §5.3.2, such that the process zone is likely to be larger than the separated region, then Williams' solution will still dominate the contact-edge stress state. An estimate of when the process zone will be larger than the separation zone can be obtained using the analysis from §3: namely, when the process zone is dominated by the mode II eigensolution (which implies closure), the separation zone will be negligibly small.

The boundary between a mode II dominated process zone and a mixed-mode process zone for case (b) is shown in Figure 6.1(a) for $2\alpha = 270°$ (assuming that $\nu = 0.3$ and that a correlation coefficient of $\rho_{II} = 0.7$ marks this boundary). Note that this curve has been obtained using the approach described in §3.2. Therefore, for loading to the left of this curve, the response is likely to remain similar to that of case (a). In contrast, for loading to the right of this curve, the extent of slip and separation is implied to be moderate or large, so the analysis from §3 should not be applied.

For case (c), K_{II} implies opening, and K_I implies closure. However, recall from §5.5 that when the contact-edge response is case (c), Williams' solution can only be applied when the loading is mainly mode I in character. Thus, following a similar argument to that for case (b) given above, the contact edge will certainly remain closed when the bilateral analysis from §3 implies that the process zone is dominated by the mode I eigensolution (which implies closure). The boundary between a mode I dominated process zone and a mixed-mode process zone for case (c) is shown in Figure 6.1(a) for $2\alpha = 270°$ (assuming that $\nu = 0.3$ and that a correlation coefficient of $\rho_I = 0.7$ marks this boundary).

Thus, the analysis from §3 can be applied to complete contacts if: (i) the values of mS_I and mS_{II} resulting from the applied loads lie to the left side of the curves shown in Fig-

ure 6.1(a), and (ii) the coefficient of friction is sufficient to prevent edge slip. Figure 6.1(b) includes additional curves in this region of the plot, which have also been obtained using the techniques presented in §3. This figure displays the boundaries that separate mainly mode I, mainly mode II, and mixed-mode process zones for $2\alpha = 270°$ (assuming that $\nu = 0.3$ and that $\rho_I, \rho_{II} = 0.7$ mark these boundaries). The small-scale-yielding limit of $r_p^{\max}/a = 0.15$ is also included on this plot. In addition, Figure 6.1(b) is annotated to show the regions where separation is implied to occur, so Williams' solution should not be applied.[1]

6.2 Juoksukangas' fretting tests

As the clamped-cantilever test rig (see Figure 4.2) developed by Juoksukangas et al. [2013] is used as a working example to demonstrate the calculations presented in §4-§5, the techniques presented in these chapters have already been applied to it. However, one aspect of analysis that has been been neglected thus far is the assessment of the mode-mixity and small-scale yielding of the process zone at the contact edges during these tests. Indeed, for part of the analysis in §4.4-§4.5, it is simply assumed that the variation of the mode I generalized stress intensity factor, ΔK_I, controls crack nucleation, and the fretting data are then shown in Figure 4.6 on a plot of ΔK_I vs. cycles to macroscopic crack initiation, N_1.

6.2.1 Mode-mixity and small-scale yielding

Recall from §3 that assuming ΔK_I controls crack nucleation amounts to assuming: (i) the process zone is within the limits of small-scale yielding (i.e. $r_p \ll a$), and (ii) the process zone is dominated by the mode I eigensolution (i.e. $r_p \ll d_0$). These two assumptions can be examined by plotting the generalized stress intensity factors arising during the fretting tests on Figure 6.1(b). This is done in Figure 6.2, which includes

[1] Note that although interior separation may not actually occur even when the response is to the right of the curve in the case (c) region of this plot, the analysis presented here still should not be applied because higher-order terms in Williams' series expansion will become important.

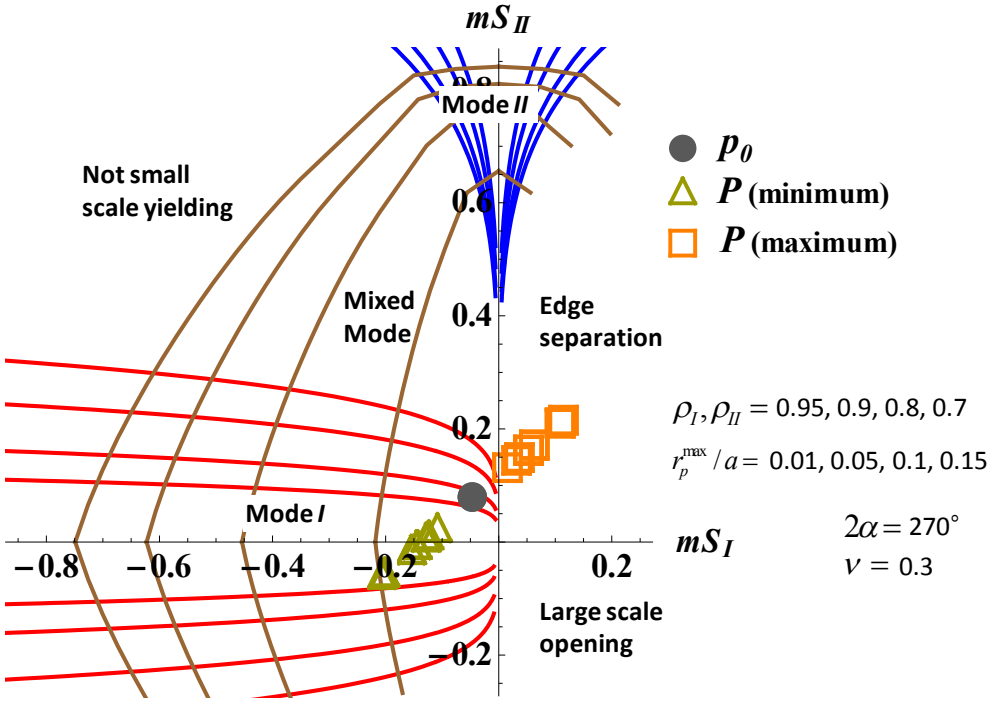


Figure 6.2: Plot of mS_I, mS_{II} at the extreme points of the load cycle for the fretting tests reported by Juoksukangas et al. [2013]. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown for $2\alpha = 270°$ (all calculated assuming $\nu = 0.3$).

curves of $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ for characterizing the mode-mixity of the process zone and also includes small-scale-yielding limits of $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ (see §3 for more details). As Poisson's ratio, ν, for the 'quenched and tempered 34CrNiMo6 steel' that was used in these tests was not reported by Juoksukangas et al. [2013], a value of $\nu = 0.3$ has been assumed for calculating the mode-mixity and small-scale-yielding information in Figure 6.2.[2] Also note that all analysis is for the top corner (as in §4-§5).

The friction coefficient, f, during these experiments was not reported by Juoksukangas et al. [2013], but it is assumed that $f > g_{r\theta}^I = 0.543$ for the following analysis. Thus, the contact edge will remain stuck as long as the loading trajectory remains to the left of the curves shown in Figure 6.1(a). In Figure 6.2, the values of mS_I, mS_{II} arising when the clamping load of $p_0 = 100MPa$ is first applied are displayed with a circle. This shows that the application of p_0 results in full stick and is well within the 1% small-scale-

[2] In any case, Figure 3.9 shows that the results are very insensitive to variations in Poisson's ratio.

yielding limit included in this figure. Furthermore, this figure shows that the correlation coefficient between the process zone arising due to p_0 and a pure mode I process zone (i.e. ρ_I) is between 90% and 95%.

The cantilever tip loads that were applied during the fretting tests were obtained from the data reported in [Juoksukangas et al., 2013], and more details on how these data were processed is given in §A.2.1. The extreme values of mS_I, mS_{II} (i.e. when P takes on its maximum or minimum value during the loading regime shown in Figure 5.1(a)) are plotted in Figure 6.2. This figure shows that when P is at its minimum value during the tests (i.e. when P is applied upwards), the process zone at the upper corner is even more strongly mode I dominated in all cases, such that $\rho_I > 0.95$. In addition, the process zone is within the 1% small-scale-yielding limit for all of these tests. Note that when P is at its minimum value, the upper corner response is in the case (a) region of the diagram for some of the tests, but it extends into the case (c) region for others. However, as the process zone remains heavily mode I dominated, interior separation will not occur, and the edge will remain closed and stuck (see §5.5 for more details).

On the other hand, the results presented in §4.3.1 show that edge separation is implied to have occurred in all the fretting tests that were conducted. This is also evident in Figure 6.2, which shows that all the tests result in a case (b) response when P is at its maximum value. Note that the calibrations of the generalized stress intensity factors have been obtained from a bilateral model of the contact, which does not account for the influence of slip or separation. Thus, as separation is implied for case (b), the points in Figure 6.2 corresponding to when P is at its maximum value should not be taken as a measure of the stress state at the contact edge. Instead, these points should only be interpreted as implying that edge separation will occur.

6.2.2 Implications for fretting performance

These results show that the top corner is subjected to a mode I dominated stress state for approximately half of the load cycle (i.e. when P is applied upwards), but during

the other half of the cycle (i.e. when P is applied downwards), the corner lifts off and slips (as in Figure 4.1(c)). Thus, it is not clear from this figure whether ΔK_I will control crack nucleation because this depends on which type of edge condition results in the most damage to the specimen. There are three possibilities:

1. Most fretting damage occurs when P is applied upwards due to the stress intensification associated with the sharp corner at the contact edges.

2. Most fretting damage occurs when P is applied downwards due to the partial-slip conditions at the contact edges.

3. Both partial-slip conditions and closed-and-stuck conditions result in similar levels of damage.

If hypothesis 1 is correct, ΔK_I can be expected to control fatigue life. Alternatively, if hypothesis 2 or 3 is correct, other factors are likely to be more (or at least equally) important.

Of course, it is not possible to discriminate between these hypotheses based on the single set of tests reported by Juoksukangas et al. [2013]. To determine which of these hypotheses is correct, more experiments would need to be carried out with different loading regimes and test geometries (but maintaining the same material and contact angle of $\phi = 90°$). All these results could then be plotted together on Figure 4.6 (i.e. on a plot of ΔK_I vs. cycles to macroscopic crack initiation, N_1). If ΔK_I does in fact control crack nucleation, the fatigue limit obtained in §4.4 should hold for these subsequent tests, and all the results should fall approximately on the same curve on this plot.

Another important issue that could be investigated is whether partial-slip conditions or closed-and-stuck conditions are more detrimental for fretting performance. One way to investigate this issue would be to use Figure 6.2 and the other calculations presented in §4-§5 to design several groups of experiments that excite distinct contact-edge responses. For example, one set of tests could be designed to excite closed-and-stuck contact edges throughout the load cycle, and a separate group could excite open-and-slipping contact

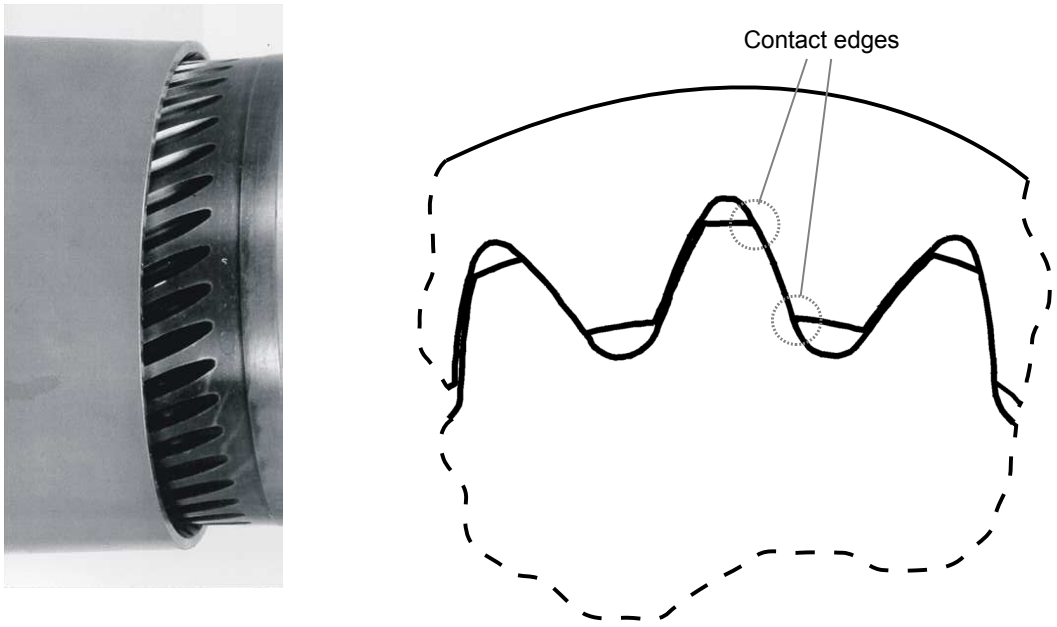
edges. The fretting performance of these two sets of tests could then be compared, and this would reveal which of these contact-edge states results in more severe damage.

Alternatively, this could be investigated with a single set of tests using the same clamped-cantilever rig shown in Figure 4.2 but with a modified loading regime. Specifically, instead of applying P in a fully reversing sense, i.e. from $-P$ to $+P$, the load could be applied from zero to $+P$ (i.e. downwards). Doing this would cause the stress state at the *top corner* to alternate between the circle (when only p_0 is applied) and the squares in the case (b) region of Figure 6.2. Because of the nature of the test configuration, applying $+P$ (i.e. downwards) results in a stress state at the *bottom corner* that is effectively identical to that which arises at the *top corner* when applying $-P$ (i.e. upwards). Therefore, this zero to $+P$ loading regime would subject the top corner to open-and-slipping conditions (i.e. the squares in Figure 6.2) and the bottom corner to closed-and-stuck conditions (i.e. the triangles in Figure 6.2).

If such tests were carried out and failure consistently initiated from the top corner, this would suggest that open-and-slipping conditions cause more damage than closed-and-stuck conditions. Conversely, closed-and-stuck conditions would appear to cause greater damage if failure consistently initiated from the bottom corner. Additionally, in order to check that failure initiated at the corner with the more severe edge conditions and not due to some other effect associated with the test rig, these tests could be performed again but this time with the loading regime from zero to $-P$. For this loading regime, failure would be expected to initiate from the opposite corner from the zero to $+P$ loading regime (i.e. for these tests, the top corner would be closed and stuck and the bottom corner would be open and slipping).

6.3 Mugadu's fretting tests

In this section, a different set of fretting-fatigue experiments is examined, which were carried out by Mugadu [2002] for his doctoral dissertation. These tests were motivated by the need to predict the fretting performance of the spline couplings that are used to join



(a) Image of a spline coupling.

(b) Cross-sectional diagram of a spline coupling.

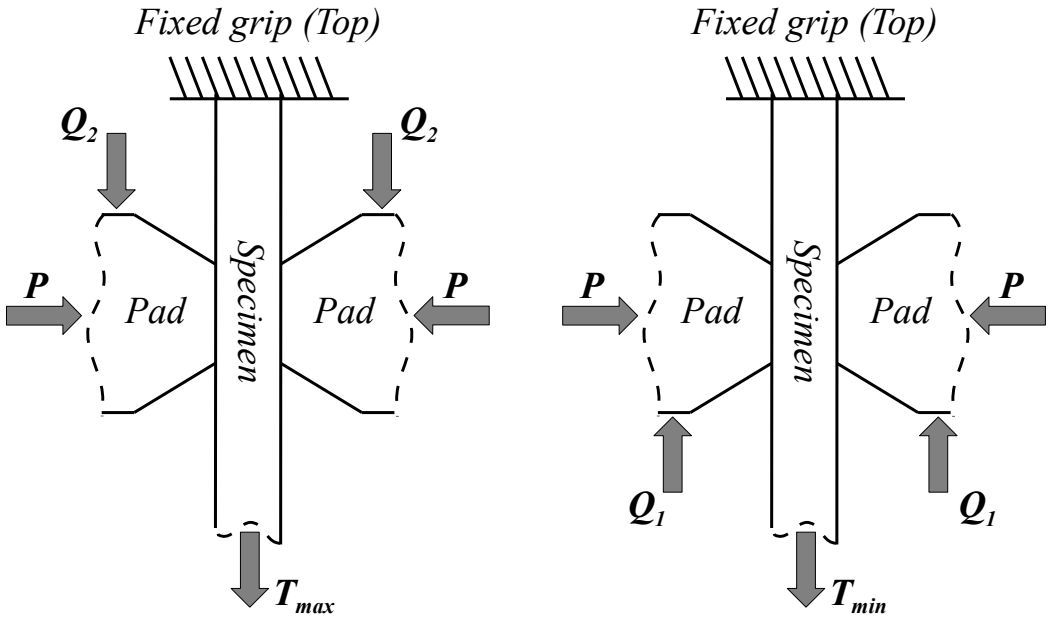
Figure 6.3: Spline coupling used to connect split shafts in gas-turbine aero-engines.

shafts in Rolls-Royce's gas-turbine aero-engines. An image of a spline coupling is shown in Figure 6.3(a), and a diagram of its cross section is shown in Figure 6.3(b). The object of these experiments was to develop an understanding of the key factors that influence the fatigue performance of spline couplings.

As noted by Limmer et al. [2001], although in-service failure of spline couplings is rare, when it does occur, failure usually initiates from the contact edge that is formed between the spline teeth, which is shown in Figure 6.3(b). Since failure initiates near the contact edge, the asymptotic techniques presented in §2-§6.1 may be able to describe the stress state in this region, depending on the type of behaviour arising at the contact edges during these tests. Thus, in this section, these fretting experiments are examined using this asymptotic approach.

6.3.1 Experimental configuration

Mugadu [2002] performed these fretting-fatigue tests at the University of Oxford using a multi-axial test rig, which has been developed and refined over several decades. The



(a) Maximum load.

(b) Minimum load.

Figure 6.4: Idealized diagrams of the complete contact test geometry at the maximum and minimum times in the load cycle.

earliest version of this rig was designed by Bramhall [1973] and incorporated a single hydraulic actuator, which was used to apply a bulk load to a tensile specimen. Cylindrical contact pads were then fastened onto the tensile specimen and were connected to the test rig with springs. These springs caused shear forces to develop at the contact interface when the bulk load was applied, and this resulted in partial-slip and fretting damage. Mugadu [2002] updated this test rig by incorporating a second hydraulic actuator to apply a shear force to the contact (i.e. as a replacement for the springs).[3]

Idealized diagrams of the complete contact formed between the tensile specimen and the contact pads are shown in Figure 6.4. This figure also shows the loading condition at the maximum and minimum points in time during the load cycle, where P is the normal load that forms the contact, T is a bulk tension that is applied to the specimen, and Q_1, Q_2 are shear loads that are applied to the contact pads. Note that P, T, Q_1, Q_2 all have units of force per unit depth, i.e. $[FL^{-1}]$, where F denotes force, and L denotes length.

The loading regime used in these tests was as follows. First, *before* the contact pads

[3] See Chapter 6 of [Mugadu, 2002] for more details on the test rig.

were fastened onto the specimen, a bulk tension, T_{mean}, was applied. This initial bulk tension was set to the mean value of the bulk load during the tests, i.e. $T_{mean} = (T_{max} - T_{min})/2$. The contact pads were then fastened to the specimen with a normal load of P, which was held constant for the duration of the test. The bulk and shear loads were then cycled between their maximum and minimum values. Note that the ranges of the bulk and shear loads that were applied were chosen so as to begin the tests in the *gross-sliding* regime. However, Mugadu [2002] reports that the friction coefficient increased after the first several thousand cycles, so some of these tests may eventually have *shaken down* to a partially adhered state (see §9 for more details on frictional shakedown).

The tests concluded either when the specimen failed or when the specimen was deemed to have 'run out' (i.e. if it did not fail within approximately 10^6 load cycles). Note that most of these tests were carried out in displacement control, but some were carried out in force control. Also note that the tests were started at a frequency of $1Hz$, but this frequency was slowly increased to approximately $4Hz$ during the first 1,000-2,000 load cycles. In the following analysis, it is assumed that this constitutes a quasi-static loading condition, such that inertial effects are not important.

In his dissertation, Mugadu [2002] reports the results of fretting tests that he performed on complete contacts with an edge angle of $\phi = 121°$, which he refers to as Group 1, and on complete contacts with an edge angle of $\phi = 76°$, which he refers to as Group 2. However, only the Group 1 tests on contacts with an edge angle of $\phi = 121°$ are examined here,[4] and it is noted that this angle was chosen because it arises at the edge of the complete contact formed between the spline teeth shown in Figure 6.3(b). For these tests, both the contact pads and the tensile specimens were made of a proprietary material developed by Rolls-Royce, which is referred to as 'CMV steel', and it is noted that the specimens had a square cross section, which was $10mm$ thick. Mugadu [2002] further divided the Group 1 tests into five subsets: namely, Groups A, B, C, D, and E. The

[4]This is because the mode II term in Williams' solution is bounded for the Group 2 tests. Thus, if the contacts were adhered, the contact-edge stress state was almost certainly dominated by the mode I term in Williams' solution, so the plots of K_{II} vs. K_I are less useful for analysing these tests.

CMV steel used for the contact pads and the specimens in test Groups A-D was nitrided, whereas it was non-nitrided for the tests in Group E. Thus, only the tests with nitrided specimens in Groups A-D are considered for consistency in the subsequent analysis.

6.3.2 Mugadu's analysis of the fretting tests

Before applying the asymptotic analysis developed in §2-§6.1, it should be noted that Mugadu [2002] performs an analysis of these tests in Chapter 7 of his doctoral thesis. The first part of this chapter focuses on a special singular finite element that was developed by Akin [1976], which provides more rapid convergence near to singular points in a stress field than the more common linear or quadratic isoparametric elements. Following from this analysis, Mugadu [2002] uses a finite element model with these singular elements to calculate the generalized stress intensity factors arising at the contact edges during the fretting tests.

Because Mugadu [2002] assumes that the corners are slipping for these calculations, the generalized stress intensity factors he obtains are not the same as K_I or K_{II}, which are defined in (2.25). Instead, they are the multiplier corresponding to a one-term expansion of the asymptotic solution for a semi-infinite wedge sliding on a half-plane, which is due to Gdoutos and Theocaris [1975] and Comninou [1976]. Because the corner asymptote generally has a different order of stress singularity when the contact edge is adhered and when it is sliding, much of Mugadu's subsequent analysis focuses on the variation of the singularity as the contact-edge state varies during the load cycle.

Mugadu [2002] reports the values of the (slipping) generalized stress intensity factors, denoted K^*, and notes that contacts that were subjected to greater values of K^* tended to have a shorter fatigue life (as expected). In addition, Mugadu [2002] notes that contacts that were subjected to a greater range of bulk tension, T, tended to have a shorter life, but that varying the contact pressure, P, did not have a significant influence. Finally, Mugadu [2002] notes that every specimen that failed did so from the *top corner*, i.e. on the side furthest away from the hydraulic actuator that applied the bulk tension, T.

6.3.3 Finite element model

The complexity of the analysis pursued by Mugadu [2002] is significantly increased because of his effort to determine whether the sliding asymptote or the adhered asymptote should be applied at different times during the load cycle. This is especially true with regard to the sliding asymptote because its form (i.e. the order of the singularity) is sensitive to the coefficient of friction. As the complexity of such an approach may make it difficult to apply and interpret in practice, a simpler asymptotic analysis is performed here, which is based entirely on Williams' solution. Thus, the analysis pursued here begins by assuming that the contact interface is fully adhered, but the techniques described in §4-§6.1 can be used to determine whether slip and/or separation are implied to occur.

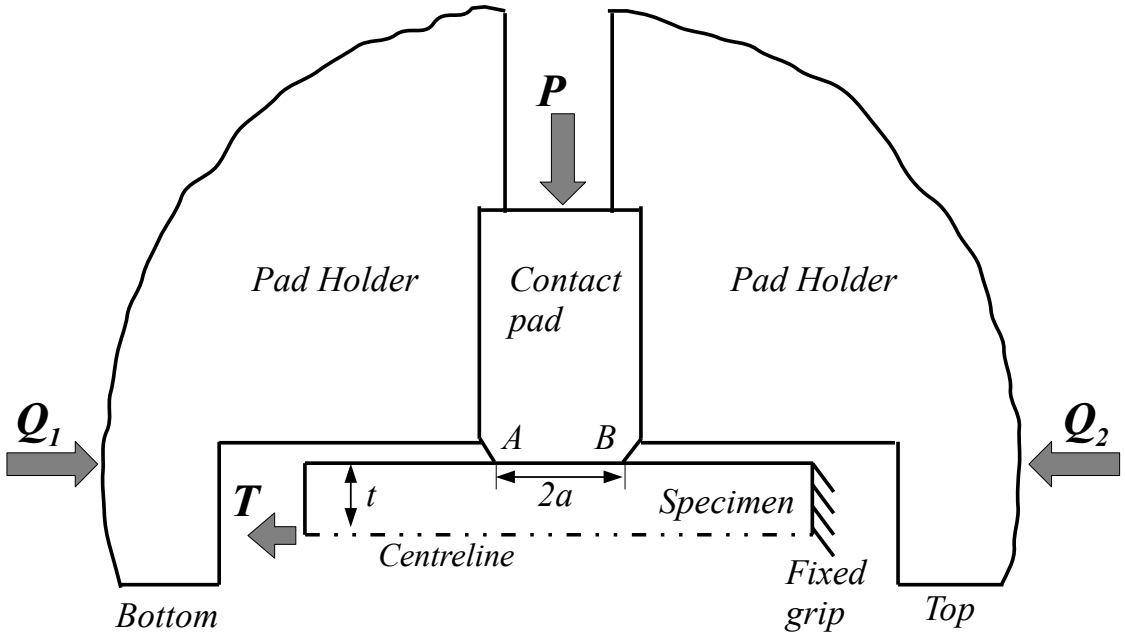
The first step in this analysis is to create a bilateral finite element model of the contact and to obtain calibrations of K_I and K_{II}. Note that the calibrations for Q_1, Q_2 were quite sensitive to seemingly remote geometric features and to the way in which the loads were applied. Thus, the finite element model that was used to obtain K_I and K_{II} includes some details of a special D-shaped holder that was used to hold the contact pads.[5]

An idealized diagram of the finite element model is shown in Figure 6.5(a), and an image of the model is shown in Figure 6.5(b).[6] Notice that only the left half of the test rig shown in Figure 6.4 is modelled here because the test rig is symmetrical about the centreline of the test specimen (i.e. about a vertical centreline in Figure 6.4). This figure also shows the applied loads, P, Q_1, Q_2, T, and the boundary conditions that were applied to the specimen. Also note that a is the half-width of the contact interface, t is the half-width of the specimen, and $t/a = 1$ for the test rig used by Mugadu [2002].

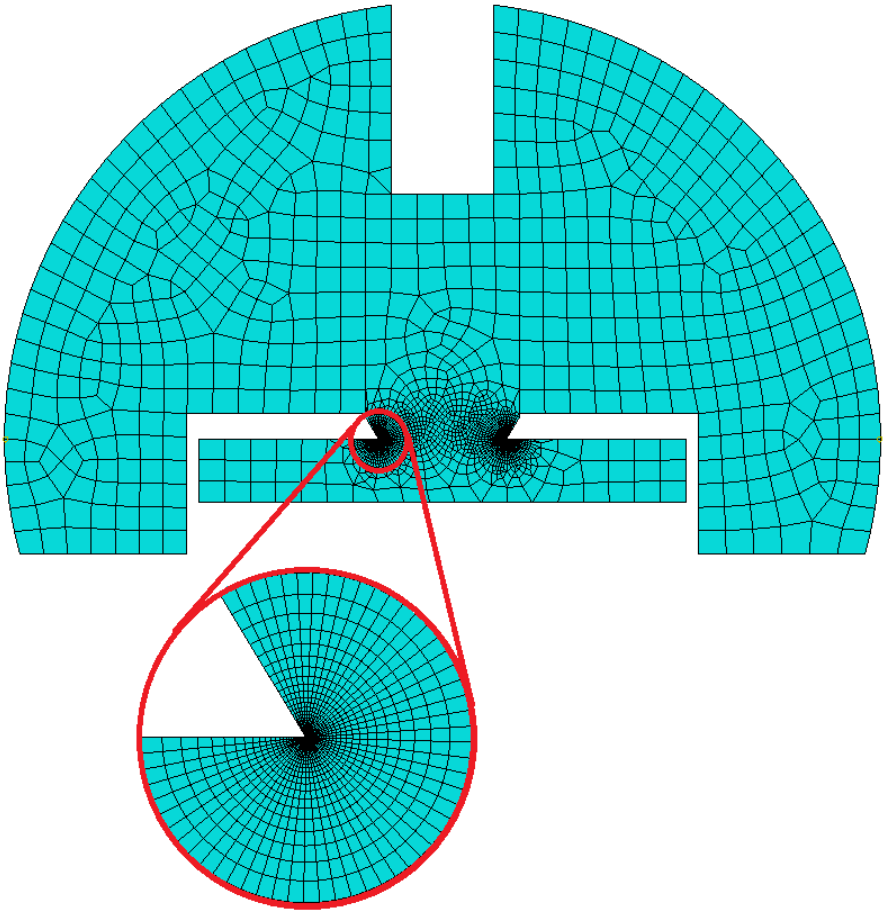
It is noted that the contact pads and the specimens were made of 'CMV steel', and the pad holder was made of brass. Thus, although the finite element model shown in

[5]This pad holder was designed to allow the contact pads to rotate, so they would be self-aligning when the normal load, P, was first applied. It also ensured that the shear load was applied along the line passing through the contact interface, so no tipping moment was produced. See Chapter 6 of [Mugadu, 2002] for more details.

[6]The contact was modelled as a two-dimensional, plane-strain, bilateral geometry in ABAQUS/CAE v6.11, which incorporated a combination of `CPE4R` and `CPE3` elements.



(a) Idealized diagram of the finite element model.



(b) Image of the finite element model.

Figure 6.5: Model of the test rig used by Mugadu [2002].

Figure 6.5 is a bilateral model, separate partitions were included in it for the contact pads, the specimen, and the pad holder. This enables each of these partitions to be assigned different material properties. Thus, Young's modulus, E, and Poisson's ratio, ν, were set to $E = 200GPa$ and $\nu = 0.3$ for the contact pad and the specimen, whereas these were set to $E = 100GPa$ and $\nu = 0.34$ for the pad holder.

6.3.4 Type A and type B contact edges

Notice that because of the nature of the boundary conditions and the way in which the loads were applied, the two corners of the test rig that are shown in Figure 6.5 do not possess any further symmetry properties. Hence, calibrations must be obtained for both of these corners. However, before obtaining these calibrations, it is helpful to introduce the sign convention shown in Figure 6.6 to distinguish between them.

To help interpret this figure, let a polar coordinate set (r, θ) be defined at each contact edge, where $\theta = 0$ lies along the bisector of the notch formed by the contact, and θ is taken to be positive in the counter-clockwise direction (as in Figure 4.3). Notice that for the bottom corner (i.e. the left side of Figure 6.5(a)), the interface between the pad and the specimen lies at a positive value of θ, and this type of corner is referred to as a type A corner. For this type of corner, the interface lies along the horizontal line shown in Figure 6.6 (which corresponds to $\theta = 29.5°$ when $\phi = 121°$). Conversely, for the top corner, the interface between the pad and the specimen lies at a negative value of θ, i.e. $\theta = -29.5°$ when $\phi = 121°$, and this is referred to as a type B corner.

All of the analysis presented in §4-§5 pertains to type A corners (since the upper corner of the clamped-cantilever test rig is a type A corner). However, it is important to distinguish between type A and type B corners because the sign of θ_{int} (i.e. the angle corresponding to the contact interface) affects the implications for contact-edge behaviour that are described in the previous chapters. Note that the implications for type A and type B corners are different because of the properties of the $f(\theta)$ functions of Williams' solution, which are plotted in Figure 2.4 for the example notch angle of $2\alpha = 270°$.

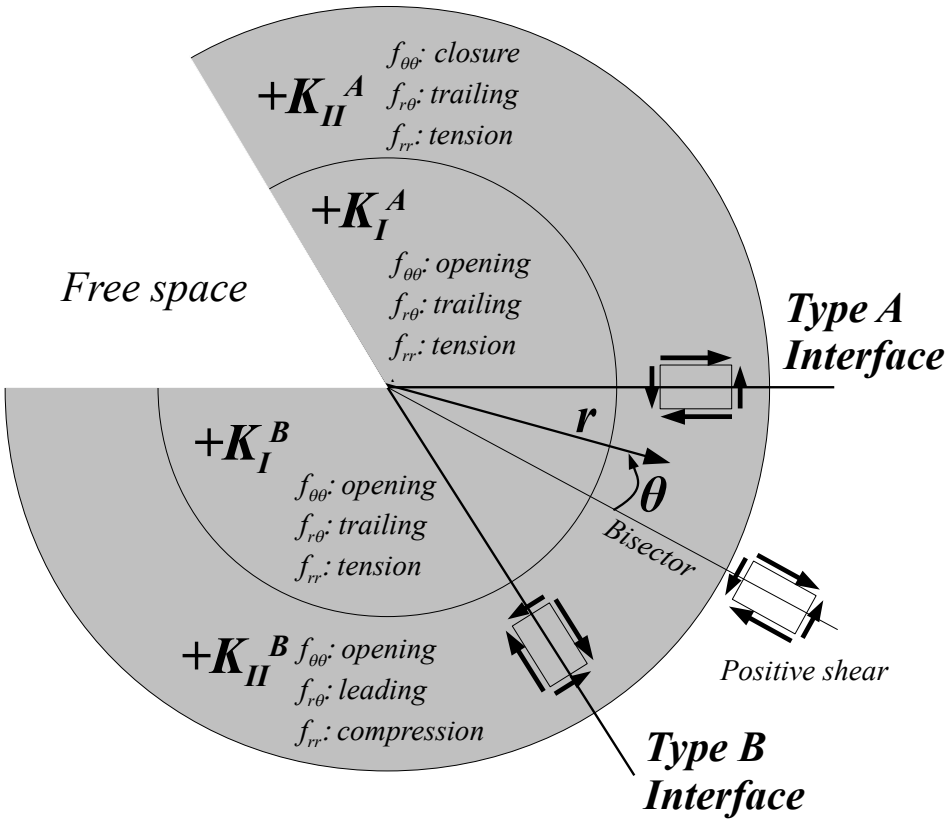


Figure 6.6: Sign convention for type A and type B contact edges.

The implications of the $f(\theta)$ functions for type A corners are shown in Figure 6.6. A positive value of K_I^A implies: opening, a trailing-edge shear traction, and a tensile radial stress (in the mode I dominated region near the contact edge). In addition, a positive value of K_{II}^A implies: closure, a trailing-edge shear traction, and a tensile radial stress (in the mode II dominated region of the contact).[7] Note that these implications can be inferred simply by examining the $f(\theta)$ functions, which are shown in Figure 2.4.

Figure 6.6 shows that while the implications for $+K_I^A$ and $+K_I^B$ are identical, those for $+K_{II}^B$ are the opposite of those for $+K_{II}^A$. This is because of the anti-symmetric properties of the $f(\theta)$ functions for the mode II eigensolution (see Figure 2.4(b)). This illustrates that a positive value of K_{II}^B has the same implications for contact behaviour as a negative value of K_{II}^A. For this reason, when experimental results for both type A and

[7] Of course, these implications reverse for negative values of K_I^A, K_{II}^A: that is, an implication of closure changes to opening, a leading-edge shear traction changes to a trailing-edge shear traction, and a tensile radial stress changes to a compressive radial stress.

type B corners are shown on plots of K_{II} vs. K_I, it is most convenient to multiply the K_{II}^B values by -1, so the implications derived in the previous sections remain unchanged. In other words, it is most convenient to use plots of $K_{II}^A, -K_{II}^B$ vs. K_I^A, K_I^B.

6.3.5 Asymptotic analysis of the test rig

Calibrations of the generalized stress intensity factors were obtained for both contact edges from the bilateral finite element model shown in Figure 6.5, which are given by[8]

$$\left\{\begin{array}{c} K_I^A a^{\lambda_I - 1} \\ K_{II}^A a^{\lambda_{II} - 1} \\ \hline K_I^B a^{\lambda_I - 1} \\ K_{II}^B a^{\lambda_{II} - 1} \end{array}\right\} = \left[\begin{array}{cccc} -0.1537 & 0.1223 & -0.0609 & 0.0751 \\ 0.1280 & 0.1610 & -0.1109 & 0.1562 \\ \hline -0.1534 & -0.2116 & 0.2731 & 0.0752 \\ -0.1285 & 0.4236 & -0.4737 & -0.1564 \end{array}\right] \left\{\begin{array}{c} P/a \\ Q_1/a \\ Q_2/a \\ T/t \end{array}\right\}. \quad (6.1)$$

The calibrations for the bottom corner are denoted K_I^A, K_{II}^A, and those for the top corner are denoted K_I^B, K_{II}^B. Recall that P, Q_1, Q_2, T all have units of force per unit depth, i.e. $[FL^{-1}]$.

Note that $(\lambda_I, \lambda_{II}) = (0.5116, 0.7258)$, and $(g_{r\theta}^I, g_{r\theta}^{II}) = (0.296, -0.877)$ when $\phi = 121°$. Thus, for this contact angle, the minimum friction coefficient required to prevent edge slip is $f > g_{r\theta}^I = 0.296$. However, as the contact angle used in these tests is greater than $101.7°$, $|g_{r\theta}^{II}| > g_{r\theta}^I$ (see §4.2.2). Therefore, if the multiplier on the mode II term is large enough (i.e. if $d_0 \ll a$, see §5.4), slip may in fact initiate from the interior (mode II) region of the contact and propagate out to the contact edge – even if the response is case (a) and $f > 0.296$.

Mugadu [2002] reported the friction coefficient, f, present during each of the fretting experiments he conducted. One surprising feature of these results is that the friction coefficient was found to be different for the maximum and minimum halves of the load cycle. In general, f was found to be greater during the minimum half of the load cycle than during the maximum half of the cycle, and Mugadu [2002] hypothesized that this

[8]Note that when one unit of force is applied to the specimen, only half of this force is transferred into the half of the test rig that is shown in Figure 6.5. Thus, the calibration for T corresponds to 0.5 units of force being applied to the half of the specimen that is included in the finite element model.

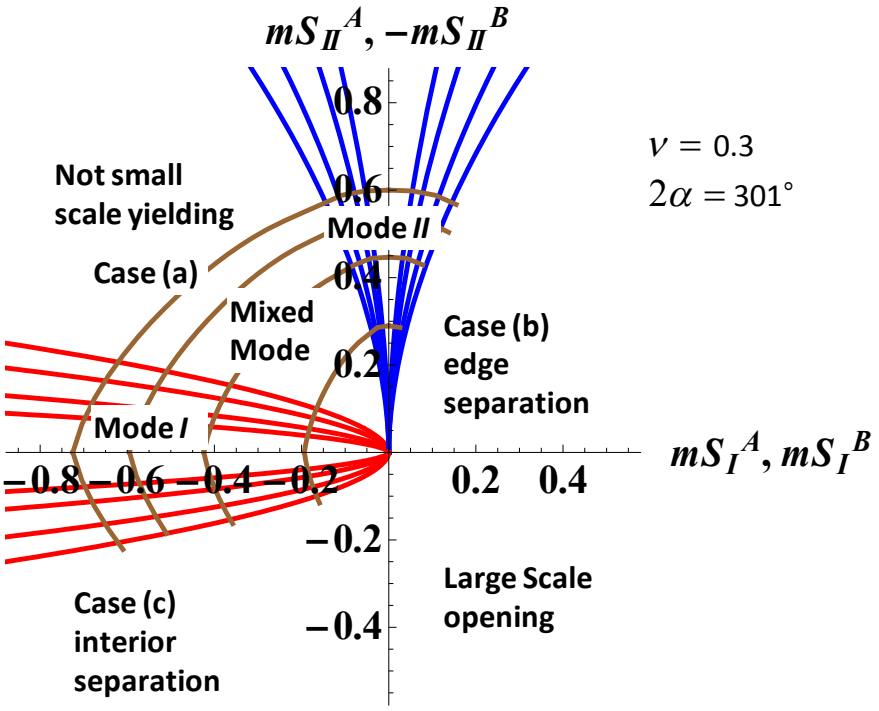


Figure 6.7: Contact-edge behaviour for type A and type B corners when $2\alpha = 301^\circ$. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown (all calculated assuming $\nu = 0.3$).

effect occurred because of the lack of symmetry in the way in which the shear load was applied. In addition, very large variation was observed in the friction coefficient from test to test: that is, f ranged from being almost 0 to as high as 1.2.

6.3.6 Analysis of fretting tests

A good way to begin the analysis of these fretting tests is to plot the generalized stress intensity factors at the maximum and minimum points in the load cycle for both corners on a plot of the form of Figure 6.1. As the curves that display the small-scale-yielding and mode-mixity information in Figure 6.1 are for a contact angle of $\phi = 90^\circ$, they must be recalculated for the contact angle that was used in Mugadu's fretting tests. In addition, since both type A and type B corners are to be examined on this plot, its axes now become $mS_{II}^A, -mS_{II}^B$ vs. mS_I^A, mS_I^B, so the implications of each quadrant remain the same as those in Figure 6.1. This type of plot is shown in Figure 6.7 for $\phi = 121^\circ$ (i.e. $2\alpha = 301^\circ$).

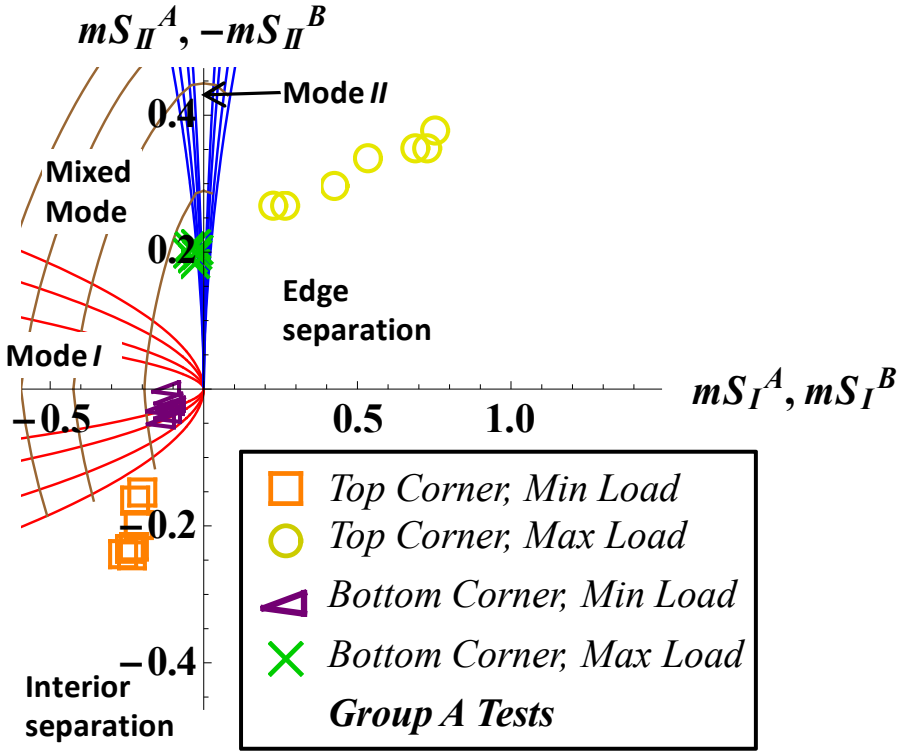
Before plotting the extreme values of K_I, K_{II} during the load cycle for Mugadu's

tests, recall that prior to the formation of the contact, T was set to its mean value. Notice that this initial mean bulk load only resulted in a uni-axial stress state prior to the formation of the contact, whereas application of the normal load resulted in a singular stress state at the contact edges. Subsequent application of the bulk and shear loads also added further singular contributions to the stress state near the contact edges. Clearly, the (bounded) contribution of the initial bulk load should not be included on a plot in dimensionless-generalized-stress-intensity-factor space, so the state of generalized stress intensity is assumed to have been controlled by $P + T_{max} + Q_2 - T_{mean}$ at the maximum point in the cycle and by $P + T_{min} + Q_1 - T_{mean}$ at the minimum point in the cycle.

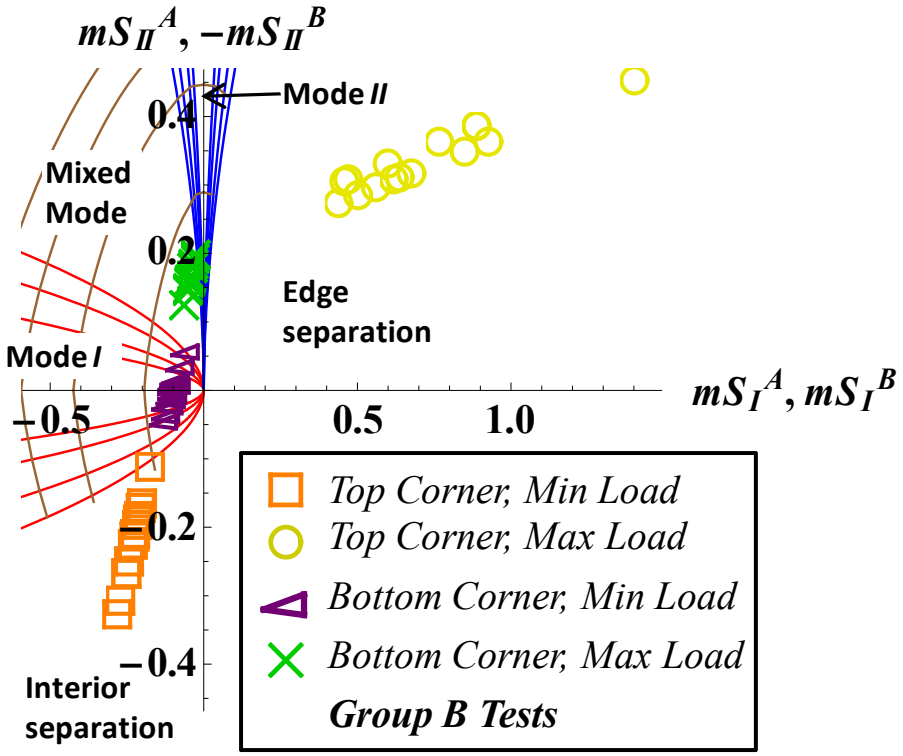
It is noted that a total of 50 fretting tests were performed in Groups A, B, C, and D with 7 tests in Group A, 15 in Group B, 14 in Group C, and 14 in Group D. The values of $mS_{II}^A, -mS_{II}^B, mS_I^A, mS_I^B$ at the extreme points in the load cycle are shown in Figure 6.8 for the tests in Groups A and B and in Figure 6.9 for the tests in Groups C and D. Note that the information shown in Figure 6.7 is also included in these figures. The results plotted in these figures were computed using the values of the applied normal, bulk, and shear loads reported in Appendix A of [Mugadu, 2002], and the values of mS_I, mS_{II} were calculated using the calibrations given in (6.1) with the contact half-width, $a = 5mm$, selected as the characteristic length dimension.

The most obvious remark that can be made about the data shown in Figures 6.8-6.9 is that the top contact edges were subjected to more severe conditions than were the bottom contact edges: that is, both the magnitude of the stresses at the extreme points in the cycle and the stress range during the cycle were significantly greater at the top contact edges. This feature of the data is in good agreement with the fact that all of the specimens that failed did so from the top contact edges.

Another feature of these tests that is clearly illustrated by Figures 6.8-6.9 is that significant amounts of slip and separation occurred at the top contact edges during the maximum point in the load cycle because the response is quite far into the case (b) region of the plot. Moreover, interior initiated slip (and possibly separation) *may* also have oc-

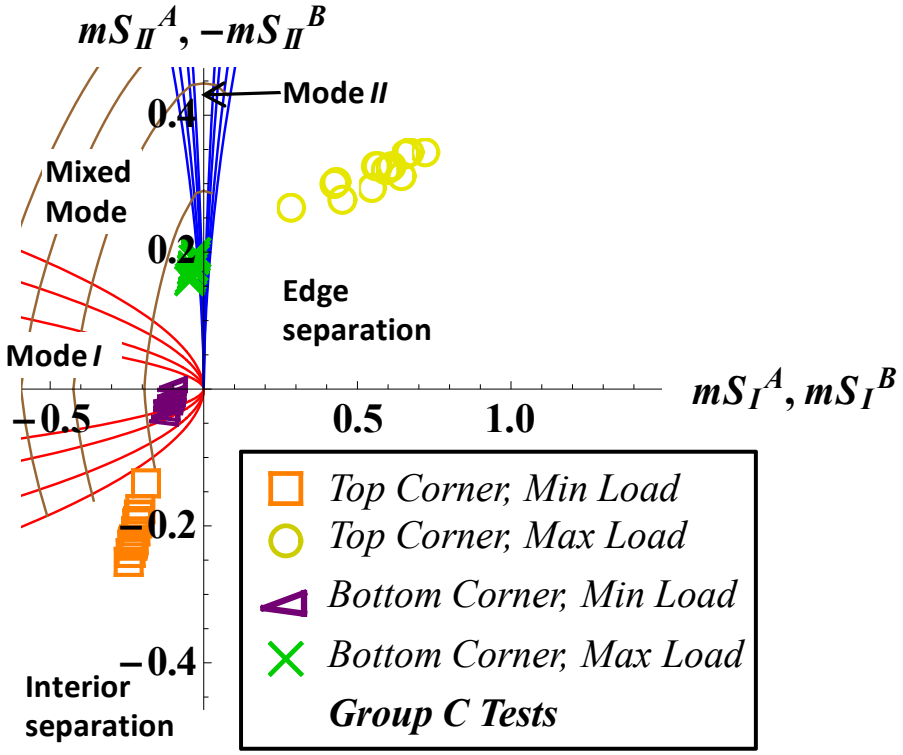


(a) Group A results.

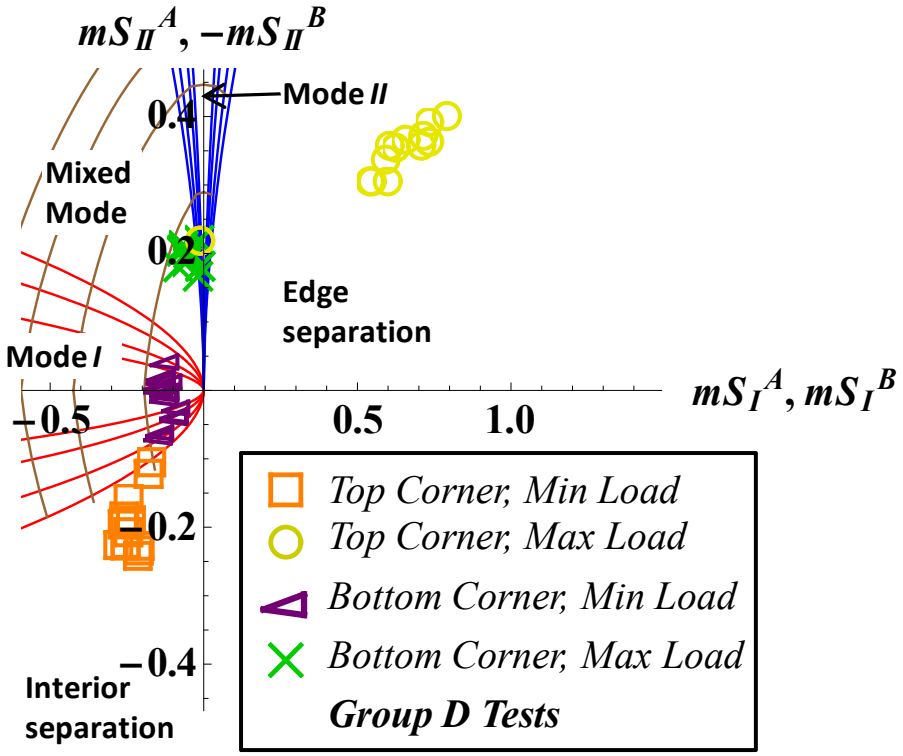


(b) Group B results.

Figure 6.8: Plots of mS_I, mS_{II} at the extreme points of the load cycle for test Groups A and B of the data reported by Mugadu [2002]. These plots also include the mode-mixity and small-scale-yielding information shown in Figure 6.7.



(a) Group C results.



(b) Group D results.

Figure 6.9: Plots of mS_I, mS_{II} at the extreme points of the load cycle for test Groups C and D of the data reported by Mugadu [2002]. These plots also include the mode-mixity and small-scale-yielding information shown in Figure 6.7.

curred at the top contact edges during the minimum point in the load cycle because the response is quite far into the case (c) region. However, higher-order terms in Williams' series expansion are likely to become important in these cases, so the actual contact-edge behaviour at this point in the load cycle cannot be determined using the current approach.

On the other hand, Figures 6.8-6.9 show that the lower contact edges probably remained closed during these tests because the data lie, for the most part, within the case (a) region of the plot. Although some points do stray slightly into the case (c) region, it is still unlikely that separation occurred in these cases because the contact-edge stress state was dominated by the mode I eigensolution, which implies closure (see §5.5).

6.3.7 Examination of fatigue performance

Because failure always initiated from the top contact edges, these contact edges obviously controlled the fatigue performance of these fretting tests. Unfortunately, Figures 6.8-6.9 suggest that a significant amount of slip and separation occurred at the top contact edges during the loading regime, especially when the maximum load was applied. Consequently, the calculations presented in §5 cannot be used to estimate the extent of slip or separation. In addition, the behaviour at the minimum point in the load cycle cannot be analysed with the tools presented in §5 because the data are outside the region where the mode I term is dominant.

Although the stress state at the top contact edges certainly is not accurately described by Williams' solution at the maximum or minimum points in the load cycle, the severity of the loading is still captured, at least to some extent, by the range of the generalized stress intensity factors applied during these tests, i.e. ΔK_I or ΔK_{II}.[9] Given that the loading regime subjects the top contact edges to a wide range of both K_I and K_{II}, it is not clear what the relative importance of K_I and K_{II} might be for fatigue performance (if any, given that so much slip and separation occurred in these tests). Thus, for the sake of simplicity, the full range of the mode I dimensionless generalized stress intensity factor

[9]Notice from (6.1) that ΔK_I and ΔK_{II} are linearly related to the load range in terms of P, Q_1, Q_2, T, so a larger ΔK_I corresponds to a more severe loading regime.

at the top corner, ΔmS_I^B, is plotted against the base-10 logarithm of the number of cycles to failure, $log_{10}(N)$, in Figure 6.10.[10]

In this figure, tests that eventually failed from the top corner are shown as a cross, and those that did not fail, i.e. tests that were 'run outs', are shown as a circle. One aspect of these plots that should be pointed out immediately is that the value of N that is shown is the number of cycles to *failure*, not to macroscopic crack initiation. Therefore, the value of N that is shown in these plots is the sum of two contributions: namely, (i) some number of cycles during which no macroscopic cracks were present in the specimen, but microscopic cracks were nucleating; and (ii) some number of cycles during which one or several macroscopic cracks were present and were in the crack propagation phase.

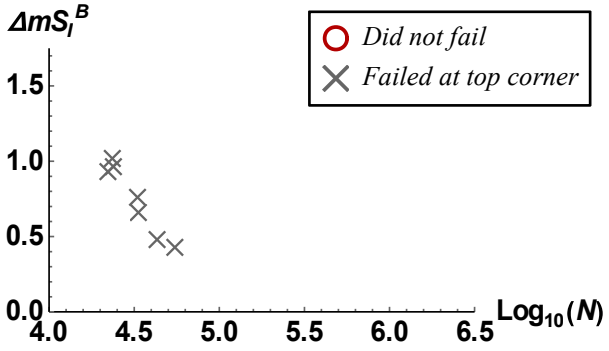
The asymptotic approach employed here can only be used to analyse the crack nucleation phase during which the critical region where damage is accumulating is near to the contact edge. Once the crack begins propagating into the specimen, it will move into a region that is outside the influence of the contact-edge stress state. Thus, during the crack propagation phase, a standard fracture mechanics approach becomes appropriate for predicting the number of cycles to failure using a Paris-style crack propagation law, e.g. [Pugno et al., 2006].

Although this is one possible source of scatter in the data shown in Figure 6.10, it probably is not the main source. This is because practical engineering assemblies that are subject to high-cycle fatigue, e.g. gas-turbine engines, frequently spend the majority of their service life in the crack nucleation phase and only a short fraction of their life in the crack propagation phase [Nicholas, 1999]. Hence, failure probably occurred shortly after macroscopic cracks appeared.[11]

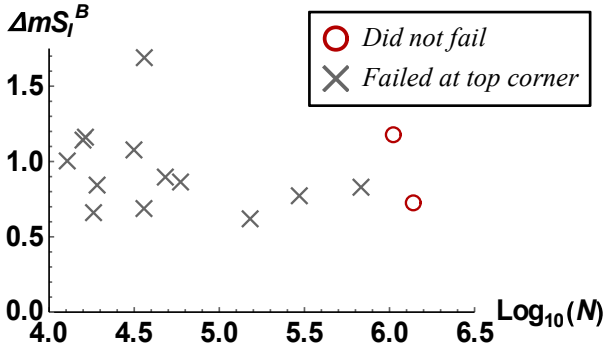
Leaving aside the issue of the aggregation of the crack nucleation and crack propaga-

[10] It is noted that ΔmS_{II}^B or some other weighted combination of ΔmS_I^B and ΔmS_{II}^B could also be used to correlate against fatigue life, and their peak or mean values could also be taken into account. These more sophisticated analyses are not pursued here, but it is noted that plots of ΔmS_{II}^B vs. $log_{10}(N)$ look similar to plots of ΔmS_I^B vs. $log_{10}(N)$ in the sense that the level of scatter and the trends that are present are similar in both types of plots.

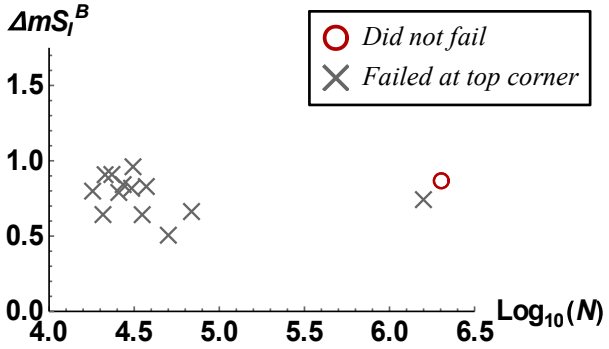
[11] For example, Figures 7 and 8 in [Juoksukangas et al., 2013] show that failure occurred quite shortly after macroscopic cracks initiated in their fretting tests.



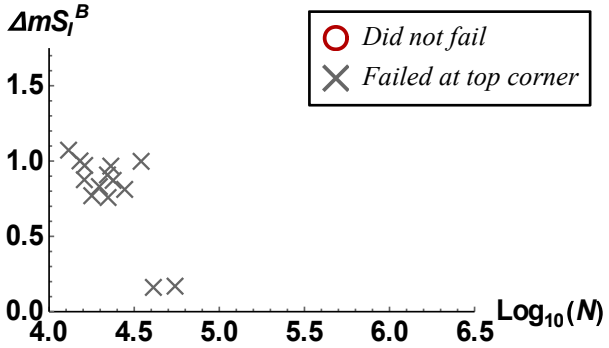
(a) Group A results.



(b) Group B results.



(c) Group C results.



(d) Group D results.

Figure 6.10: Plots of ΔmS_I^B vs. $log_{10}(N)$ for the data reported by Mugadu [2002].

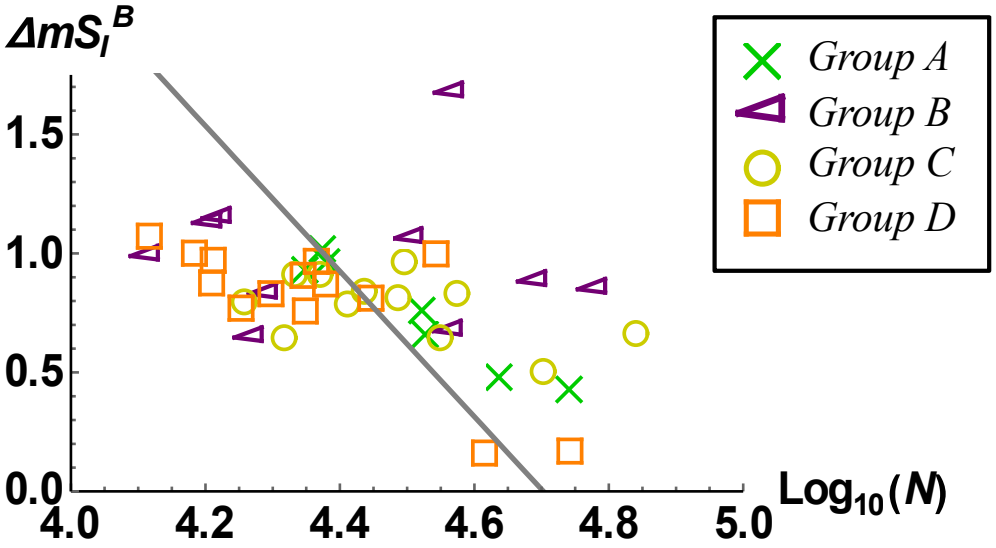


Figure 6.11: Plot of ΔmS_I^B vs. $log_{10}(N)$ for tests that failed within 10^5 cycles. The best-fit line based on a least-squares linear regression is also shown.

tion phases, the most notable feature of these data is the large amount of scatter that is present. The main exception to this is the Group A tests, which exhibit quite a strong, negative, linear trend. However, the trend that is found in case A is not maintained in the other test groups. Since Group A only comprises 7 out of the 50 tests that were carried out, the strength of this trend is probably due to chance rather than to the particular conditions arising during these tests, especially since the conditions are quite similar across the test groups (see Figures 6.8-6.9).

Figure 6.10 also shows that all the tests in Groups A and D failed in less than 10^5 cycles. In contrast, 5 tests in Groups B and 2 tests in Group C lasted for more than 10^5 cycles. However, the better fatigue performance of these 7 tests does not appear to bare any relationship to ΔmS_I^B. In addition, a qualitative examination of the friction coefficients during these tests, which are reported in Appendix A of [Mugadu, 2002], does not reveal any obvious pattern.

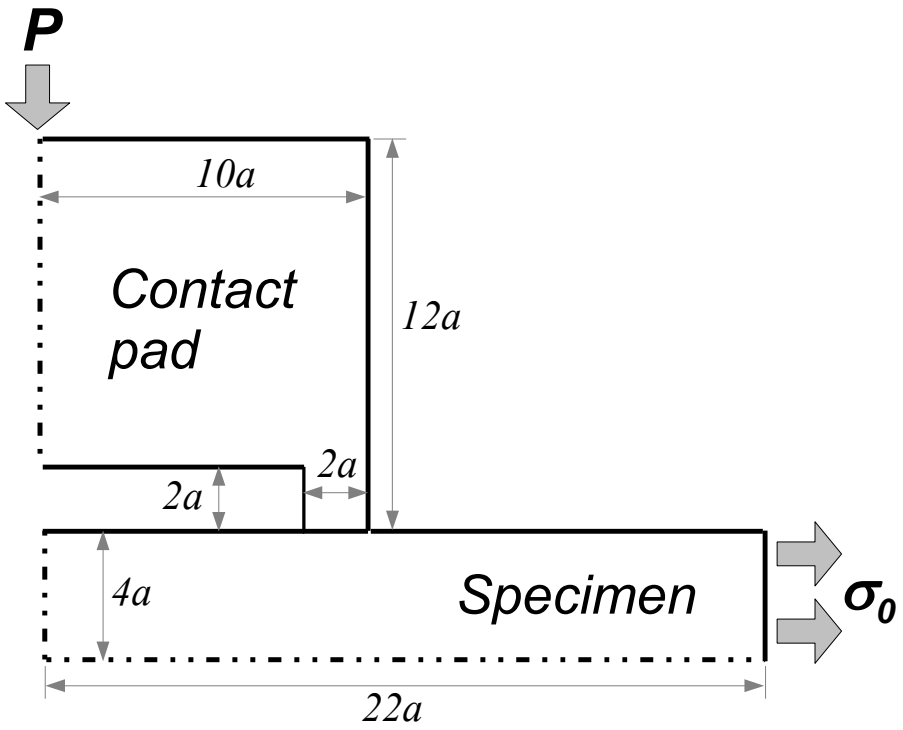
However, from the point of view of component safety, the possibility of some components lasting significantly longer than predicted is much less problematic than the converse. Thus, for the purpose of predicting the fatigue life of the spline couplings that these tests were designed to simulate, these 7 tests can safely be neglected. Thus, the 43 tests

that failed within 10^5 cycles are plotted together in Figure 6.11 on a plot of ΔmS_I^B vs. $log_{10}(N)$. This figure also includes the best-fit line based on a least-squares linear regression, which is given by the equation $log_{10}(N) = 4.70 - 0.327\ \Delta mS_I^B$. This regression has an R^2 of 0.204, which illustrates that there is a very large amount of variability in the number of cycles to failure, N, that is not explained by the range of the mode I generalized stress intensity factor, ΔmS_I^B. However, this regression does capture a statistically significant negative trend in the data ($F_{1,41} = 10.5$, $p < .005$), which suggests that larger values of ΔmS_I^B tended to result in shorter fatigue lives.

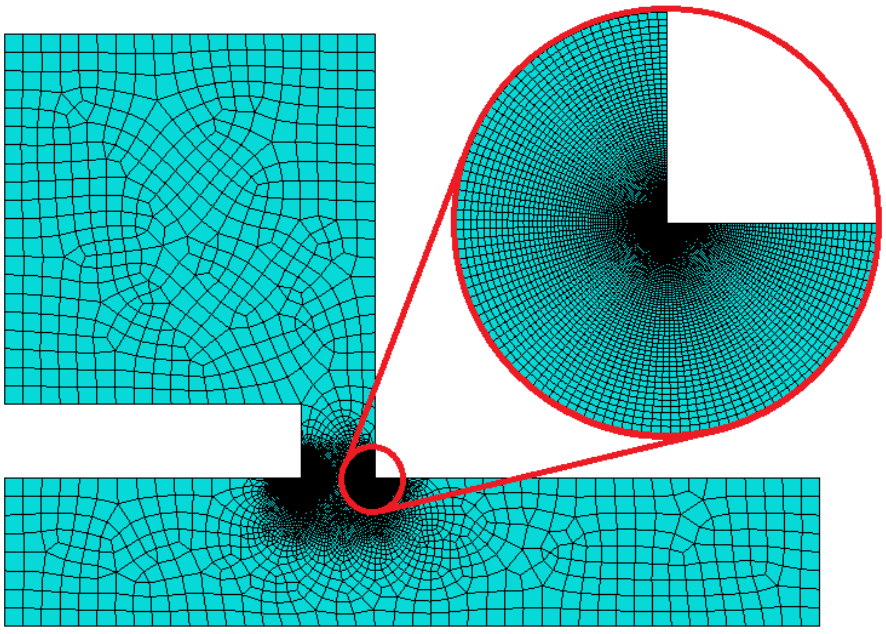
This result is not surprising because larger values of ΔmS_I^B are associated with more severe loading regimes in terms of the applied loads, viz. P, Q_1, Q_2, T. On the other hand, this figure does not provide any evidence of a fatigue limit, i.e. a value of ΔmS_I^B below which failure never occurs, which would be expected to be present for the steel alloy that was used in these tests. Thus, these results suggest that ΔmS_I^B is not the right parameter to correlate against the fatigue life of these tests. This is probably because large amounts of slip and separation occurred (see Figures 6.8-6.9), which means that the bilateral assumption cannot be used to analyse these tests.

6.4 Noraphaiphipaksa's fretting tests

In this section, fretting-fatigue tests that were carried out by Noraphaiphipaksa et al. [2013] are examined using the asymptotic approached developed in §2-§6.1. These tests were performed by fastening 'bridge' type contact pads onto a tensile specimen and then cycling the bulk load in the specimen. The specimens were machined from a SM430A steel rod, and the contact pads were machined from a SM430A steel plate. Both materials had a Young's modulus of $E = 210GPa$ and a Poisson's ratio of $\nu = 0.3$. However, the specimens had a yield stress of $\sigma_y = 467MPa$, whereas the contact pads had a yield stress of $\sigma_y = 270MPa$.



(a) Idealized diagram of the test rig.



(b) Image of the finite element model.

Figure 6.12: Model of one quarter of the test rig used by Noraphaiphipaksa et al. [2013].

6.4.1 Asymptotic analysis of test rig

As before, a bilateral finite element model of the test rig was created to obtain calibrations of the generalized stress intensity factors at the contact edges. Note that this test rig incorporated four contact interfaces (i.e. two on each side of the test specimen). Since the test rig is symmetrical about the centreline of the contact pads and about the centreline of the tensile specimen, only one quarter of it must be modelled. An idealized diagram of the finite element model is shown in Figure 6.12(a), and an image of the model is shown in Figure 6.12(b). Using this model, the following calibrations[12] were obtained for the mean contact pressure, p_0, and for the bulk load in the specimen, σ_0:

$$\begin{Bmatrix} K_I^A a^{\lambda_I - 1} \\ K_{II}^A a^{\lambda_{II} - 1} \\ \hline K_I^B a^{\lambda_I - 1} \\ K_{II}^B a^{\lambda_{II} - 1} \end{Bmatrix} = \left[\begin{array}{cc} -0.1693 & -0.1669 \\ 0.3656 & 0.5092 \\ \hline -0.5617 & 0.6163 \\ -0.3034 & -0.2770 \end{array}\right] \begin{Bmatrix} p_0 \\ \sigma_0 \end{Bmatrix}, \tag{6.2}$$

where K_I^A, K_{II}^A are for the inner corner (i.e. the left corner in Figure 6.12), and K_I^B, K_{II}^B are for the outer corner. Note that p_0 was applied using a proving ring, which is accounted for by the load P that is shown in Figure 6.12(a), such that $p_0 = P/2a$, where a is the half-width of the contact.

Since the contact-edge angle of this test rig is $\phi = 90°$, the results obtained in §4.3 apply: namely, $(g_{r\theta}^I, g_{r\theta}^{II}) = (0.543, -0.219)$, and $(\lambda_I, \lambda_{II}) = (0.5445, 0.9085)$. Thus, the minimum friction coefficient required to prevent the contact edges from slipping is $f > g_{r\theta}^I = 0.543$. Note that Noraphaiphipaksa et al. [2013] found the friction coefficient to be approximately 0.8 based on the results of a gross-sliding friction test. Hence, whenever the contact edges were closed, i.e. $K_I > 0$, they were almost certainly stuck as well.

The load ratio at which edge separation first occurs can also be calculated based on the condition that $K_I > 0$ for closure at the contact edges (as in §4.3). This calculation

[12] The test rig was modelled in ABAQUS/CAE v.6.11-1 as a bilateral, two-dimensional, plane-strain geometry, which incorporated `CPE4R` and `CPE3` elements. Note that Young's modulus was set to $E = 210GPa$, and Poisson's ratio was set to $\nu = 0.3$ in this model.

reveals that the contact edges will remain stuck at the inner corners when $\sigma_0/p_0 > -0.99$, and the outer corners will remain stuck when $\sigma_0/p_0 < 1.1$. Thus, provided that the ratio of the applied loads remains within the range $-0.99 < \sigma_0/p_0 < 1.1$, the contact edges will remain stuck, and fretting damage will not occur.

6.4.2 Application to fretting tests

For the fretting-fatigue tests reported by Noraphaiphipaksa et al. [2013], the contact pads were first fastened onto the tensile specimen using a proving ring, resulting in a mean contact pressure of $p_0 = 60MPa$. A bulk load, σ_0, was then applied to the specimen, which was cycled about a mean bulk load of zero with amplitudes (i.e. half the total range) within the range $150MPa$ to $300MPa$, e.g. from $-150MPa$ to $+150MPa$. This loading regime was applied until the specimen failed or 'ran out' (i.e. did not fail within 10^7 cycles), and the total number of cycles during the test was recorded.

Since the loading regime just described results in load ratios of between $\sigma_0/p_0 = \pm 2.5$ and $\sigma_0/p_0 = \pm 5$, it is clear that significant amounts of separation occurred at the contact edges during all of these tests. The implications of the loading trajectory on contact-edge behaviour can also be seen in more detail by showing the extreme points in the loading history on a plot of the form of Figure 6.1, and this is done in Figure 6.13. Note that the generalized stress intensity factors that are shown on this plot were calculated using the value of σ_y from the specimen, i.e. $\sigma_y = 467MPa$.

This figure shows that when the normal load, p_0, is first applied, both contact edges remain closed and stuck since they are in the case (a) region of the plot (see the triangle and the circle in Figure 6.13). In addition, the contact-edge response due to p_0 alone is within 1% small-scale yielding and results in a mode I dominated process zone at both contact edges.

Figure 6.13 also includes 4 additional data points for each contact edge that show the contact-edge response when bulk loads of $\sigma_0 = \pm 150MPa$ and $\sigma_0 = \pm 300MPa$ are applied. These data points show that when the maximum bulk load, $\sigma_{\max}$, (i.e. a positive,

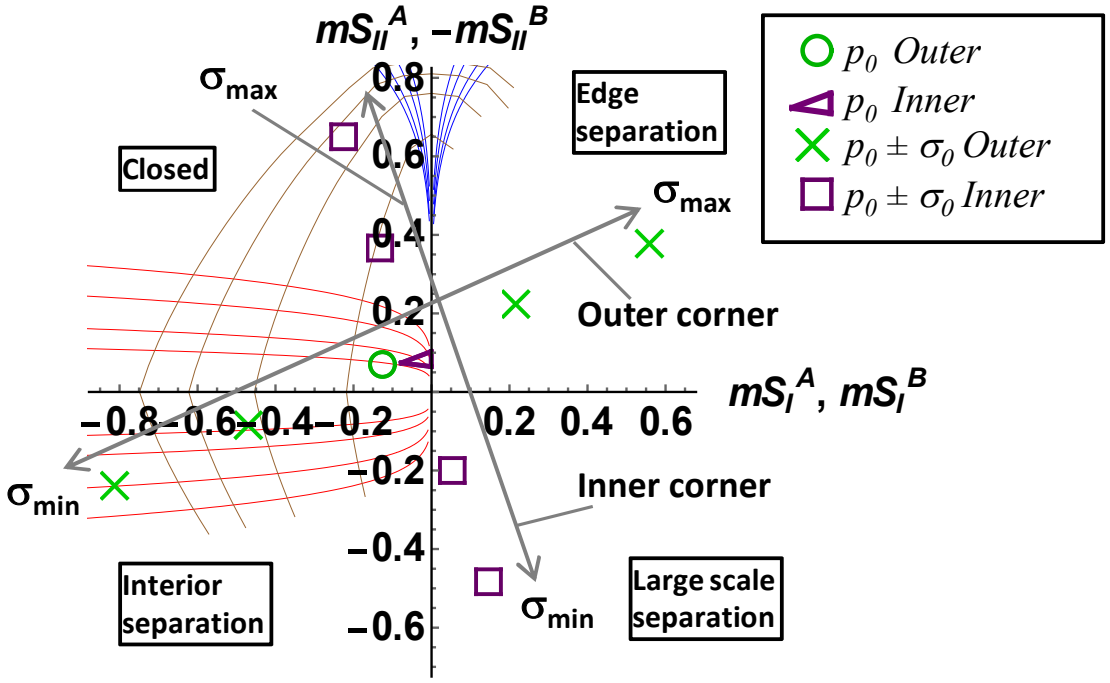


Figure 6.13: Plot of mS_I, mS_{II} at the extreme points of the load cycle for the fretting tests reported by Noraphaiphipaksa et al. [2013]. Mode-mixity results for $\rho_I, \rho_{II} = 0.95, 0.9, 0.8, 0.7$ and small-scale-yielding results for $r_p^{\max}/a = 0.01, 0.05, 0.1, 0.15$ are also shown for $2\alpha = 270°$ (all calculated assuming $\nu = 0.3$).

tensile load) is applied, the inner contact edges remain in the case (a) region, although the process zones become mixed-mode in character. Note that the process zones at the inner contact edges remain within 1% small-scale yielding when $\sigma_0 = +150MPa$, but they are only within 10% small-scale yielding when $\sigma_0 = +300MPa$. Conversely, when $\sigma_{\max}$ is applied, the response of the outer contact edges is very far into the case (b) region, implying that significant amounts of slip and separation occurred.

Figure 6.13 also shows that when the minimum bulk load, $\sigma_{\min}$, is applied, the response of the inner contact edges moves into the region where both K_I and K_{II} imply separation. On the other hand, the response of the outer contact edges moves very slightly into the case (c) region of the plot. However, since the response is still dominated by the mode I term, interior separation will not occur. Instead, the outer contact edges will remain closed at the minimum point in the load cycle.

It is noted that Noraphaiphipaksa et al. [2013] performed a detailed unilateral finite element analysis to determine the contact behaviour arising during these fretting tests.

Some of these results are shown in Figure 3 of [Noraphaiphipaksa et al., 2013] for contact pressures of $p_0 = 30MPa, 60MPa, 90MPa$ and bulk loads of $\sigma_0 = \pm 180MPa$. In particular, the results for the case when $p_0 = 60MPa$ show that approximately 50% of the contact interface separates at both extremes of the load cycle. Specifically, the outer contact edges separate at the maximum bulk load, and they are closed and slipping at the minimum bulk load. Conversely, the inner contact edges separate at the minimum bulk load, but they are closed and stuck at the maximum bulk load. Thus, these unilateral finite element results are in good agreement with the qualitative implications of the asymptotic results presented in Figure 6.13.

6.4.3 Implications for fretting performance

A significant portion of the analysis presented in [Noraphaiphipaksa et al., 2013] focuses on predicting the fatigue life of the specimens by numerically modelling the propagation of fatigue cracks during the tests. It is noted that Noraphaiphipaksa et al. [2013] reported that cracks invariably initiated from the outer contact edges of the test rig. In addition, for their analysis, these authors assumed that the number of cycles that were spent in the crack nucleation phase was small in comparison to the number of cycles that were spent in the crack propagation phase. Thus, they ignored the influence of crack nucleation in their predictions of fatigue performance.

The asymptotic analysis developed here can only be applied during the crack nucleation phase for which the contact-edge stress field dominates behaviour in the region where damage is accumulating. Once a crack begins to propagate deep into the specimen, the contact-edge stress state loses relevance, and a standard fracture mechanics approach should be used. Thus, since the majority of the fatigue life of these specimens was spent in the crack propagation phase, this asymptotic analysis should not be used.

In addition, the unilateral finite element results obtained by Noraphaiphipaksa et al. [2013] and Figure 6.13 both show that significant amounts of separation occurred throughout the load cycle. Thus, these tests were conducted using a loading regime that makes an

asymptotic analysis of the contact edge very difficult. As a consequence, very little can be determined about these tests using this asymptotic analysis (aside from the qualitative implications discussed above). Indeed, Figure 6.13 does not help explain why cracks always initiated from the outer contact edges as opposed to the inner contact edges since this figure gives very similar implications for both contact edges.

For these reasons, it is not appropriate to construct plots of ΔmS_I vs. $log_{10}(N)$ for these tests since it is unlikely that ΔK_I or ΔK_{II} controlled their fatigue life. This means that if new tests were conducted with the same contact angle and an identical loading regime in terms of ΔK_I or ΔK_{II}, but the details of the test rig were significantly modified, it is unlikely that the new fatigue data would fall along the same curve on a plot of ΔmS_I vs. $log_{10}(N)$. This is also probably the case for the tests performed by Mugadu [2002]. On the other hand, there is a good chance that the results obtained by Juoksukangas et al. [2013] could be reproduced with quite a different test rig, provided that the same contact angle were used, and the loading regime were similar (in terms of ΔK_I or ΔK_{II}).

6.5 Discussion

The analysis presented in this chapter reveals that a very wide range of contact-edge conditions were explored in the three sets of fretting-fatigue tests that have been considered here. As a result, the asymptotic analysis developed in §2-§6.1 does not apply equally well to all three test groups. The validity of applying this asymptotic approach is most easily assessed by plotting the extreme values of the dimensionless generalized stress intensity factors, mS_I, mS_{II}, on an annotated plot of mS_{II} vs. mS_I, such as that shown in Figure 6.1. This is done in Figure 6.2 for the tests performed by Juoksukangas et al. [2013], in Figures 6.8-6.9 for those by Mugadu [2002], and in Figure 6.13 for those by Noraphaiphipaksa et al. [2013].

These figures show that in the tests performed by Juoksukangas et al. [2013], the contact edges were subjected to only a moderate amount of separation. In contrast, in the tests performed by Mugadu [2002] and Noraphaiphipaksa et al. [2013], the contact

edges were subjected to a much more severe loading regime, which resulted in significant amounts of slip and separation. Thus, the asymptotic analysis used here, which is based on the assumption of a bilateral contact interface, is much more useful for the tests performed by Juoksukangas et al. [2013] than for those by Mugadu [2002] and Noraphaiphipaksa et al. [2013]. This is evidenced by the large amount of scatter in Mugadu's data (see Figure 6.10), which is not present in Juoksukangas' data (see Figure 4.6).

If the asymptotic analysis presented in this chapter were carried out prior to running the tests that were performed by Mugadu [2002] and Noraphaiphipaksa et al. [2013], it would have been clear that these tests would excite a significant amount of separation at the contact edges. It should be noted that some asymptotic forms have been developed to describe the bounded stress state arising at the edge of an incomplete contact, e.g. [Hills et al., 2012b]. However, the results presented in this chapter suggest that neither the 'notch' nor the 'bounded' asymptotic forms can fully describe the behaviour near the contact edges during these tests. This is because the notch asymptote cannot be applied when such a large amount of separation occurs, but the sharp edge of the contact also makes it difficult to use the bounded asymptotes.

Of course, performing an analysis that incorporates both the singular 'notch' asymptote and the bounded 'incomplete' asymptote at different points during the load cycle would be quite complicated. Thus, since it had been decided that such a severe loading regime was to be applied during the tests by Mugadu [2002] and Noraphaiphipaksa et al. [2013], it may have been more informative to use flat-and-rounded contact pads instead of sharp-edged contact pads. This would have enabled the bounded asymptote to be used (provided that the friction coefficient was sufficiently high, and the loading regime remained within small-scale yielding), which would have enabled the test data to be compared with that obtained by other researchers (see [Hills et al., 2012b]). Conversely, sharp-edged contact pads could have been used but with a modified loading regime that would have produced only a minimal amount of edge separation, which would have enabled the asymptotic analysis developed in §2-§6.1 to have been applied.

6.6 Summary

In this chapter, the tools developed in §2-§5 are applied to analyse several groups of fretting-fatigue tests that were performed on complete contacts. First, the conditions under which the analysis described in §3 can be applied to complete contacts is examined. General plots in dimensionless-generalized-stress-intensity-factor space are then constructed, which can be used to plot experimental data and to determine the behaviour of the contact edges during fretting experiments.

The data reported by Juoksukangas et al. [2013] are then examined using these plots. The results reveal that during half of the load cycle, the contact edges were closed, and the process zones at the contact edges were heavily mode I in character and were within 1% small-scale-yielding. However, during the other half of the cycle, a moderate amount of edge separation occurred. Nevertheless, this asymptotic approach can be used to analyse these tests, and this is evidenced by the minimal scatter in the plot of the generalized stress intensity factor range vs. cycles to failure.

In contrast, for the tests performed by Mugadu [2002], a significant amount of slip and separation is implied to have occurred at the top contact edges from which failure initiated. Hence, these tests are more difficult to analyse using this asymptotic approach, and this may be evidenced by the much greater level of scatter in the plots of the generalized stress intensity factor range vs. cycles to failure. However, this analysis did reveal that the top contact edges were subjected to a more severe load range than the bottom contact edges, which explains why failure always initiated from the top contact edges.

Similarly, for the tests conducted by Noraphaiphipaksa et al. [2013], the loading regime resulted in a significant amount of separation throughout the load cycle. Thus, it is not appropriate to use the asymptotic approach presented in this thesis to quantify the fatigue life of these tests since a bilateral model does not represent the contact-edge behaviour accurately. However, the load range in which this asymptotic analysis could be applied is found, and it is hoped that future tests will explore this load range.

Chapter 7

Static reduction of finite element models

In this chapter, a sub-structuring procedure for significantly reducing the number of degrees of freedom in finite element models of two-dimensional frictional contacts is presented. This is done by using a quasi-static model-reduction technique in which all the internal degrees of freedom are eliminated from the global stiffness matrix. The reduced contact stiffness matrix, which contains only the degrees of freedom along the contact interface, is then obtained along with vectors that account for the influence of the applied loads. Using this reduced model, transient simulations of the evolution of contact behaviour can be run at least 10 times more quickly than with a full finite element model. The reduced contact stiffness matrix also has several other more advanced applications, one of which is examined in §9.

Note that parts of this chapter have been published in [Thaitirarot et al., 2014][1] and [Flicek, 2013].

7.1 Introduction

Up until this point, the calculations that have been presented for analysing complete contacts have been based on Williams' solution for a semi-infinite notch. Indeed, ob-

[1] Note that the work published in Thaitirarot et al. [2014] was carried out in collaboration with another DPhil candidate, A. Thaitirarot. The author's main contribution to this work was the development and verification of the static-reduction procedure, and its implementation in MATLAB. Indeed, the author wrote all the MATLAB code provided in the appendix of Thaitirarot et al. [2014]. A. Thaitirarot's main contribution to this publication was obtaining the results in the 'Example Problems' section, so these results are not included in this chapter.

taining calibrations of the generalized stress intensity factors is the only aspect of these calculations that requires the finite element method (or some other numerical technique) to be employed. However, even in this case, a bilateral model is used, which imposes a continuity of displacement condition along the interface. Thus, the computational expense of imposing frictional contact conditions is avoided.

Although asymptotic approaches are very useful for determining the response at the contact edges, there are some aspects of contact behaviour that they cannot be used to describe. First, they provide no information on contact behaviour at the interior of the contact, where the stress state is heavily influenced by the finite geometry of the particular problem under consideration. In addition, no asymptotic approach has yet been presented in the literature for quantifying the evolution of the slip or separation zones at the edge of a complete contact subjected to cyclic loading: that is, all the calculations presented thus far, such as those discussed in §5 (including both those based on violations and those that employ the distributed dislocation technique), only consider the case of monotonic loading. However, the motivation for modelling complete contacts comes from the need to predict their performance when they are subject to fatigue conditions, which is of course a situation that involves cyclic loading.

Thus, it is often necessary to use the information obtained from corner asymptotes in conjunction with numerical results obtained from a unilateral finite element model. These are generally made by creating two (or more) separate domains and then connecting them with contact elements, which use either the penalty method or the augmented Lagrange multiplier method. However, such models usually require a large number of degrees of freedom to obtain sufficient accuracy in the contact solution, so they frequently require quite a significant amount of computational resources for each simulation. This is especially so for cyclic loading problems, which can require numerous load cycles to be simulated before the contact solution reaches a steady-state response. Moreover, testing multiple load cases is extremely computationally expensive if a full model is used, e.g. when studying the influence of initial conditions on the steady-state response.

One way to improve the efficiency of such analyses is to reduce the number of degrees of freedom in the finite element model by eliminating the internal degrees of freedom from the global stiffness matrix, $\boldsymbol{K}$, and obtaining the *contact stiffness matrix*, $\boldsymbol{K}^{\mathrm{C}}$. This procedure is referred to as *sub-structuring* or *static reduction*, and it reduces the contact problem to that of a system of massless, rigid blocks connected with a general stiffness matrix, $\boldsymbol{K}^{\mathrm{C}}$, where each block is potentially in frictional contact with a rigid plane obstacle. In addition to $\boldsymbol{K}^{\mathrm{C}}$, vectors that account for the influence of the applied loads, $\boldsymbol{f}^{w}$, can also be determined.

With this reduced model, the evolution of the contact interface with time can be solved by iteratively applying the contact inequalities (see Ahn and Barber [2008]) or by using linear-complementarity methods (see Bertocchi [2009]), so contact elements that employ the penalty or augmented Lagrange multiplier method are not required. In addition, the contact stiffness matrix has several more advanced applications, such as determining the *shakedown limit* for cyclically loaded frictional contacts (i.e. the value of cyclic load above which the contact is guaranteed to dissipate energy due to frictional slip), which is examined in §9.

The concept of sub-structuring and static reduction is well known; indeed, many commercial finite element packages allow the user to apply this technique to eliminate degrees of freedom during certain analyses. However, it is not currently possible to perform the more advanced calculations that require the contact stiffness matrix, e.g. calculating the shakedown limit, within these software packages. In addition, once $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ have been obtained, the evolutionary contact problem can be solved in any data analysis package such as MATLAB, which gives the user much greater control over how the contact law is implemented.

7.2 Single elastic body in contact with a rigid body

To explain how to perform the static-reduction procedure, it is most straightforward to begin with the simplest case of a single elastic body in contact with a rigid surface, such

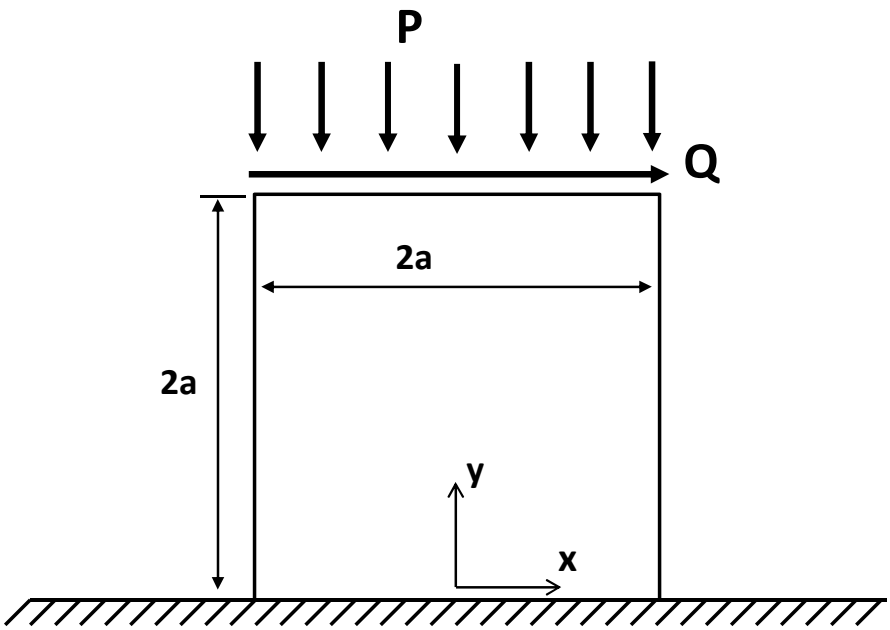


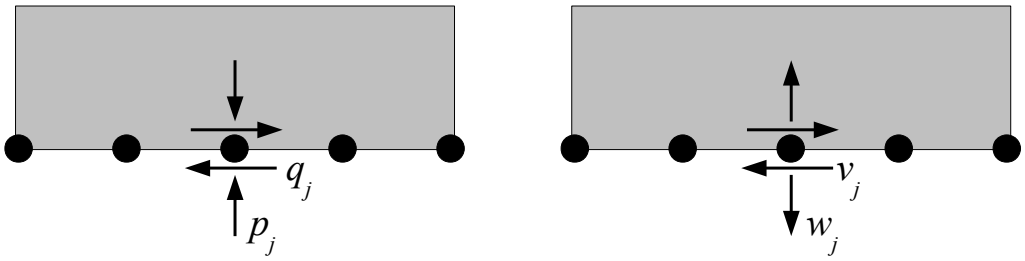
Figure 7.1: Diagram of complete contact between an elastic square and a rigid surface.

as that shown in Figure 7.1. In §7.3, it is shown that the more relevant case of frictional contact between two elastic bodies can be solved by performing the procedure described in this section for each body separately and then merging the results.

Thus, suppose that the two-dimensional elastic square shown in Figure 7.1 has been meshed using some finite element software package but that no boundary conditions or loads have been applied. Most software packages, e.g. ABAQUS/CAE, contain commands allowing the user to export the global stiffness matrix, $\boldsymbol{K}$, into a text file. The details of this procedure are of course program specific, so these are omitted here, but it is noted that the steps required to do this in ABAQUS/CAE are detailed in Flicek [2013].[2]

Once $\boldsymbol{K}$ is obtained for the problem described above, each node in the finite element model must be categorized as belonging to one of three sets: the contact nodes, $\mathcal{S}_C$; the externally loaded nodes, $\mathcal{S}_E$; or the unloaded nodes, $\mathcal{S}_I$. For example, for the model shown in Figure 7.1, the nodes along the top edge of the elastic square are in $\mathcal{S}_E$, those along the bottom edge are in $\mathcal{S}_C$, and all other nodes are in $\mathcal{S}_I$. Note that the nodes in

[2]Note that more detailed instructions, including sample MATLAB code that performs the static reduction, can be found in [Flicek, 2013] or in the appendix of Thaitirarot et al. [2014] and can also be downloaded from the following URL: `http://www.eng.ox.ac.uk/stress/`.



(a) Nodal forces: normal, p_j, and shear, q_j.

(b) Nodal displacements: normal, w_j, and tangential, v_j.

Figure 7.2: Sign convention showing the positive direction for the forces and displacements at the contact nodes.

$\mathcal{S}_E$ can either be loaded with a prescribed displacement or a prescribed force. Since the reduction procedure must be carried out slightly differently for these two load cases, $\mathcal{S}_E$ must be further partitioned into: $\mathcal{S}_U$, where non-zero nodal displacements are prescribed, and $\mathcal{S}_T$, where non-zero nodal forces are prescribed. However, in most practical problems, the loads are either specified in force control or in displacement control, so either $\mathcal{S}_U$ or $\mathcal{S}_T$ is null.

7.2.1 Sign convention

For the two-dimensional case, each node in the model is associated with two degrees of freedom. Thus, a vector of the reaction forces at each contact node, j, can be written as

$$\boldsymbol{f}_j^{\mathrm{C}} = \{q_j, p_j\}^{\mathrm{T}}, \tag{7.1a}$$

where $j \in \{1, 2, 3, ..., N\}$, N is the number of contact nodes, p_j is the force normal to the contact surface, and q_j is the shear component tangential to the surface. The sign convention adopted for the nodal forces is shown in Figure 7.2(a), where p_j is taken to be positive in compression. Similarly, a vector of the relative displacement at each contact node, j, can be written as

$$\boldsymbol{u}_j^{\mathrm{C}} = \{v_j, w_j\}^{\mathrm{T}}, \tag{7.1b}$$

where w_j is the normal (opening) component, and v_j is the tangential (slip) component. The sign convention for the relative displacements is shown in Figure 7.2(b), where a positive w_j contributes to a positive gap between the two bodies. Vectors containing the forces and displacements along the contact interface can then be written as

$$\boldsymbol{f}^{\mathrm{C}} = \{\boldsymbol{f}_1^{\mathrm{C}}, \boldsymbol{f}_2^{\mathrm{C}}, \boldsymbol{f}_3^{\mathrm{C}}, ..., \boldsymbol{f}_N^{\mathrm{C}}\}^T \tag{7.2a}$$

$$\boldsymbol{u}^{\mathrm{C}} = \{\boldsymbol{u}_1^{\mathrm{C}}, \boldsymbol{u}_2^{\mathrm{C}}, \boldsymbol{u}_3^{\mathrm{C}}, ..., \boldsymbol{u}_N^{\mathrm{C}}\}^T. \tag{7.2b}$$

The analysis here concerns contact between linear-elastic materials subjected to quasi-static loading, so the relationship between the nodal forces and nodal displacements along the contact interface can be written as

$$\boldsymbol{f}^{\mathrm{C}} = \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{\mathrm{C}} + \boldsymbol{f}^{w}, \tag{7.3}$$

where $\boldsymbol{K}^{\mathrm{C}}$ is a positive semi-definite and symmetric *contact stiffness matrix*, and $\boldsymbol{f}^{w}$ is a vector that accounts for the influence of the applied loads. Specifically, $\boldsymbol{f}^{w}$ corresponds to the contact nodal forces that would be generated by the applied loads if the contact nodal displacements were constrained to be zero (i.e. if $\boldsymbol{u}^{\mathrm{C}} = \boldsymbol{0}$).

7.2.2 Eliminating the unloaded nodes

The global stiffness matrix, $\boldsymbol{K}$, relates the forces, $\boldsymbol{f}$, and displacements, $\boldsymbol{u}$, at all the nodes in the finite element model as

$$\boldsymbol{f} \equiv \begin{Bmatrix} \boldsymbol{f}^{\mathrm{I}} \\ \boldsymbol{f}^{\mathrm{E}} \\ \boldsymbol{f}^{\mathrm{C}} \end{Bmatrix} = \boldsymbol{K}\boldsymbol{u} \equiv \begin{bmatrix} \boldsymbol{K}^{\mathrm{II}} & \boldsymbol{K}^{\mathrm{IE}} & \boldsymbol{K}^{\mathrm{IC}} \\ \boldsymbol{K}^{\mathrm{EI}} & \boldsymbol{K}^{\mathrm{EE}} & \boldsymbol{K}^{\mathrm{EC}} \\ \boldsymbol{K}^{\mathrm{CI}} & \boldsymbol{K}^{\mathrm{CE}} & \boldsymbol{K}^{\mathrm{CC}} \end{bmatrix} \begin{Bmatrix} \boldsymbol{u}^{\mathrm{I}} \\ \boldsymbol{u}^{\mathrm{E}} \\ \boldsymbol{u}^{\mathrm{C}} \end{Bmatrix}, \tag{7.4}$$

where the superscripts $(\mathrm{I}, \mathrm{E}, \mathrm{C})$ correspond to nodes in the set $(\mathcal{S}_I, \mathcal{S}_E, \mathcal{S}_C)$, respectively. Notice that a majority of the elements of $\boldsymbol{K}$ will usually relate to nodes in $\mathcal{S}_I$ since most

nodes are located in the interior of the body (i.e. an area), whereas $\mathcal{S}_E$ and $\mathcal{S}_C$ only contain nodes on the surface of the body (i.e. an edge). Thus, the number of degrees of freedom in the stiffness matrix can be significantly reduced by eliminating the nodes in $\mathcal{S}_I$. Since the loads are applied quasi-statically, the body must be in equilibrium at all times during the loading history. Thus, Newton's third law requires that the net force at the unloaded nodes, $\mathcal{S}_I$, is zero, i.e.

$$\boldsymbol{f}^{\mathrm{I}} = \boldsymbol{0}, \tag{7.5}$$

and this condition can be used to eliminate the unloaded nodes.

To do this, the first row of (7.4) can be separated from the other two rows, which gives

$$\boldsymbol{f}^{\mathrm{I}} = \boldsymbol{K}^{\mathrm{II}}\boldsymbol{u}^{\mathrm{I}} + \boldsymbol{K}^{\mathrm{IE}}\boldsymbol{u}^{\mathrm{E}} + \boldsymbol{K}^{\mathrm{IC}}\boldsymbol{u}^{\mathrm{C}}, \tag{7.6}$$

and

$$\left\{ \begin{array}{c} \boldsymbol{f}^{\mathrm{E}} \\ \boldsymbol{f}^{\mathrm{C}} \end{array} \right\} = \left[\begin{array}{ccc} \boldsymbol{K}^{\mathrm{EI}} & \boldsymbol{K}^{\mathrm{EE}} & \boldsymbol{K}^{\mathrm{EC}} \\ \boldsymbol{K}^{\mathrm{CI}} & \boldsymbol{K}^{\mathrm{CE}} & \boldsymbol{K}^{\mathrm{CC}} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{u}^{\mathrm{I}} \\ \boldsymbol{u}^{\mathrm{E}} \\ \boldsymbol{u}^{\mathrm{C}} \end{array} \right\}. \tag{7.7}$$

Using (7.5) with (7.6), an expression for the displacements at the unloaded nodes, $\boldsymbol{u}^{\mathrm{I}}$, can be obtained as

$$\boldsymbol{u}^{\mathrm{I}} = -\left[\boldsymbol{K}^{\mathrm{II}}\right]^{-1} \left[\begin{array}{cc} \boldsymbol{K}^{\mathrm{IE}} & \boldsymbol{K}^{\mathrm{IC}} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{u}^{\mathrm{E}} \\ \boldsymbol{u}^{\mathrm{C}} \end{array} \right\}. \tag{7.8}$$

This expression for $\boldsymbol{u}^{\mathrm{I}}$ can then be substituted into (7.7), which gives

$$\begin{aligned} \left\{ \begin{array}{c} \boldsymbol{f}^{\mathrm{E}} \\ \boldsymbol{f}^{\mathrm{C}} \end{array} \right\} = \Bigg[-\left[\begin{array}{c} \boldsymbol{K}^{\mathrm{EI}} \\ \boldsymbol{K}^{\mathrm{CI}} \end{array} \right] \left[\boldsymbol{K}^{\mathrm{II}}\right]^{-1} \left[\begin{array}{cc} \boldsymbol{K}^{\mathrm{IE}} & \boldsymbol{K}^{\mathrm{IC}} \end{array} \right] \\ + \left[\begin{array}{cc} \boldsymbol{K}^{\mathrm{EE}} & \boldsymbol{K}^{\mathrm{EC}} \\ \boldsymbol{K}^{\mathrm{CE}} & \boldsymbol{K}^{\mathrm{CC}} \end{array} \right] \Bigg] \left\{ \begin{array}{c} \boldsymbol{u}^{\mathrm{E}} \\ \boldsymbol{u}^{\mathrm{C}} \end{array} \right\}, \end{aligned} \tag{7.9}$$

and by defining a new partially reduced matrix, $\boldsymbol{L}$, as

$$\boldsymbol{L} = \left[-\left[\begin{array}{c} \boldsymbol{K}^{\mathrm{EI}} \\ \boldsymbol{K}^{\mathrm{CI}} \end{array} \right] \left[\boldsymbol{K}^{\mathrm{II}}\right]^{-1} \left[\begin{array}{cc} \boldsymbol{K}^{\mathrm{IE}} & \boldsymbol{K}^{\mathrm{IC}} \end{array} \right] + \left[\begin{array}{cc} \boldsymbol{K}^{\mathrm{EE}} & \boldsymbol{K}^{\mathrm{EC}} \\ \boldsymbol{K}^{\mathrm{CE}} & \boldsymbol{K}^{\mathrm{CC}} \end{array} \right] \right], \tag{7.10}$$

(7.9) can be rewritten as

$$\left\{ \begin{array}{c} \boldsymbol{f}^{\mathrm{E}} \\ \boldsymbol{f}^{\mathrm{C}} \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{L}^{\mathrm{EE}} & \boldsymbol{L}^{\mathrm{EC}} \\ \boldsymbol{L}^{\mathrm{CE}} & \boldsymbol{L}^{\mathrm{CC}} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{u}^{\mathrm{E}} \\ \boldsymbol{u}^{\mathrm{C}} \end{array} \right\}. \tag{7.11}$$

To obtain $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$, (7.11) must be reduced further, but the next steps in the procedure depend on the way in which the external loads are applied. Thus, the force-controlled and displacement-controlled cases are considered separately below.

7.2.3 Force-controlled loading

The aim is now to obtain $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ by further reducing the $\boldsymbol{L}$ matrix. Since the loads are applied in force control, $\mathcal{S}_U$ is null, so $\mathcal{S}_E = \mathcal{S}_T$. Furthermore, $\boldsymbol{f}^{\mathrm{E}}$ is a known function of time, t, while $\boldsymbol{u}^{\mathrm{E}}$ is unknown. It is noted that (7.11) can be written in equation form as

$$\boldsymbol{f}^{\mathrm{E}} = \boldsymbol{L}^{\mathrm{EE}}\boldsymbol{u}^{\mathrm{E}} + \boldsymbol{L}^{\mathrm{EC}}\boldsymbol{u}^{\mathrm{C}} \tag{7.12a}$$

$$\boldsymbol{f}^{\mathrm{C}} = \boldsymbol{L}^{\mathrm{CE}}\boldsymbol{u}^{\mathrm{E}} + \boldsymbol{L}^{\mathrm{CC}}\boldsymbol{u}^{\mathrm{C}}. \tag{7.12b}$$

Thus, (7.12a) can be solved for $\boldsymbol{u}^{\mathrm{E}}$ as

$$\boldsymbol{u}^{\mathrm{E}} = \left[\boldsymbol{L}^{\mathrm{EE}}\right]^{-1} \left[\boldsymbol{f}^{\mathrm{E}} - \boldsymbol{L}^{\mathrm{EC}}\boldsymbol{u}^{\mathrm{C}}\right]. \tag{7.13}$$

This expression for $\boldsymbol{u}^{\mathrm{E}}$ can then be substituted into (7.12b), which gives

$$\begin{aligned} \boldsymbol{f}^{\mathrm{C}} &= \boldsymbol{L}^{\mathrm{CE}} \left[\left[\boldsymbol{L}^{\mathrm{EE}}\right]^{-1} \left[\boldsymbol{f}^{\mathrm{E}} - \boldsymbol{L}^{\mathrm{EC}}\boldsymbol{u}^{\mathrm{C}}\right]\right] + \boldsymbol{L}^{\mathrm{CC}}\boldsymbol{u}^{\mathrm{C}} \\ &= \boldsymbol{K}^{\mathrm{E}}\boldsymbol{f}^{\mathrm{E}} + \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{\mathrm{C}}, \end{aligned} \tag{7.14}$$

where

$$\boldsymbol{K}^{\mathrm{E}} = \boldsymbol{L}^{\mathrm{CE}} \left[\boldsymbol{L}^{\mathrm{EE}}\right]^{-1} \tag{7.15a}$$

$$\boldsymbol{K}^{\mathrm{C}} = \left[\boldsymbol{L}^{\mathrm{CC}} - \boldsymbol{L}^{\mathrm{CE}} \left[\boldsymbol{L}^{\mathrm{EE}}\right]^{-1} \boldsymbol{L}^{\mathrm{EC}}\right], \tag{7.15b}$$

and

$$\boldsymbol{f}^{w} = \boldsymbol{K}^{\mathrm{E}} \boldsymbol{f}^{\mathrm{E}}. \tag{7.16}$$

7.2.4 Displacement-controlled loading

If the external loads are instead applied in displacement control, $\mathcal{S}_T$ is null, so $\mathcal{S}_E = \mathcal{S}_U$. Hence, $\boldsymbol{u}^{\mathrm{E}}$ is a known function of time, and $\boldsymbol{f}^{\mathrm{E}}$ is unknown. In this case, inspection of (7.12b) reveals that it is already in the desired form, where

$$\boldsymbol{K}^{\mathrm{E}} = \boldsymbol{L}^{\mathrm{CE}} \tag{7.17a}$$

$$\boldsymbol{K}^{\mathrm{C}} = \boldsymbol{L}^{\mathrm{CC}}, \tag{7.17b}$$

and in this case

$$\boldsymbol{f}^{w} = \boldsymbol{K}^{\mathrm{E}} \boldsymbol{u}^{\mathrm{E}}. \tag{7.18}$$

Notice that not only is $\boldsymbol{f}^{w}$ influenced by whether the loads are applied in force or displacement control but so is the contact stiffness matrix, $\boldsymbol{K}^{\mathrm{C}}$. This is because $\boldsymbol{K}^{\mathrm{C}}$ is determined by replacing the prescribed loading conditions by the equivalent homogeneous conditions. Hence, when the loads are prescribed in force or displacement control, zero-force or zero-displacement conditions, respectively, are introduced into this problem.

For this reason, $\boldsymbol{K}^{\mathrm{C}}$ will be identical for all force-controlled problems, even if the nodes in $\mathcal{S}_T$ change (assuming $\mathcal{S}_C$ remains the same). This is because the zero-force condition is the same condition that applies to the nodes in $\mathcal{S}_I$ in the reduction procedure, and these conditions do not modify the system's stiffness. In contrast, modifying the nodes in $\mathcal{S}_U$ (while keeping the same nodes in $\mathcal{S}_C$) will modify $\boldsymbol{K}^{\mathrm{C}}$. This is because a

zero-displacement boundary condition is applied to each node in $\mathcal{S}_U$ when determining $\boldsymbol{K}^{\mathrm{C}}$, and this will of course modify the system's stiffness.

7.3 Contact between two elastic bodies

As mentioned in the previous section, $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ can be determined for contact problems between two elastic bodies by first performing the steps described in §7.2 for each body separately and then performing several steps to merge the results. Note that each body must have the same number of contact nodes, spaced in the same way, and arranged in the same order. In addition, the sign convention shown in Figure 7.2 should be adopted *for each body separately*.

Newton's third law requires that the forces along the contact interface are equal and opposite. Thus, with the sign convention mentioned above, this implies that $p_j^1 = p_j^2$, $q_j^1 = q_j^2$, and hence $\boldsymbol{f}_1^{\mathrm{C}} = \boldsymbol{f}_2^{\mathrm{C}}$, where the index '1' or '2' denotes the body to which the symbol applies. However, since these are identical, the index can be dropped on these terms. In addition, the sign convention of Figure 7.2(b) implies that the relative normal (opening) displacement between each contact node, j, is given by

$$w_j = w_j^1 + w_j^2, \tag{7.19a}$$

and the relative tangential (slip) displacement is given by

$$v_j = v_j^1 + v_j^2. \tag{7.19b}$$

Hence, the relative displacement vector is

$$\boldsymbol{u}^{\mathrm{C}} = \boldsymbol{u}_1^{\mathrm{C}} + \boldsymbol{u}_2^{\mathrm{C}}. \tag{7.20}$$

By applying the steps in §7.2 to each body separately, $\boldsymbol{K}_1^{\mathrm{C}}, \boldsymbol{K}_2^{\mathrm{C}}$ and $\boldsymbol{f}_1^{w}, \boldsymbol{f}_2^{w}$ can be determined for each body in contact with a rigid surface. Relations of the form of (7.3)

can then be written as

$$\boldsymbol{f}^{\mathrm{C}} = \boldsymbol{f}_1^w + \boldsymbol{K}_1^{\mathrm{C}} \boldsymbol{u}_1^{\mathrm{C}} \tag{7.21a}$$

$$\boldsymbol{f}^{\mathrm{C}} = \boldsymbol{f}_2^w + \boldsymbol{K}_2^{\mathrm{C}} \boldsymbol{u}_2^{\mathrm{C}} \tag{7.21b}$$

for these two separate contact problems.

The contact stiffness matrix corresponding to contact between these two bodies can be obtained using $\boldsymbol{K}_1^{\mathrm{C}}$, $\boldsymbol{K}_2^{\mathrm{C}}$ and $\boldsymbol{f}_1^w$, $\boldsymbol{f}_2^w$. However, the following procedure involves a matrix inversion of the contact stiffness matrix of one of the two bodies, so it is necessary for at least one of these to be non-singular and invertible: that is, one of the bodies must not have any rigid-body degrees of freedom. This is not usually problematic because in practice at least one of the two elastic bodies is usually externally supported. Thus, in the following calculations, it is assumed that body 2 does not have rigid-body degrees of freedom, so its contact stiffness matrix, $\boldsymbol{K}_2^{\mathrm{C}}$, is positive definite and invertible. However, if a problem is to be considered in which neither body is externally supported, then the loads for body 2 must be applied in displacement control (since this imposes zero-displacement boundary conditions when determining $\boldsymbol{K}_2^{\mathrm{C}}$ and makes it positive definite).

There are several equivalent ways to obtain $\boldsymbol{K}^{\mathrm{C}}$ for contact between the two elastic bodies using $\boldsymbol{K}_1^{\mathrm{C}}$, $\boldsymbol{K}_2^{\mathrm{C}}$ and $\boldsymbol{f}_1^w$, $\boldsymbol{f}_2^w$. However, the approach used here begins by solving (7.21b) for $\boldsymbol{u}_2^{\mathrm{C}}$, which gives

$$\boldsymbol{u}_2^{\mathrm{C}} = \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} \left(\boldsymbol{f}^{\mathrm{C}} - \boldsymbol{f}_2^w\right). \tag{7.22}$$

This expression for $\boldsymbol{u}_2^{\mathrm{C}}$ is then substituted into (7.20), and the resulting expression is solved for $\boldsymbol{u}_1^{\mathrm{C}}$ as

$$\boldsymbol{u}_1^{\mathrm{C}} = \boldsymbol{u}^{\mathrm{C}} - \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} \left(\boldsymbol{f}^{\mathrm{C}} - \boldsymbol{f}_2^w\right). \tag{7.23}$$

Substituting (7.23) into (7.21a) and rearranging the resulting expression gives

$$\boldsymbol{f}^{\mathrm{C}} = \left[\boldsymbol{K}_1^{\mathrm{C}} \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} + \boldsymbol{I}\right]^{-1} \left[\boldsymbol{K}_1^{\mathrm{C}} \boldsymbol{u}^{\mathrm{C}} + \boldsymbol{f}_1^{w} + \boldsymbol{K}_1^{\mathrm{C}} \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} \boldsymbol{f}_2^{w}\right]. \tag{7.24}$$

Hence, this expression is in the form of (7.3), where $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ are defined as

$$\boldsymbol{K}^{\mathrm{C}} = \left[\boldsymbol{K}_1^{\mathrm{C}} \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} + \boldsymbol{I}\right]^{-1} \boldsymbol{K}_1^{\mathrm{C}} \tag{7.25}$$

$$\boldsymbol{f}^{w} = \left[\boldsymbol{K}_1^{\mathrm{C}} \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} + \boldsymbol{I}\right]^{-1} \left(\boldsymbol{f}_1^{w} + \boldsymbol{K}_1^{\mathrm{C}} \left[\boldsymbol{K}_2^{\mathrm{C}}\right]^{-1} \boldsymbol{f}_2^{w}\right). \tag{7.26}$$

Notice that if body 1 has rigid-body degrees of freedom, $\boldsymbol{K}_1^{\mathrm{C}}$ will be singular. But since (7.26) only requires the inversion of $\boldsymbol{K}_2^{\mathrm{C}}$, the required operations are still well defined. However, the resulting $\boldsymbol{K}^{\mathrm{C}}$ will be singular and will have the same rigid-body modes as $\boldsymbol{K}_1^{\mathrm{C}}$.

7.4 Partitioning the contact stiffness matrix

The previous sections specify how to obtain $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ for various types of contact problems. For the two-dimensional problems considered here, $\boldsymbol{K}^{\mathrm{C}}$ will be a $2N \times 2N$ matrix and $\boldsymbol{f}^{w}$ will be a vector of length $2N$. However, due to the definitions given in (7.1) and (7.2), the elements of $\boldsymbol{f}^{\mathrm{C}}$ and $\boldsymbol{u}^{\mathrm{C}}$ are grouped in nodal pairs, such that

$$\boldsymbol{f}^{\mathrm{C}} = \{q_1, p_1, q_2, p_2, q_3, p_3, ..., q_N, p_N\}^{\mathrm{T}} \tag{7.27a}$$

$$\boldsymbol{u}^{\mathrm{C}} = \{v_1, w_1, v_2, w_2, v_3, w_3, ..., v_N, w_N\}^{\mathrm{T}}. \tag{7.27b}$$

In many situations, it is more convenient to group the elements of these vectors into normal and tangential components as

$$\boldsymbol{u}^{\mathrm{C}} = \{\boldsymbol{v}, \boldsymbol{w}\}^{\mathrm{T}} \tag{7.28a}$$

$$\boldsymbol{f}^{\mathrm{C}} = \{\boldsymbol{q}, \boldsymbol{p}\}^{\mathrm{T}}, \tag{7.28b}$$

where

$$\boldsymbol{v} = \{v_1, v_2, v_3, ..., v_N\}^{\mathrm{T}} \tag{7.29a}$$

$$\boldsymbol{w} = \{w_1, w_2, w_3, ..., w_N\}^{\mathrm{T}} \tag{7.29b}$$

$$\boldsymbol{q} = \{q_1, q_2, q_3, ..., q_N\}^{\mathrm{T}} \tag{7.29c}$$

$$\boldsymbol{p} = \{p_1, p_2, p_3, ..., p_N\}^{\mathrm{T}} . \tag{7.29d}$$

Thus, $\boldsymbol{K}^{\mathrm{C}}$ must be rearranged and partitioned into sub-matrices, such that

$$\left\{ \begin{array}{c} \boldsymbol{q} \\ \boldsymbol{p} \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{v} \\ \boldsymbol{w} \end{array} \right\} + \left\{ \begin{array}{c} \boldsymbol{q}^w \\ \boldsymbol{p}^w \end{array} \right\}, \tag{7.30}$$

where

$$\boldsymbol{K}^{\mathrm{C}} = \left[\begin{array}{cc} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} \end{array} \right]. \tag{7.31}$$

Notice that since $\boldsymbol{K}^{\mathrm{C}}$ is a stiffness matrix, it must be symmetric. Therefore, $\boldsymbol{A}$ and $\boldsymbol{C}$ must also be symmetric, whereas $\boldsymbol{B}$ is not usually symmetric.

7.5 Evolution of the contact interface with time

After performing the static-reduction procedure, the contact interface becomes equivalent to a system of N massless blocks (i.e. nodes) connected with some general stiffness matrix. At each instant during the evolution of the contact interface with time, each of these nodes must be in one of the following states: stuck, slipping forwards, slipping backwards, or opening. In addition, each of these four states requires that several conditions must hold. These are:

$S_j = 1$	Stick	$w_j = 0$;	$\dot{v}_j = 0$;	$p_j \geqslant 0$; $\lvert q_j \rvert \leqslant f p_j$
$S_j = 2$	Separation	$p_j = 0$;	$q_j = 0$;	$w_j > 0$
$S_j = 3$	Forward slip	$w_j = 0$;	$q_j = -f p_j$;	$\dot{v}_j > 0$; $p_j \geq 0$
$S_j = 4$	Backward slip	$w_j = 0$;	$q_j = f p_j$;	$\dot{v}_j < 0$; $p_j \geq 0$,

where S_j is an integer state variable that specifies the contact state of node j.

These conditions illustrate that each of these four states demands that two equations are satisfied. Therefore, if the state of each node, $\boldsymbol{S}$, is known, two of the four nodal forces and displacements (i.e. q_j, p_j, v_j, w_j) are determined (assuming that the slip displacements, $\boldsymbol{v}$, from the previous instant in time are also known). With this information, the two unknown forces or displacements can then be calculated using (7.30).

7.5.1 Gauss-Seidel solution algorithm

To solve for the evolution of the contact interface with time, both the full time history of the external loads, $\boldsymbol{f}^w(t)$, and the contact stiffness matrix, $\boldsymbol{K}^{\mathrm{C}}$, must be known. However, once these have been determined, multiple algorithms can be used to solve this problem. For example, Ahn and Barber [2008] present a Gauss-Seidel algorithm that involves applying the conditions specified in §7.5 iteratively in each time-step. For this procedure, the initial displacement condition of all the nodes, $\boldsymbol{u}^{\mathrm{C}}(t_0)$, must be known, where t_0 denotes the time just prior to the onset of loading. The external loads, $\boldsymbol{f}^w(t_1)$, for the first time-step, t_1, are then applied.

This procedure begins by assuming that the nodal displacements are equal to their initial values, i.e. $\boldsymbol{u}^{\mathrm{C}}(t_1) = \boldsymbol{u}^{\mathrm{C}}(t_0)$, and that all the nodes are stuck, i.e. $S_j = 1$ for $j \in \{1, ..., N\}$. Each contact node is then examined *one at a time* to check whether the inequalities required for the assumed state are still satisfied (since the external loads have changed from the previous time-step). Since all the nodes are assumed to be stuck, these inequalities are $p_j \geqslant 0$ and $|q_j| \leqslant f p_j$. The algorithm begins at node $j = 1$, and checks these conditions. If they are satisfied, the state at this node remains the same, i.e. $S_1 = 1$, so the displacements remain unchanged, the nodal forces are recalculated, and the algorithm proceeds to the next node.

If one of these inequalities is violated at node j, this implies that the assumed state is incorrect: that is, the state at this node has changed from the previous time-step. Therefore, a new contact state must be assumed for this node. The particular inequality that is

violated suggests what the correct state is likely to be. For instance, if the node is assumed to be stuck, but the condition $p_j \geqslant 0$ is not satisfied, this suggests the node is actually in a state of separation, i.e. $S_j = 2$. On the other hand, if the condition $|q_j| \leqslant fp_j$ is violated, this suggests the node is actually slipping, i.e. $S_j = 2$ or $S_j = 3$, and the sign of the shear traction provides information on the slip direction. Hence, a new assumption on the node state is made based on which inequality is violated.

The algorithm then checks the inequalities associated with the newly assumed state for node j. If these inequalities are satisfied, the algorithm recalculates the contact forces and displacements and moves to the next node. If node j is assumed to be in state 2, 3, or 4, but one of the inequalities for the assumed state is violated, then the node is assumed to be stuck. The algorithm proceeds in this way until the correct state is found for every contact node. A convergence-check loop is then performed in which the forces and displacements are recalculated until the change in the contact displacements is less than a specified tolerance. Once this convergence criterion is met, the algorithm proceeds to the next time-step, t_2. In the next time-step, the external loads are updated to $\boldsymbol{f}^w(t_2)$, the nodal displacements, $\boldsymbol{v}, \boldsymbol{w}$, and states, $\boldsymbol{S}$, are assumed to be the same as at the end of the previous time-step, and this process is repeated.

Note that further details on this algorithm, including a block diagram of its operation, are given in Ahn and Barber [2008]. In addition, another common algorithm, which poses the evolution of the system with time as a linear-complementarity problem, is discussed by Bertocchi [2009].

7.6 Incomplete contact problems

The calculations described thus far apply to complete contacts (e.g. Figure 7.1) for which no initial gap is present between the bodies in the absence of external loads. For these contacts, the contact area cannot increase in size, so the selection of the nodes that are in $\mathcal{S}_C$ is simply defined by the contact geometry. However, for incomplete contacts, such as that shown in Figure 7.3, there is a gap between the bodies in the absence of

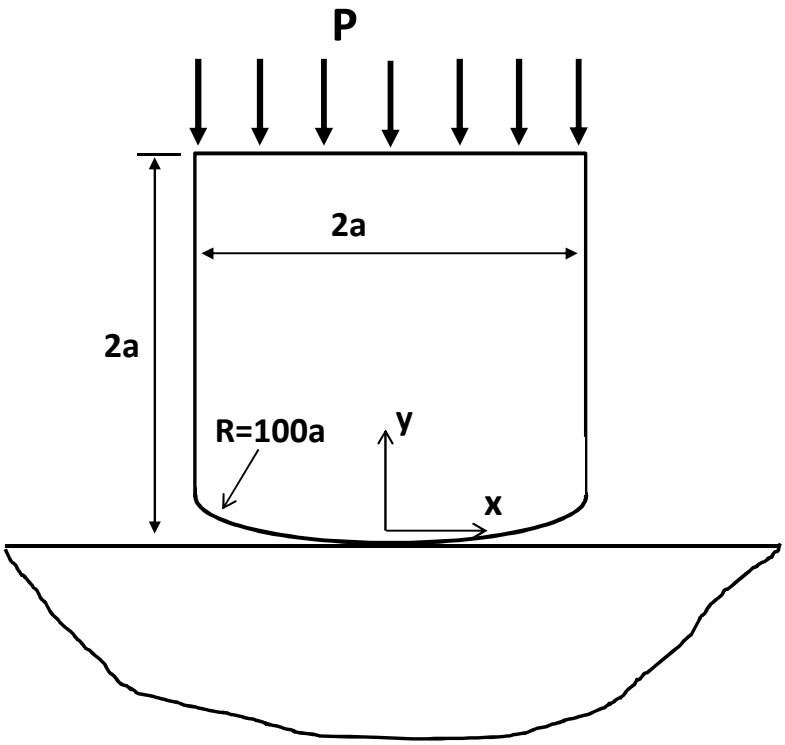


Figure 7.3: Diagram of an incomplete contact.

external loads, and the contact area increases as the normal load is applied.

The first step when performing static reduction on an incomplete contact is to define the set $\mathcal{S}_C$ as the maximum *potential* contact area, which comprises all the nodes that might come into contact under the expected loading regime. The static-reduction procedure can then be performed exactly as described in §7.2-§7.3. However, the procedure described in §7.4 for partitioning $\boldsymbol{K}^{\mathrm{C}}$ into normal and tangential components may need to be modified depending on the nature of the initial gap, $g(s)$, where s is a spatial coordinate defined along the potential contact area. Specifically, no modifications must be made to the procedure if the curvature of the contact interface is small. More precisely, the steps in §7.4 can be applied without modification when: (i) the initial gap, $g(s)$, is small in comparison to the contact area, and (ii) its derivative is also small (i.e. $dg/ds \ll 1$).

The incomplete contact shown in Figure 7.3 is an example for which the above two conditions hold. However, if the interface were highly curved, the above approximation would not be valid, so a rotational transformation would need to be applied to the contact

stiffness matrix to obtain the local normal and tangential coordinates at each node along the interface. Such a transformation is applied by Liu et al. [2014] in their analysis of an automotive brake drum, but this calculation is not discussed further here.

7.6.1 Modification of the Gauss-Seidel algorithm

If the solution algorithm specified in §7.5 is to be applied to incomplete contact problems, it must first be modified slightly. This is because the conditions given in §7.5 state that contact is implied at node j when $w_j = 0$, whereas separation is implied when $w_i > 0$. But these conditions are only applicable to the special case when there is no initial gap between the bodies, i.e. $\boldsymbol{g} = \boldsymbol{0}$. For incomplete contacts, contact is implied when the normal displacement compensates for the initial gap, i.e. when $g_j + w_j = 0$.

Thus, the contact conditions must be modified to account for the initial gap, which gives

$$
\begin{array}{lllll}
S_j = 1 & \text{Stick} & w_j + g_j = 0 \; ; & \dot{v}_j = 0 \; ; & p_j \geq 0 \; ; \;\; |q_j| \leq f p_j \\
S_j = 2 & \text{Separation} & p_j = 0 \; ; & q_j = 0 \; ; & w_j + g_j > 0 \\
S_j = 3 & \text{Forward slip} & w_j + g_j = 0 \; ; & q_j = -f p_j \; ; & \dot{v}_j > 0 \; ; \;\; p_j \geq 0 \\
S_j = 4 & \text{Backward slip} & w_j + g_j = 0 \; ; & q_j = f p_j \; ; & \dot{v}_j < 0 \; ; \;\; p_j \geq 0 \; .
\end{array}
$$

Once the solution algorithm is modified in this way, the contact stiffness matrix, $\boldsymbol{K}^{\mathrm{C}}$, and the external loading vector, $\boldsymbol{f}^w$, can be used to solve the evolutionary contact problem as before.

7.6.2 Modification of the loading vector

If the user would like to avoid modifying the solution algorithm as described above, an alternative approach can be used, which requires the external loading vector, $\boldsymbol{f}^w$, to be modified. Recall that for the complete contact problem, $\boldsymbol{f}^w$ is defined as the reaction forces generated at the contact nodes due to the applied loads when the relative contact displacements are constrained to be zero, i.e. $\boldsymbol{u}^{\mathrm{C}} = \boldsymbol{0}$. For the incomplete contact problem, $\boldsymbol{f}^w$ can be redefined as the contact forces required to cause all the nodes in $\mathcal{S}_C$ to be

in contact. These forces are given by the superposition of:

1. $\boldsymbol{f}^{w0}$: The forces that would be generated by the external loads if there had been no initial gap, i.e. if $\boldsymbol{g} = \boldsymbol{0}$.

2. $\boldsymbol{f}^{wg}$: The forces that are required to close the initial gap in the absence of external loads.

Hence, $\boldsymbol{f}^{w}$ is now given by

$$\boldsymbol{f}^{w} = \boldsymbol{f}^{w0} + \boldsymbol{f}^{wg}, \tag{7.32}$$

where $\boldsymbol{f}^{w0}$ can be calculated exactly as for the complete contact problem, and $\boldsymbol{f}^{wg}$ can be calculated using (7.30) as

$$\boldsymbol{f}^{wg} \equiv \left\{ \begin{array}{c} \boldsymbol{q^{wg}} \\ \boldsymbol{p^{wg}} \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{0} \\ \boldsymbol{-g} \end{array} \right\}. \tag{7.33}$$

To show that modifying the solution algorithm is equivalent to modifying the external loading vector, recall that when the modified solution algorithm is used, the relationship between the contact forces and displacements is given by (7.3) as

$$\boldsymbol{f}^{\mathrm{C}} = \boldsymbol{f}^{w} + \boldsymbol{K}^{\mathrm{C}} \boldsymbol{u}^{\mathrm{C}},$$

and contact occurs when $w_i + g_i = 0$. Now consider a displacement vector, $\boldsymbol{u}^*$, which is defined as

$$\boldsymbol{u}^* = \boldsymbol{u}^{\mathrm{C}} + \boldsymbol{u}^{g}, \tag{7.34}$$

where $\boldsymbol{u}^{\mathrm{C}}$ is the nodal displacement excluding the initial gap, and $\boldsymbol{u}^{g}$ is the nodal displacement corresponding to the initial gap, which is defined as

$$\boldsymbol{u}^{g} = \left\{ \begin{array}{c} \boldsymbol{v} \\ \boldsymbol{w} \end{array} \right\} = \left\{ \begin{array}{c} \boldsymbol{0} \\ \boldsymbol{-g} \end{array} \right\}. \tag{7.35}$$

Thus, (7.32) can be written as

$$\boldsymbol{f}^{w} = \boldsymbol{f}^{w0} + \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{g}. \tag{7.36}$$

This modified external loading vector can be substituted into (7.3) as

$$\begin{aligned}\boldsymbol{f}^{\mathrm{C}} &= (\boldsymbol{f}^{w0} + \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{g}) + \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{\mathrm{C}} \\ &= \boldsymbol{f}^{w0} + \boldsymbol{K}^{\mathrm{C}}(\boldsymbol{u}^{\mathrm{C}} + \boldsymbol{u}^{g}) \\ &= \boldsymbol{f}^{w0} + \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{*}. \end{aligned} \tag{7.37}$$

In this case, contact occurs when $w_i = 0$, which corresponds to $w_i^* = -g_i$. Thus, the gap can either be accounted for in the solution algorithm or in the external loading vector.

7.7 Reduction in simulation run-time

Thaitirarot et al. [2014] have shown that effectively identical results are obtained for the transient evolution of the contact interface by using $\boldsymbol{K}^{\mathrm{C}}$ and $\boldsymbol{f}^{w}$ with the Gauss-Seidel solution algorithm implemented in MATLAB or the full finite element model with contact elements in ABAQUS/CAE. This was demonstrated with three example problems: (i) the complete contact problem shown in Figure 7.1, (ii) complete contact between an elastic square and an elastically similar half-plane, and (iii) the incomplete contact problem shown in Figure 7.3. In addition, the solution was obtained 60, 65, and 40 times more quickly for problems (i), (ii), and (iii), respectively, by using MATLAB rather than ABAQUS.

It is noted that the loading regime for the two complete contact problems consisted of first applying a normal load, P, and then monotonically applying a shear load, Q, which was statically equivalent to a shear passing through the contact interface. Conversely, for the incomplete contact problem, only a normal load, P, was monotonically applied. It is noted that the contact interface incorporated 101 nodes (and hence 202 degrees of

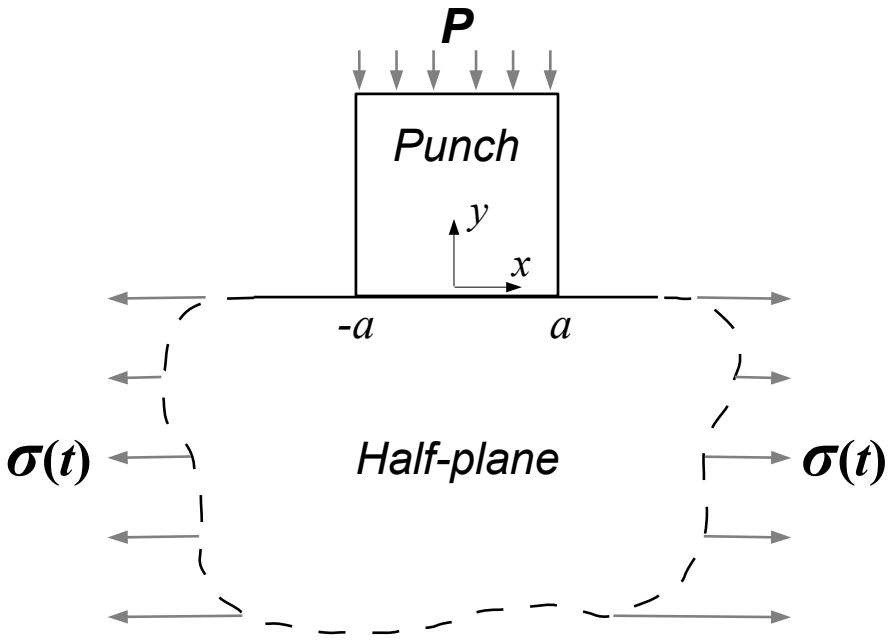


Figure 7.4: Complete contact between a square elastic punch and an elastically similar half-plane showing: the static normal load, P; the time-varying load, $\sigma(t)$; the size of the contact, $2a$; and the (x, y) coordinate set.

freedom) in all three example problems, and the simulations performed in MATLAB generally finished in under 10 seconds (on the particular computer that was used).

This raises the question of whether this same dramatic improvement in run-time would be maintained for longer simulation times. This is because ABAQUS is a commercial software package with a large number of features, so it is likely that a large number of error checking and option selection operations are performed at the beginning of each simulation. Conversely, the solution algorithm implemented in MATLAB is simply interpreted, compiled, and run. Thus, it is likely that the overhead associated with each ABAQUS simulation is greater than that associated with each MATLAB simulation.

To investigate the influence of simulation duration on computational savings, several transient contact simulations were performed on the complete contact shown in Figure 7.4 using the cyclic loading regime shown in Figure 7.5. For these simulations, a null initial displacement condition was assumed, i.e. $\boldsymbol{v}(0) = \mathbf{0}$. The normal load, P, was then monotonically applied and subsequently was held constant in time. Finally, the bulk load, $\sigma(t)$, was cycled in tension within the range $0 \leqslant \sigma(t) \leqslant \sigma^{\max}$ for as many cycles as was

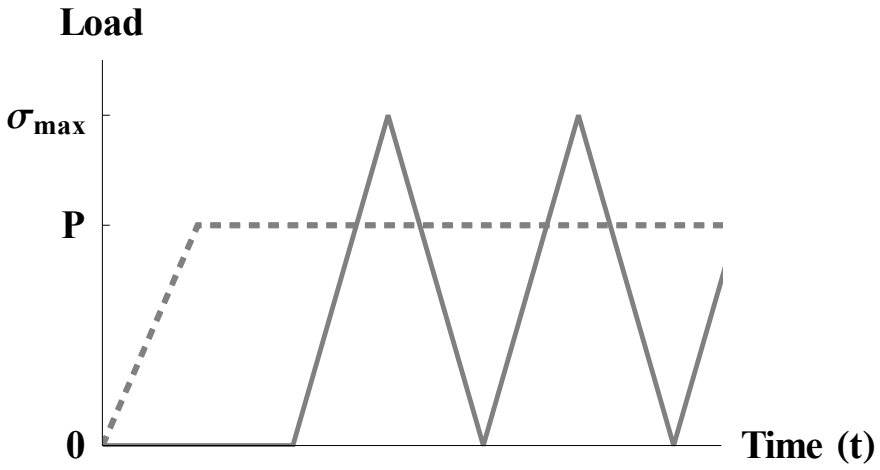
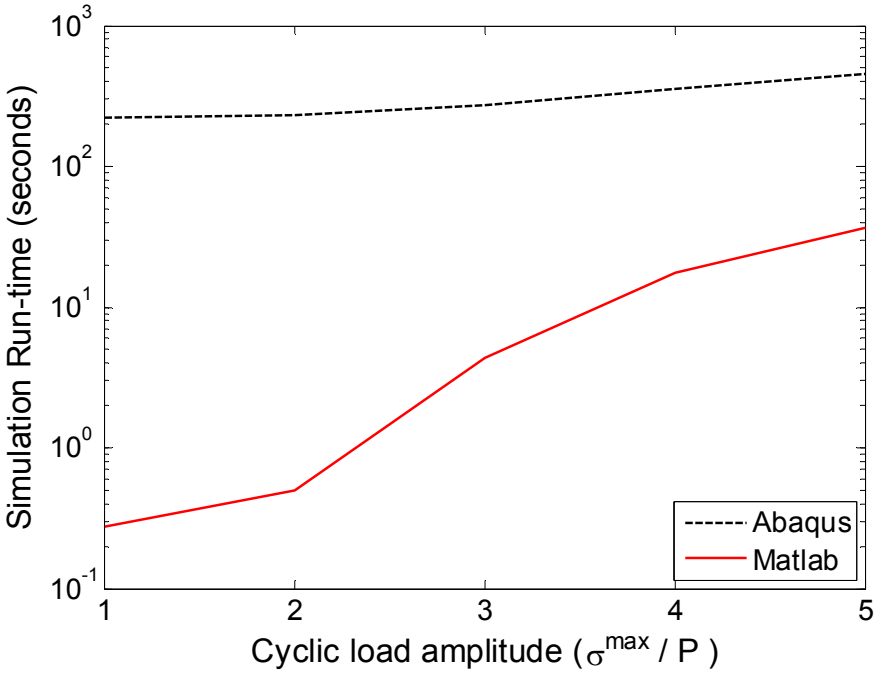


Figure 7.5: Cyclic loading regime applied to the complete contact shown in Figure 7.4.

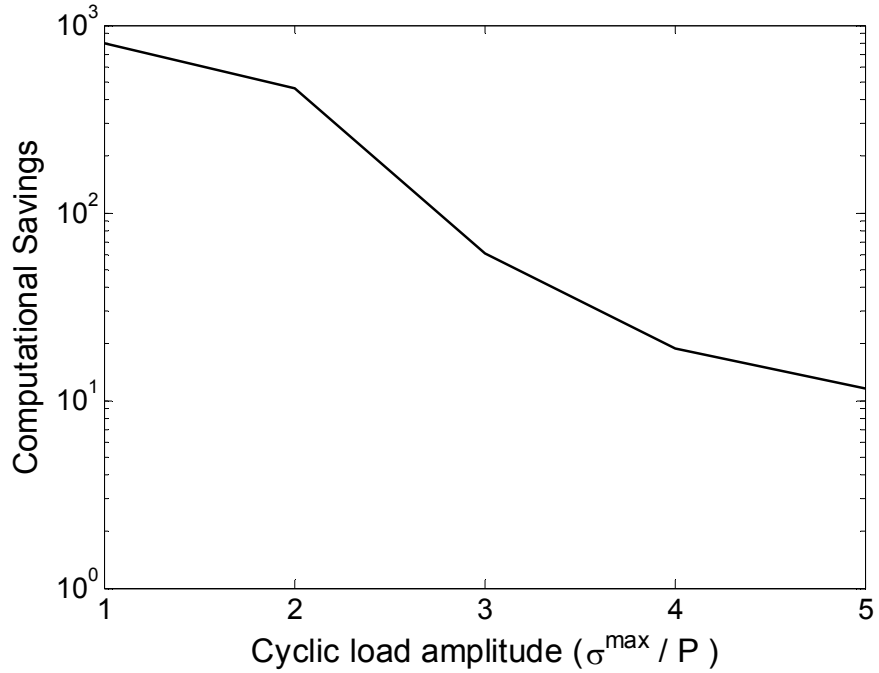
required for the contact to reach a steady-state response. See §9.5.3 for more details on the criteria used to determine when a steady state has been achieved.

Details of the finite element model used for these simulations are provided in §A.3.1, but here it is noted that the model incorporated 128 uniformly spaced nodes along the contact interface. These simulations were run on a Dell Optiplex 780 with an Intel Q9550 2.83GHz quad-core processor and 8GB of RAM running the Window 7 Enterprise 64-bit operating system. Although the models were created in ABAQUS/CAE version 6.12-3, the ABAQUS simulations presented here were performed in ABAQUS/CAE version 6.11-1. The static-reduction procedure and Gauss-Seidel simulations were performed in MATLAB version R2013a.

Notice that as the magnitude of $\sigma^{\max}$ is increased, the number of load increments required for each load cycle also increases, hence so does the total time required to complete the simulation. Thus, the influence of simulation run-time on computational savings was investigated by performing simulations with ABAQUS and MATLAB at five levels of cyclic load, viz. $\sigma^{\max}/P = 1, 2, 3, 4, 5$, and comparing the simulation times. It is noted that effectively identical solutions were obtained using ABAQUS and MATLAB, which is in agreement with the results reported by Thaitirarot et al. [2014]. The actual simulation run-times (in seconds) are shown in Figure 7.6(a), and the computational savings, which



(a) Semi-log plot of simulation run-time (log-scale) plotted against cyclic load amplitude, $\sigma^{\max}/P$, for the full model in ABAQUS and the reduced model in MATLAB.



(b) Semi-log plot of computational savings (log-scale) plotted against cyclic load amplitude, $\sigma^{\max}/P$.

Figure 7.6: Comparison of the computational savings achieved by running transient frictional contact simulations in MATLAB vs. in ABAQUS.

are defined as

$$\text{Computational savings} = \frac{\text{ABAQUS run-time}}{\text{MATLAB run-time}} - 1, \tag{7.38}$$

are shown in Figure 7.6(b).

Figure 7.6(b) indicates that, as suspected, the improvement in run-time in MATLAB relative to ABAQUS decreases as the simulation time increases. For example, when $\sigma^{\max}/P = 1$, the solution was obtained approximately 800 times faster in MATLAB than in ABAQUS. However, Figure 7.6(a) reveals that the MATLAB simulation required only 0.3 seconds to complete, whereas the ABAQUS simulation required 221 seconds to complete. Conversely, when $\sigma^{\max}/P = 5$, the MATLAB simulation finished in 36 seconds, whereas the ABAQUS simulation finished in 458 seconds, corresponding to a computational savings factor of 11.7. In addition, when $\sigma^{\max}/P = 3$, the MATLAB simulation finished in approximately 4 seconds, giving a computational savings factor of approximately 61. As previously mentioned, the simulations performed by Thaitirarot et al. [2014] generally finished in under 10 seconds, and computational savings of approximately 60 were reported. Hence, the results shown in Figure 7.6 are consistent with those reported by Thaitirarot et al. [2014], but they also illustrate that the computational savings are heavily problem dependent.

In addition, Figure 7.6 supports the hypothesis that each ABAQUS simulation is associated with greater overhead. This is evidenced by the ABAQUS simulation time being less sensitive to increases in the bulk load amplitude, $\sigma^{\max}$, but requiring over 200 seconds to finish at even the lowest level of bulk load considered. However, even for the largest simulations considered, the solution was still obtained just over 10 times faster with MATLAB than with ABAQUS, which still represents very significant computational savings. Moreover, it is very likely that significant gains could be achieved by implementing the Gauss-Seidel algorithm in a faster programming environment, e.g. Fortran or C++, instead of in MATLAB. Thus, the static-reduction approach remains a very attractive option for solving quasi-static, linear-elastic, two-dimensional frictional contact problems.

7.8 Summary

In this chapter, a procedure for significantly reducing the number of degrees of freedom in finite element models of two-dimensional frictional contacts is discussed. In particular, the steps required to obtain the reduced contact stiffness matrix and a vector that accounts for the influence of the applied loads are provided. An algorithm developed by Ahn and Barber [2008] for solving the evolution of the contact interface with time is then discussed. The calculations presented in this chapter are first explained in the context of complete contact problems, but the modifications required to apply them to incomplete frictional contact problems are subsequently discussed.

Finally, the computational savings that can be achieved by solving the evolutionary contact problem using the static-reduction procedure instead of using the full finite element model is investigated. It is found that the computational savings are strongly influenced by the size of the problem under consideration. However, the simulations performed using the reduced model with the solution algorithm developed by Ahn and Barber [2008] (and implemented in MATLAB) finished at least 10 times faster than the same model run in ABAQUS for all the problems that were considered.

In addition, performing simulations with the reduced model using the solution algorithm developed by Ahn and Barber [2008] provides the user with greater control over how the frictional law is implemented. Furthermore, the reduced contact stiffness matrix has several other applications, such as calculating the shakedown limit of cyclically loaded frictional contacts, which is examined in §9.

Chapter 8

The transition from normal load to sliding

In this chapter, the static-reduction technique described in §7 is used to study the complete contact formed between a square elastic block and an elastically similar half-plane, which is subjected to a monotonically increasing shear load that is raised until sliding is achieved. Numerical results show that for moderate levels of shear force (i.e. less than 50% of the sliding condition) very different qualitative behaviour is observed for high and low coefficients of friction. Conversely, for high shear forces (i.e. near to the sliding condition), the response for high and low friction coefficients is similar.

Note that parts of this chapter have been published in [Flicek et al., 2015c].

8.1 Introduction

In this chapter, the static-reduction technique described in §7 is used to investigate the complete frictional contact formed between a square elastic block and an elastically similar half-plane, which is shown in Figure 8.1. Here, a quasi-static loading regime is considered in which the normal load, P, is first applied to form the contact, and then the shear load, Q, is monotonically increased until sliding is achieved. Although an elasticity formulation for a rectangular contact with simple boundary conditions (e.g. constant normal displacement and a frictionless contact interface) is feasible (see [Khadem and O'Connor, 1969a,b]), such an approach is not possible with the complex boundary con-

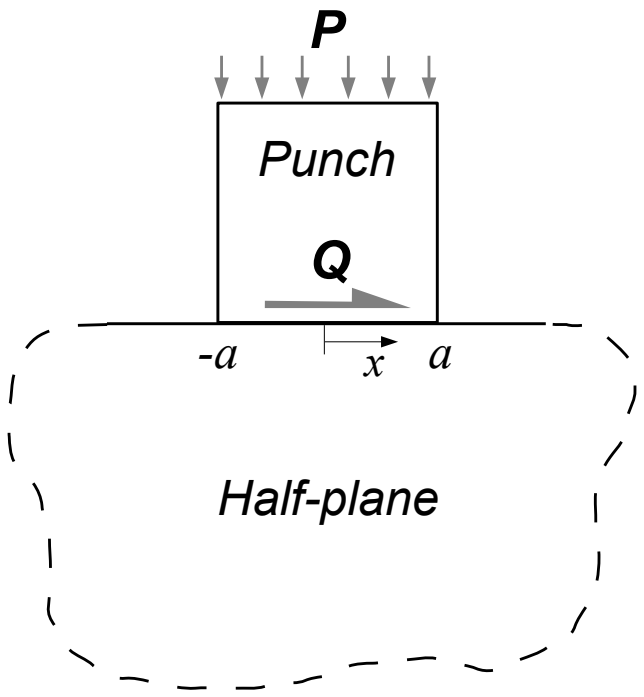


Figure 8.1: Diagram of complete contact between a square punch and an elastically similar half-plane.

ditions considered here, i.e. a frictional interface and a more complex loading regime. Thus, this problem must be solved numerically.

This problem has been studied in some detail by Churchman and Hills [2006b] and later by Karuppanan and Hills [2008a]. The approach adopted by these authors was to use a combination of the finite element method and an asymptotic analysis of the contact edges. Churchman and Hills [2006b] investigated the implications of Williams' solution (see §2) for the contact shown in Figure 8.1, and derived the behaviour arising on the initial application of the normal load, P. Making use of a result due to Hills and Dini [2004], Churchman and Hills [2006b] demonstrated that the contact remains fully adhered on the application of the normal load if the coefficient of friction, f, is greater than $g^I_{r\theta} = 0.543$ (see §4.2.1 and §4.3); otherwise both contact edges slip outwards.

Finite element results obtained by Churchman and Hills [2006b] revealed that when $f > g^I_{r\theta} = 0.543$, slip does not initiate until a shear force of approximately 55% of that required to cause full sliding is applied and that slip initiates from the interior of

the contact in these cases. Because slip initiates from the interior of the contact, corner asymptotes provide no information on their evolution, so a purely numerical approach must be used. On the other hand, when slip initiates from the contact edge, the asymptotic solution for a wedge sliding on a half-plane developed by Gdoutos and Theocaris [1975] and Comninou [1976] can be employed.

Karuppanan and Hills [2008b] investigated the implications of the sliding asymptote for complete contacts and discussed the numerous contact-edge states that may occur depending on the contact-edge angle and the coefficient of friction. Karuppanan et al. [2008] also presented more detailed results on contact-edge behaviour for a square block sliding on a half-plane. The results for this contact-edge angle revealed that the tractions at the *leading edge* (i.e. the right edge in Figure 8.1) are always power-order singular, but that three types of contact-edge behaviour are possible at the *trailing edge*, depending on the coefficient of friction. These are:

1. If $f < 1/\pi$, the tractions are power-order singular, but note that the order of singularity is weaker than at the leading edge.

2. If $1/\pi < f < 0.392$, the trailing edge experiences a power-order bounded response but is in intimate contact.

3. If $0.392 < f$, the trailing edge separates, and the contact pressure is locally square-root bounded.

Thus, if the coefficient of friction is sufficiently large (i.e. if $f > g^I_{r\theta} = 0.543$), Williams' solution can be used to model contact-edge behaviour on the initial application of the normal load and up to moderate values of shear load. Conversely, the behaviour at high shear loads near the sliding condition is accurately described by the formulation due to Gdoutos and Theocaris [1975] and Comninou [1976]. The transition between these two states, however, must be tackled numerically. Churchman and Hills [2006b] investigated this transition problem by plotting out the qualitative response to be expected on a plot of the load ratio (i.e. Q/fP) vs. the coefficient of friction. Additional refinements to

this plot were later presented by Karuppanan and Hills [2008a], and these authors also examined complete contacts with punch angles of $\phi = 60°$ and $\phi = 120°$ in this way.

In this chapter, the evolution of contact behaviour during the transition from full stick to sliding is examined in detail by focusing on two particular example problems: (i) a high-friction case when $f = 0.8$, and (ii) a low-friction case when $f = 0.3$. These example values have been chosen because the asymptotic analysis described above suggests that qualitatively different behaviour should be expected for coefficients of friction above and below 0.543. But aside from this consideration, the precise values that have been chosen are somewhat arbitrary.

As this problem must be solved numerically, the static-reduction procedure described in §7 is first employed, and the evolutionary contact problem is then solved using the Gauss-Seidel algorithm described in §7.5.1, which is due to Ahn and Barber [2008]. Recall from §7 that a benefit of this approach is that there is no loss of accuracy associated with the introduction of an artificial interfacial tangential compliance, which would result if the penalty or augmented Lagrange multiplier method were used. In addition, the results are obtained over 10 times faster than with a standard finite element approach (as is shown in §7.7). As with those of Churchman and Hills [2006b] and Karuppanan and Hills [2008a], the solution obtained for this problem is geometry specific, so the results cannot be easily carried over to other geometries. Nevertheless, the general form of behaviour is likely to remain similar for contacts of similar contact-edge angle (i.e. $\phi \sim 90°$), especially in the sense that a qualitatively different response will be obtained when $f > g^I_{r\theta}$ and when $f < g^I_{r\theta}$.

8.2 Finite element model

Separate finite element models were created for the half-plane (modelled of course as a large, but finite, block) and the square punch using ABAQUS/CAE version 6.12-3.[1]

[1]Both the punch and the half-plane were modelled as two-dimensional, plane-strain geometries, which incorporated a combination of `CPE4R` and `CPE3` elements. The square punch was modelled with no external boundary conditions, and each of its edges were of length $2a$. The half-plane was approximated as a

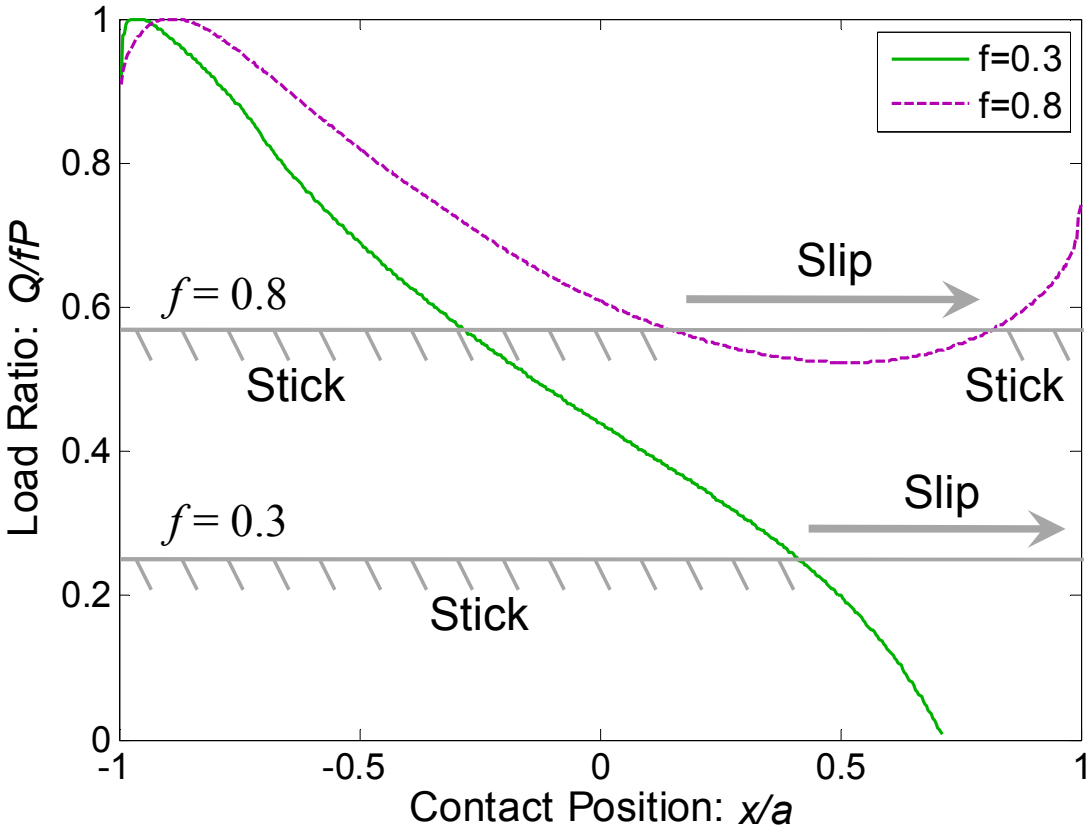


Figure 8.2: Evolution of contact behaviour during the transition from normal load to sliding for two coefficients of friction: $f = 0.3, 0.8$.

The global stiffness matrices were extracted from these models, and all subsequent calculations were performed in MATLAB R2013a (including both the static reduction and the solution of the contact problem itself). The half-plane's full stiffness matrix had 408,062 degrees of freedom, and the square punch's had 181,608 degrees of freedom. It is noted that the contact interface incorporated 512 uniformly spaced contact nodes (corresponding to 1024 degrees of freedom), such that the size of each element was $0.0039a$, where a is the half-width of the contact (see Figure 8.1). By performing the static-reduction procedure, the *reduced contact stiffness* matrix was obtained, which retained only the 1024 contact degrees of freedom. It is also noted that the model used a Young's modulus of $E = 200GPa$ and a Poisson's ratio of 0.3.

large rectangular block of width $20a$ and height $10a$, which had two boundary conditions applied to it: (i) the displacement in the vertical direction was constrained to be zero along the entirety of its bottom edge, and (ii) the displacement in the horizontal direction was constrained to be zero at one node at the centre of its bottom edge.

The normal load, P, was applied as a uniform normal traction along the top of the punch as shown in Figure 8.1; hence, it has units of $[FL^{-2}]$, where F denotes force and L denotes length. The shear load, Q, which also has units of $[FL^{-2}]$, was applied so as to be statically equivalent to a shear applied through the plane of the contact interface. To do this, a uniform shear traction was applied along the top of the punch, and, at the same time, a linearly varying normal traction was applied to counteract the resulting 'tipping' moment. This approach is precisely correct if the contact pressure distribution produces no moment about the contact centreline. However, as this contact is highly coupled (in the sense that tangential slip displacements modify the contact pressure distribution), the contact pressure distribution produces such a moment once the interface begins to slip.

8.3 Results

Figure 8.2 shows the evolution of the zones of slip and stick as the shear load, Q, is monotonically raised until sliding is achieved (i.e. $Q/fP = 1$) for the two coefficients of friction considered here, viz. $f = 0.3$ (solid line) and $f = 0.8$ (dashed line). These results are displayed on a plot of the load ratio, Q/fP, vs. the position along the contact interface, x/a. The curves in this figure mark transitions between zones of slip and stick. Two horizontal lines are included in this figure (one for each coefficient of friction that is considered), which are annotated to show the pattern of contact behaviour implied for an example load ratio. Note that the evolution of the slip zones is of course history dependent, so this particular evolution of contact behaviour only results when the shear is monotonically increased. The detailed response of the high-friction ($f = 0.8$) and low-friction ($f = 0.3$) cases are examined separately below.

8.3.1 High friction

As expected, Figure 8.2 shows that when $f = 0.8$, the contact remains fully adhered on the application of P and remains stuck until $Q/fP = 0.52$. This is in quite good agreement with the 'rule of thumb' suggested by Churchman and Hills [2006b]: namely,

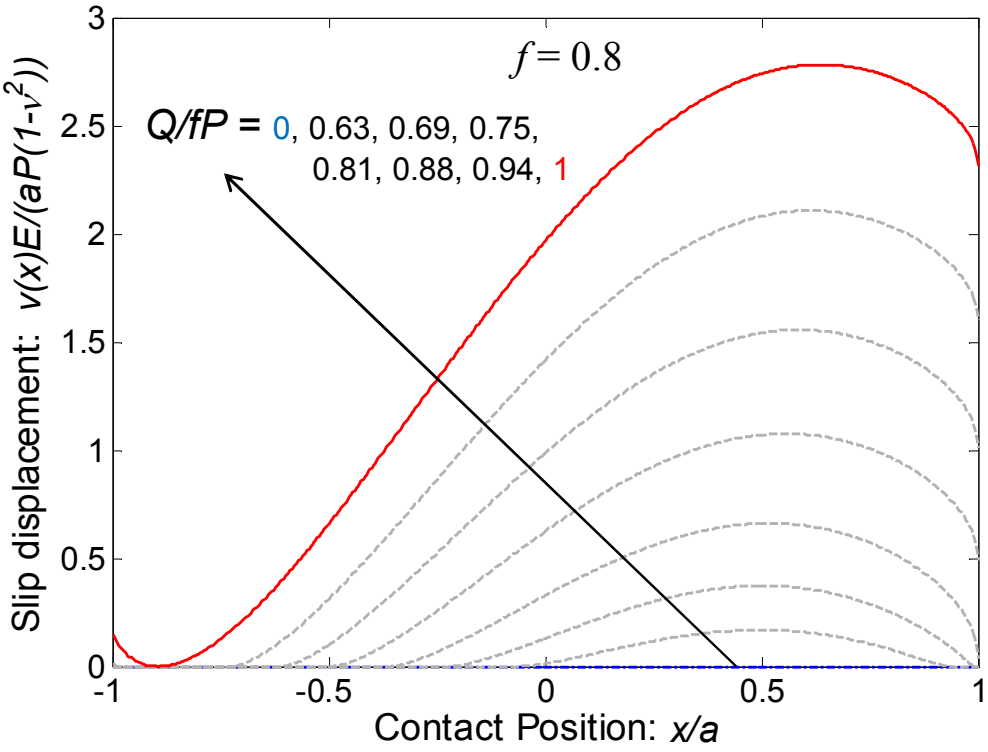
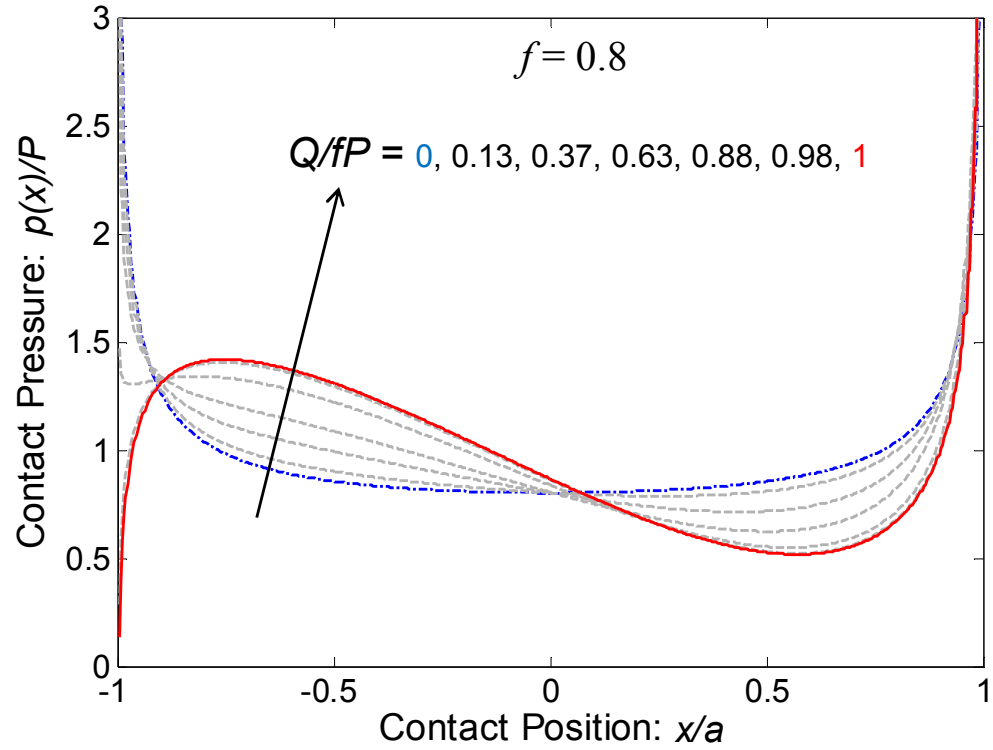


Figure 8.3: Evolution of the tangential slip displacement, $v(x)$, when $f = 0.8$.

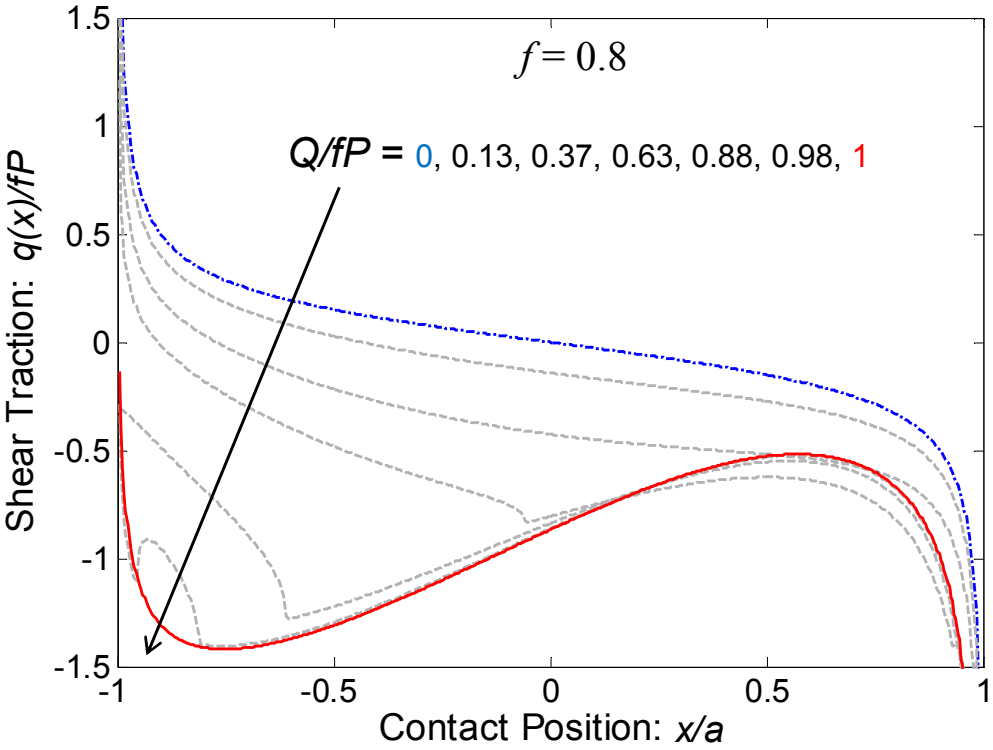
when $f > 0.543$, slip does not occur until the shear load is approximately 55% of that required to cause sliding. The position at which slip initiates is $x = 0.51a$, which is approximately midway between the leading edge and the contact centreline. This can also be seen in Figure 8.3, which shows the tangential slip displacement, $v(x)$, plotted against the position along the contact interface, x/a, at several load ratios, Q/fP, during the transition to sliding. Note that the slip displacement in this plot is normalized by $E/(aP(1-\nu^2))$.

Figures 8.2-8.3 show that as the shear force is increased further, the interior slip zone expands in both directions. As the slip zone initiated nearer to the leading edge than the trailing edge, it quickly reaches the leading edge when $Q/fP = 0.75$, i.e. at three quarters of the shear force needed to cause sliding. Progression of the slip zone towards the trailing edge is slower, so the problem becomes more complicated as the trailing edge (and therefore full sliding) is approached.

To help interpret what happens, the evolution of the contact pressure distribution,



(a) The contact pressure distribution, $p(x)$.



(b) The shear traction distribution, $q(x)$.

Figure 8.4: Evolution of the contact tractions when $f = 0.8$.

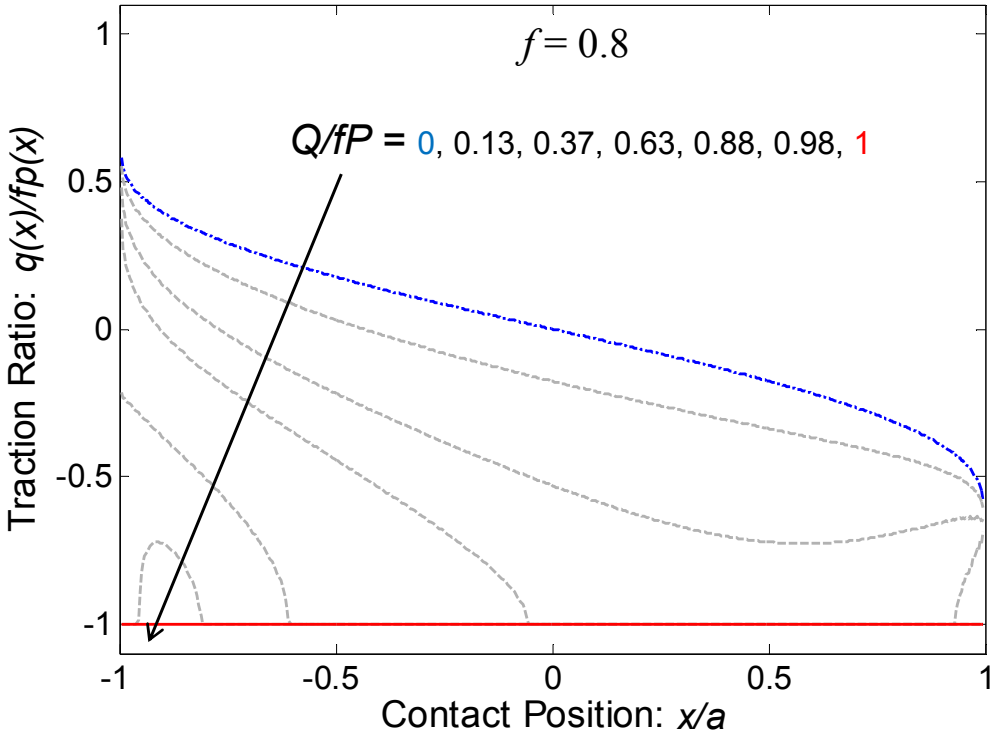


Figure 8.5: Evolution of the traction ratio, $q(x)/fp(x)$, when $f = 0.8$.

$p(x)$; the shear traction distribution, $q(x)$; and the traction ratio, $q(x)/fp(x)$, are shown in Figures 8.4-8.5. In these figures, the contact pressure distribution is normalized by the applied normal load, P, and the shear traction distribution is normalized by fP. Figure 8.4(a) shows that the contact pressure distribution becomes increasingly unsymmetrical at higher shear forces, which results in a tipping moment. In addition, Figures 8.2, 8.4(b), and 8.5 show that as the shear is increased further, a new slip zone slip initiates at the trailing edge of the contact, which occurs when $Q/fP = 0.91$.

By performing an asymptotic analysis of the trailing edge, Karuppanan et al. [2008] demonstrated that local separation must occur once sliding is achieved if $f > 0.392$ (i.e case 3 mentioned in §8.1), which means that the contact tractions in this region must be square-root bounded in form. Thus, although the numerical results suggest that slip initiates from the trailing edge in the absence of edge separation, this cannot actually be the case. Instead, a very small zone of separation must arise at the trailing edge adjacent to the slip zone, but its initial extent is too small to be captured by the finite element model.

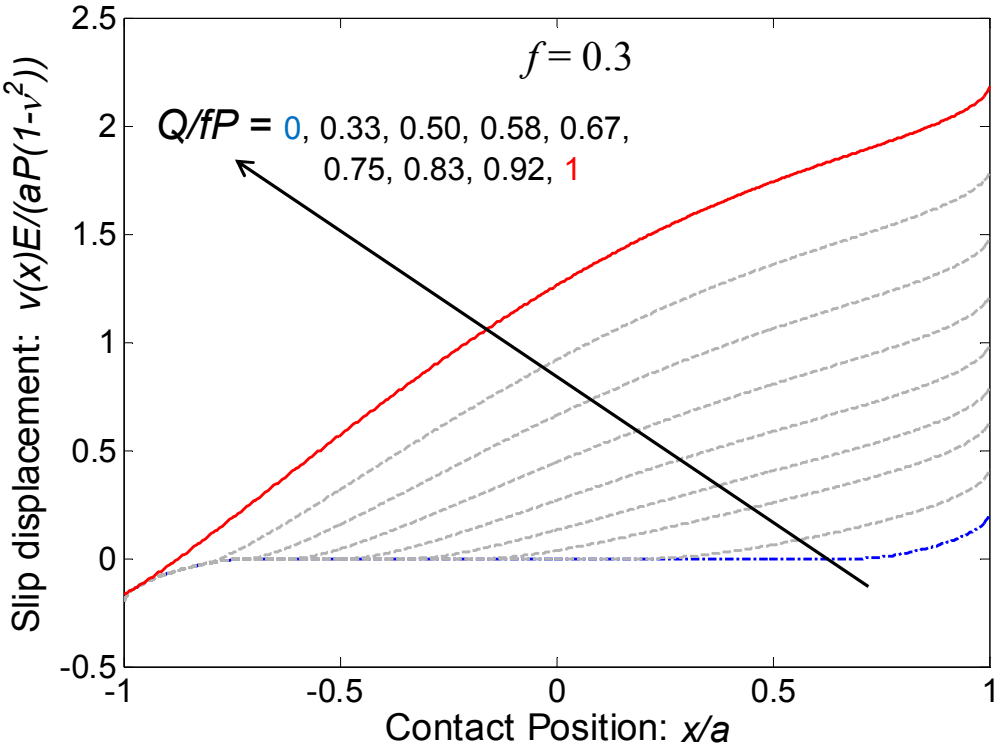
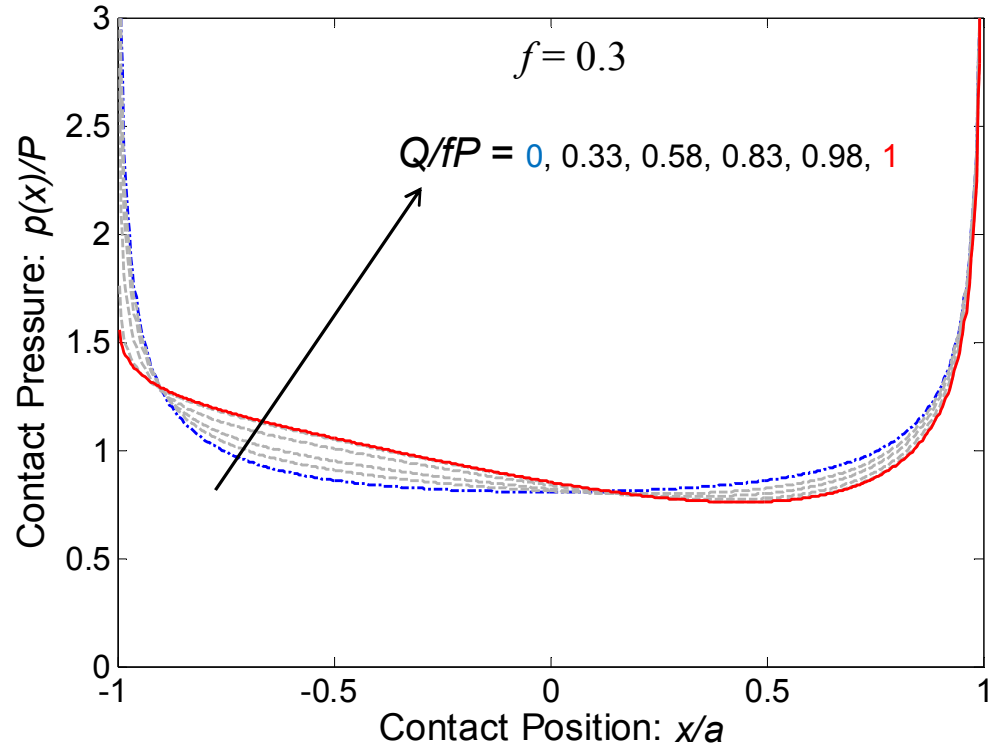


Figure 8.6: Evolution of the tangential slip displacement, $v(x)$, when $f = 0.3$.

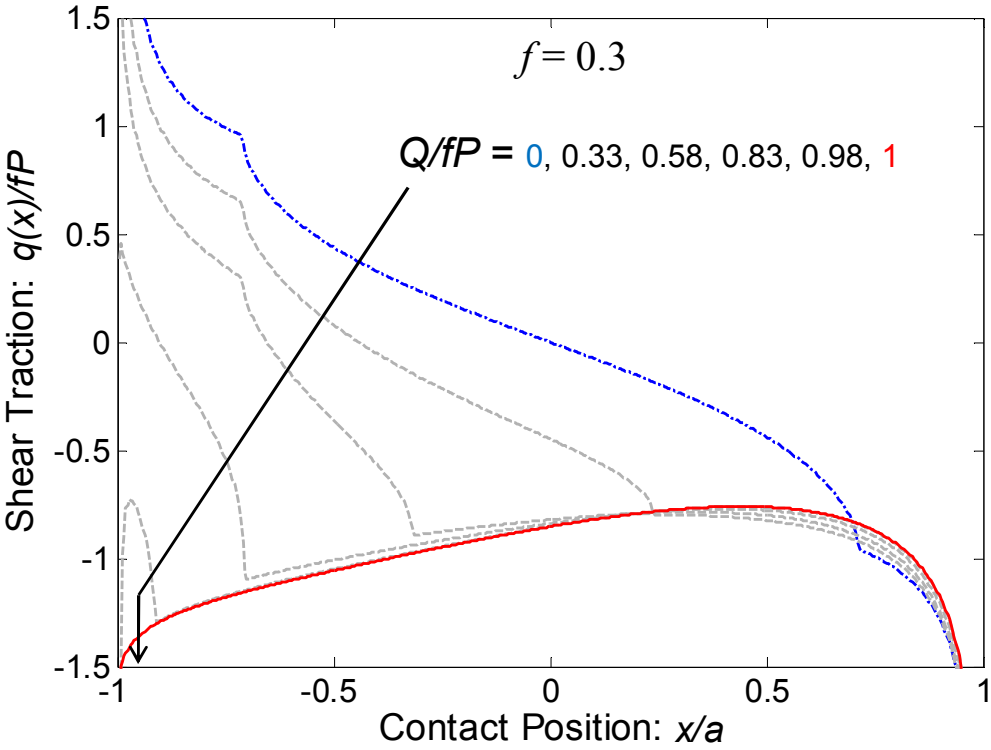
Figure 8.2 shows that further increases in the shear force cause the two slip zones to expand towards one another. Although this cannot be seen in these figures, the finite element results suggest that separation occurs at the trailing edge of contact just prior to reaching the sliding condition when $Q/fP = 0.98$, but this should simply be interpreted as the (already existing) separation zone becoming large enough to be detected by the finite element model. In fact, even when sliding is achieved, the separation zone only comprises one node at the trailing edge, which demonstrates that the separation extent is very small indeed. Finally, just before sliding is achieved, the last point to slip is at $x = -0.89a$, which is quite close to, but not at, the trailing edge.

8.3.2 Low friction

When $f = 0.3$, the coefficient of friction is insufficient to maintain adhesion at the contact edges under the application of the normal load alone (because $f < g^I_{r\theta} = 0.543$), so both contact edges initially tend to slip outwards. This behaviour is not shown in Fig-



(a) The contact pressure distribution, $p(x)$.



(b) The shear traction distribution, $q(x)$.

Figure 8.7: Evolution of the contact tractions when $f = 0.3$.

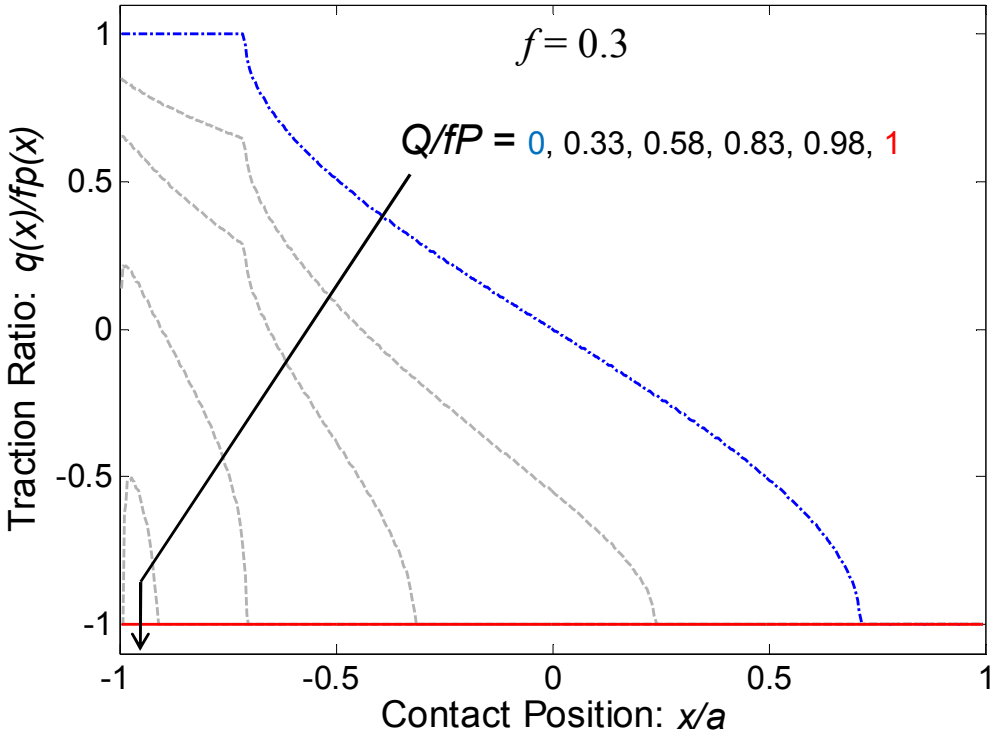


Figure 8.8: Evolution of the traction ratio, $q(x)/fp(x)$, when $f = 0.3$.

ure 8.2, but it can be seen in Figure 8.6, which shows the tangential slip displacement, $v(x)$, plotted against the position along the contact interface, x/a, at several load ratios, Q/fP, during the transition to sliding. This figure shows the antisymmetric slip distribution that arises when the normal load alone is applied (i.e. $Q/fP = 0$), where the slip zone at the trailing edge extends to $x = -0.71a$, and that at the leading edge extends to $x = 0.71a$. The extent of these initial slip lengths can also be seen by viewing the traction ratio at this load factor, which is shown in Figure 8.8. The contact pressure, $p(x)$, and the shear traction distribution, $q(x)$, are also shown in Figure 8.7 and are normalized as in Figure 8.4. Notice from Figure 8.8 that $q(x)/fp(x) = 1$ for the trailing-edge slip zone, and $q(x)/fp(x) = -1$ for the leading-edge slip zone since the corners initially slip in opposite directions.

When an infinitesimal shear load, Q, is first applied, the slip zone at the leading edge simply grows infinitesimally. Conversely, the increment of shear at the trailing edge is of the opposite sign to the sense of the slip displacement already present, so the trailing edge

adheres. Figures 8.2 and 8.6 show that as the shear force is increased, the leading-edge slip zone grows smoothly and extends towards the trailing edge. As this occurs, Figure 8.7(a) shows that the contact pressure distribution displays the same tendency to become unsymmetrical at higher shear forces, which results in a tipping moment in the same sense as for the high-friction case. As before, the slip zones at the contact edges cause the stress state to become more strongly singular at the leading edge and less strongly singular at the trailing edge. As $f = 0.3 < 1/\pi$, the response of the trailing edge remains power-order singular in form (i.e. case 1 from [Karuppanan et al., 2008] that is mentioned in §8.1).

Once a shear force of $Q/fP = 0.92$ is applied, Figures 8.2 and 8.8 show that slip initiates from the trailing edge. Churchman and Hills [2006b] showed that the *change* in tractions at the trailing edge associated with adhesion is always more strongly singular than the slipping distribution already present, leading to a tiny region of separation and attendant slip. This zone is of the order of $10^{-6}a$ in extent, so it was not captured by the finite element model. Of course, a slip zone will arise adjacent to this very small region of local separation at the trailing edge. Thus, as before, the onset of slip at the trailing edge should be interpreted as implying that both separation and slip arise. Finally, Figure 8.2 shows that the last point to slip is $x = -0.97a$, which is very near to, but not at, the trailing edge.

8.4 Discussion

The results presented in §8.3 demonstrate that when a low to moderate shear load is applied, there is a significant difference between the low-friction and high-friction cases. In fact, when the shear is at 50% of the sliding condition, approximately half of the interface is slipping when $f = 0.3$, whereas slip does not occur at this load level when $f = 0.8$. However, the responses become ever more similar as sliding is approached. One notable difference between the low-friction and high-friction cases as the sliding condition is approached can be seen in the contact pressure distributions shown in Figures 8.4(a) and 8.7(a). These figures show that the trailing edge experiences a singular contact pres-

sure for the low-friction case, whereas the contact pressure is locally square-root bounded for the high-friction case (see [Karuppanan et al., 2008]).

Some more general remarks can also be made regarding the response of elastically similar complete contacts based on the asymptotic results that have already been discussed. First, if the coefficient of friction is greater than that required to prevent edge slip on the application of the normal load, then slip cannot initiate from the contact edge unless the contact edge separates beforehand, although the separated region may be too small to resolve with a finite element model (see the trailing edge of the low-friction case in §8.3.2). However, slip that initiates from an interior point can propagate through to the contact edge (see the leading edge of the high-friction case in §8.3.1), and slip that initiates from the contact edge can grow (see the leading edge of the low-friction case in §8.3.2). Finally, irrespective of the coefficient of friction, slip cannot initiate from a trailing-edge corner without at least some small amount of separation occurring as well (see [Churchman and Hills, 2006b]), but in many cases this separated region is too small to be of practical importance.

It is perhaps worth emphasizing that in contrast to incomplete contact problems (for which coupling through elastic mismatch is light, and an exchange of elastic compliance between the two bodies produces only a small effect), the solution here depends very strongly on the distribution of the elastic stiffness between the half-pane and the elastic square. Making the square rigid, for example, will produced a very different pattern of behaviour.

8.4.1 Other considerations

It is worth taking a moment to examine the assumptions that are made in the analysis presented here. The key assumptions are: (i) the surfaces of both bodies are perfectly smooth and flat, (ii) the edges of the square are perfectly sharp, and (iii) both bodies are infinitely elastic. These assumptions greatly facilitate the numerical solution of the problem and also allow inferences to be drawn from elasticity solutions for semi-infinite

notches, which add a greater level of accuracy to the solution near the contact edges. Another assumption that is made is that the friction coefficient is constant (both spatially and temporally), though this assumption is not central to the formulation and is mainly adopted for the sake of simplicity.[2] The remainder of this section examines the conditions under which these assumptions are acceptable and provide a good description of contact behaviour.

Edge radius

The solution presented in §8.3 is probably most sensitive in the vicinity of the contact edges. As mentioned above, this solution was obtained assuming that the contact edges are perfectly sharp, which leads to the stresses being power-order singular near the contact edges. However, the contact edges will always be blunted to some degree in practice. Several authors have already investigated the influence of an edge radius on contact behaviour, e.g. [Ciavarella et al., 1998, Dini et al., 2004, Sackfield et al., 2002]. These analyses reveal that if a finite edge radius is present, the contact tractions must fall to zero local to the contact edge. However, provided that the radius is not too large, the singular solution will still dominate contact-edge behaviour.

As an example, consider Figures 8.4 and 8.7, which display the elastic stress singularities that arise at the contact edges for the sharp-edged case. These figures illustrate that the contact tractions tend to be very strongly influenced by the singular solution for at least (roughly) the 10% of the contact half-width, a, that is nearest to the contact edge (where the stress gradient is very steep). Suppose now that a very small edge radius, R, were present, such that $R \ll a$. In this case, as an observation point approaches the contact edge, the tractions would initially increase rapidly and would look very similar to those shown in Figures 8.4 and 8.7. But as the edge is approached even more closely, such that $x \to \pm a$, the stress gradient would eventually plateau and then very rapidly fall to zero at the contact edge. Thus, the edge radius controls behaviour at the very small scale

[2] On the other hand, the entire solution implicitly assumes that the Coulomb model of friction is valid in a point-wise sense along the interface.

at the contact edge, while the singular solution dominates behaviour over a larger region slightly further from the contact edge (but still sufficiently close to it that the remote geometry can be neglected). Hence, it is clear that the validity of the current solution hinges on the ratio R/a being small.

Sackfield et al. [2002] investigate this very issue by performing a very detailed asymptotic analysis that compares the solution for a perfectly sharp-edged punch and a 'flat-and-rounded' punch with a finite edge radius. These authors found that the range of edge radii for which the sharp-edged assumption remains valid increases with the friction coefficient. Sackfield et al. [2002] also provide an example application of their calculations to a contact with a half-width of $5mm$, which is typical for fretting-fatigue experiments, and selected material constants that are typical of metals. Their results showed that when the friction coefficient is 0.6 or 0.9, the maximum discrepancy between the sharp-edged and flat-and-rounded solutions remains below 5% near the contact edge even when an edge radius of 10-$20\mu m$ is present, although this does depend on the strength of the applied loads to some extent. Conversely, when $f = 0.3$, the edge radius would need to be of the order of $0.5\mu m$ for the sharp-edged solution to exert the same level of influence on contact-edge behaviour.

Corner plasticity

Figures 8.4 and 8.7 display the elastic stress singularities that result when the contact edges are assumed to be perfectly sharp. Although there will be a small edge radius in practice as discussed above, the stresses near the contact edge must still be very high for the current solution to remain valid, and these high stresses will of course be truncated by plasticity. However, if this zone of local plasticity is small in comparison to the contact half-width, a small-scale-yielding argument similar to that used in fracture mechanics can be invoked. Namely, if the yield strength of the material is sufficiently high, the zone of plasticity will be small, such that the singular solution remains sufficiently dominant near the contact edge, so the influence of plasticity can be neglected.

For the case when the contact edges are stuck, Williams' solution controls contact-

edge behaviour. Hence, the calculations described in §3 can be used to determine more precisely the range of applied loads that satisfy small-scale yielding. The results presented in §6 suggest that, at least in several cases, typical fretting-fatigue experiments, i.e. those that excite partial-slip and not gross-sliding, tend to fall within the limits of small-scale yielding. For example, fretting-fatigue tests performed by Juoksukangas et al. [2013] were found to fall within 1% small-scale yielding. In cases such as this, the current solution will remain valid. However, if the plastic zone fails to satisfy small-scale yielding, the solution presented in §8.3 will become inaccurate near the contact edges.

Surface roughness

A central assumption that is adopted in this analysis is that of complete contact between two perfectly flat surfaces. This of course represents a significant idealization since all real surfaces possess some degree of roughness. Thus, what appears to be a single contact patch at the macroscopic scale is actually a collection of many microscopic contacts between surface asperities.

Numerous authors have investigated the implications of surface roughness on contact behaviour. Perhaps most notably, Greenwood and Williamson [1966] proposed a three-parameter model for rough contacts, which considers contact between a perfectly smooth plane and a nominally flat surface that is covered with asperities. This model assumes the summits of the asperities are spherical, all the asperities have the same radius, and the height of each summit varies randomly. It is also assumed that individual asperities are far enough apart that interactions between asperities can be neglected, so each asperity contact can be treated as a classical Hertzian contact.

Handzel-Powierza et al. [1992] performed several controlled experiments on 0.45% carbon steel with three different surface finishes (turned, ground, and bead-blasted) and compared their results to the Greenwood–Williamson model. These authors found that the Greenwood–Williamson model agreed well with their data for the ground and bead-blasted surfaces when the normal load was less than half the yield stress of the material. They attributed the discrepancy for higher loads to the increasing influence of micro-

plasticity at the tallest asperities. There was a larger discrepancy for the turned surface, and the authors suggested this was because the asperities probably were not even approximately spherical due to the nature turning process. In addition, McCool [1986] compared the Greenwood–Williamson model to several other numerical models of rough contact and found fairly good agreement between the various models.

More recently, Dini and Hills [2009] studied contact between a smooth sphere and a rough half-space subjected to combined normal and cyclic shear loading. The roughness of the half-space was modelled by uniformly spaced, spherical asperities of identical height and radius. These authors then compared the behaviour of the individual asperity contacts to the contact's overall macroscopic behaviour. They found that the contact's macroscopic response was approximately given by the Catteneo–Mindlin solution for partial-slip [Cattaneo, 1938, Mindlin, 1949]: namely, there was a central stick region followed by zones of partial-slip at the contact's edges. As expected, the asperities in the partial-slip zones (from the macroscopic point of view) were subjected to full sliding at the microscopic scale. Conversely, the asperities within the permanent stick zone were not actually stuck; instead, they were subjected to partial-slip, and the behaviour of each asperity was also given by the Catteneo–Mindlin solution (at the microscopic scale).

Based on the results discussed above, it is clear that the behaviour of real contacts at the microscopic level will be different from that suggested by the macroscopic analysis performed here. One major difference is that the contact tractions will be localized in the regions where surface asperities make contact. However, based on the results obtained by Dini and Hills [2009], it seems reasonable to infer that the overall character of the solution will remain similar at the macroscopic scale, provided that the contact is much larger than the characteristic size of the surface asperities. In addition, the results of Dini and Hills [2009] suggest that asperities within slip zones will be subjected to full sliding, whereas those within stick zones will actually be subjected to partial-slip. Thus, if the contact shown in Figure 8.1 were rough, the precise location of the boundaries between (macroscopic) slip and stick zones shown in Figure 8.2 would probably be modified slightly, but

the overall shape of the curves would probably remain similar. In addition, the results of Handzel-Powierza et al. [1992] suggest that the normal load must remain below about half the yield strength of the material to prevent plasticity from occurring at the asperity level.

Variable friction coefficient

The analysis in this chapter assumes that Coulomb friction applies in a point-wise sense along the interface. In addition, the friction coefficient is assumed to be constant both spatially and temporally. However, Hills et al. [1988] note that the friction coefficient is commonly observed to increase with time during fretting-fatigue tests and that it probably also varies spatially along the interface, though its spatial variation is more difficult to study experimentally. These authors also examine the difference between 'true' and average friction in fretting-fatigue tests and argue that modification of the friction coefficient can only actually occur in zones of frictional slip. Because slip usually only occurs over part of the interface in fretting tests, this will have the tendency to increase the friction coefficient in the slip zones but leave the friction coefficient in the stick zones unchanged, resulting in a non-uniform frictional profile.

Modification of the frictional profile of the contact interface could clearly have a major influence on the results obtained in this chapter. In fact, the difference in behaviour of the low-friction and high-friction cases in §8.3 results solely from such an effect. However, the results in this chapter only consider the case of monotonic loading. Thus, since the process described by Hills et al. [1988] occurs over the course of numerous load cycles, it is unlikely to influence the results presented here. The more likely deviation from the solution presented in §8.3 is that the friction coefficient may have a more complex profile at the onset of loading than the constant distribution assumed here.

The influence of more complex frictional profiles on contact behaviour is studied by Gorbatikh et al. [2001] in the context of multiple two-dimensional cracks sliding along the same plane. These authors consider numerous piecewise distributions of the friction coefficient, and their results reveal that quite complex patterns of slip and stick may result

from such distributions. However, these authors also find that while the propagation of frictional slip is initially very sensitive to the detailed distribution of the friction coefficient, this sensitivity decreases as the slip zones increase in size. Indeed, once the slip zone has propagated over a particular region, these authors find that the result is "almost unaffected" by replacing the friction coefficient by its average value over the slip zone.

Thus, the work of Gorbatikh et al. [2001] seems to suggest that the influence of a non-uniform friction coefficient on the results presented in §8.3 is likely to be largest for low levels of applied shear force when the slip zones are first forming. However, once a larger shear is applied, and slip has propagated over a larger portion of the interface, the results would probably begin to resemble those presented here. Of course, this will only be the case provided that the average friction coefficient of the two cases being compared remains the same, and the variation in the friction coefficient along the interface is not too extreme.

8.5 Summary

The problem of a square elastic block pressed normally into an elastically similar half-plane and then sheared until sliding is achieved is examined. An asymptotic analysis of the contact edge reveals that stick persists until a finite shear force is developed if the coefficient of friction is greater than 0.543. Otherwise both edges slip outwards on the application of a normal load alone. Numerical results show that for moderate levels of shear force (i.e. less than 50% of the sliding condition) very different behaviour is observed for coefficients of friction above and below 0.543. Conversely, for high shear forces (i.e. near to the sliding condition), the response is similar for both cases, and the last point to slip is always an at interior point of the contact, but which is quite near to the trailing edge.

Chapter 9

Calculation of the shakedown limit

In this chapter, a numerical approach for determining the *shakedown limit* of discrete, elastic frictional systems subjected to quasi-static loading is presented. This technique is then used to determine the shakedown limit for a finite element model of a (coupled) complete contact with ∼50,000 total degrees of freedom and ∼250 along the contact interface. Finally, the calculated value of the shakedown limit is compared to a series of over 1,000 transient simulations, and the influence of initial conditions on steady-state frictional energy dissipation is investigated. The results demonstrate that the dissipative properties of complete contacts can be highly dependent on the initial residual slip displacement state.

Note that parts of this chapter have been published in [Flicek et al., 2015a].

9.1 Introduction

It is extremely common for frictional contacts in engineering systems to be subjected to a combination of a static load (to form the contact) and a time varying cyclic load (e.g. due to vibrations). Some examples include riveted and bolted joints [Abad et al., 2012, Law et al., 2006], dovetail joints in gas-turbine engines [Rajasekaran and Nowell, 2006, Xi et al., 2000], and spline couplings [Banerjee and Hills, 2006, Banerjee et al., 2008, Limmer et al., 2001] to name just a few. Furthermore, it is well known that even when the contact interface appears nominally to be stuck, these loading conditions frequently result in small zones of partial-slip along the contact interface, resulting in relative tangential displacements on the order of $25\text{-}100\mu m$ [Szolwinski and Farris, 1996]. The effect of

partial-slip on system performance is two-fold: on the one hand, it results in the well known damage process of fretting fatigue, which significantly reduces the service life of components [Farris et al., 2000]; but on the other hand, it provides a source of damping, which can be beneficial for system performance [Law et al., 2006].

Much of the previous work on fretting fatigue has focused on incomplete contacts, e.g. [Hills et al., 2012b, Nowell and Hills, 1987]. This is probably due to their relevance in engineering practice but also because closed-form solutions exist for many relevant loading cases, e.g. [Ciavarella, 1998a, Jäger, 1998, Mindlin and Deresiewicz, 1953]. For these contacts, the contact area is a function of the applied normal load, and the contact pressure distribution smoothly tends to zero as the contact edge is approached. As a direct consequence of this, a zone of partial-slip will almost always arise at the edge of an incomplete contact subjected to cyclic loading [Barber et al., 2008].

In contrast, numerical analysis of complete contacts subjected to cyclic loading sometimes predicts that the contact *shakes down* (see e.g. [Banerjee and Hills, 2006, Churchman and Hills, 2006b]): that is, after the first few load cycles, a favourable residual displacement state is developed that inhibits slip for all subsequent cycles. This behaviour clearly has some similarities to cyclic plasticity problems, which has prompted the question as to whether a theorem analogous to the elastic shakedown theorem due to Melan [1936] could be developed for frictional contacts. In other words, if there exists a residual displacement distribution that would inhibit slip at all times during the load cycle (i.e. a *safe displacement distribution*), is the existence of this displacement distribution sufficient to guarantee that the contact will shake down?

This question has been answered unequivocally by Klarbring et al. [2007] for discrete contact problems and by Barber et al. [2008] for continuum problems. These authors have proven that the existence of a safe displacement distribution is both a necessary and sufficient condition to guarantee that shakedown will occur if the contact is *uncoupled*: that is, if relative tangential contact displacements have no influence on the distribution of the normal contact traction. Thus, for uncoupled contacts, there exists a load factor,

i.e. a ratio of cyclic load amplitude to static load (e.g. σ/P in Figure 7.4), above which shakedown is impossible and below which it is guaranteed; hence, the contact's steady-state response never depends on initial conditions.

On the other hand, if the contact is coupled, the existence of a safe displacement distribution is still a necessary condition for shakedown to be possible, but it is no longer sufficient. In fact, for coupled contacts, there almost always exists a range of load factors for which the steady-state response depends on the initial residual displacement distribution (see [Barber et al., 2008, Klarbring et al., 2007]). In general, there are three limits on the load factor, which mark the boundaries between distinct steady-state responses:

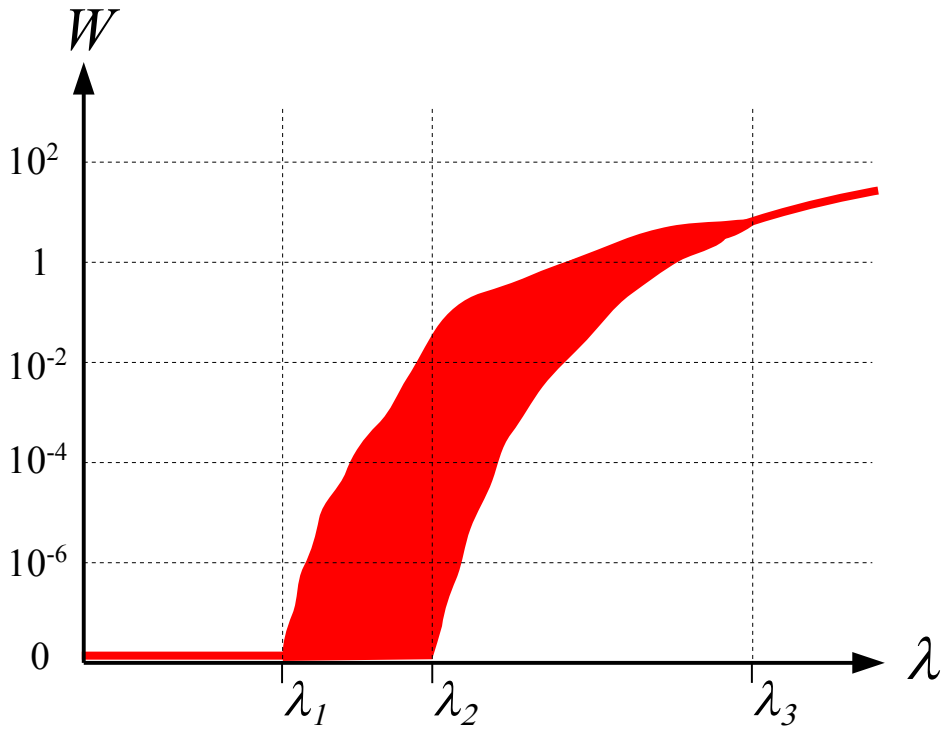
1. A load factor below which shakedown is guaranteed to occur, denoted λ_1.

2. A load factor above which shakedown is impossible, denoted λ_2.

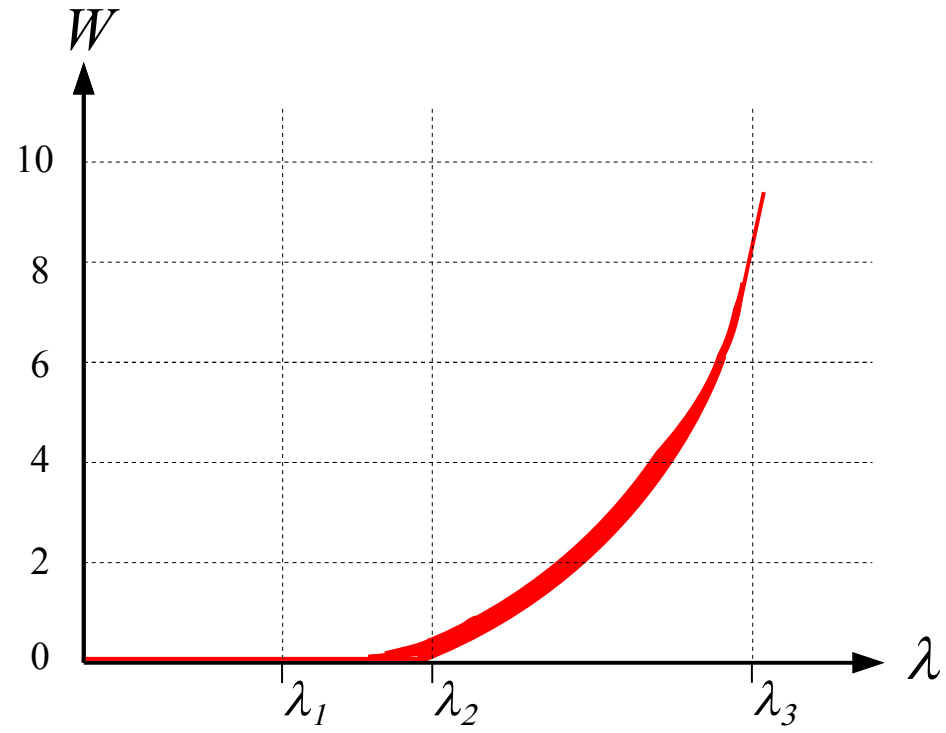
3. A load factor above which steady-state dissipation is unique, denoted λ_3.

The implications of these three load factors can be understood more easily by viewing a plot of the frictional energy dissipation that is present in the steady state, W, vs. load factor, λ. Hence, example plots are shown in Figure 9.1, where the data is shown on a semi-log scale in Figure 9.1(a) and on a linear scale in Figure 9.1(b). Notice that the semi-log scale in Figure 9.1(a) strongly emphasizes the range of dissipation that is possible for small orders of magnitude, which are significantly compressed and are difficult to distinguish between on a linear scale. For this reason, all the plots of W vs. λ in this chapter use a semi-log scale, but the reader should keep in mind how the results between λ_1 and λ_2 would be compressed if they were shown on a linear scale.

Figure 9.1 illustrates that when the load factor is less than λ_1, all possible steady-state solutions involve zero energy dissipation, irrespective of the initial conditions. For load factors above λ_1, some steady-state solutions that involve energy dissipation (i.e. cyclic slip in the steady state) become possible. However, for all load factors below λ_2, it remains possible to reach a non-dissipative (i.e. shaken down) steady state if an appropriate initial condition is chosen. Thus, for load factors between λ_1 and λ_2, initial conditions determine



(a) Semi-log scale.



(b) Linear scale.

Figure 9.1: Illustrative plots that show the range of frictional energy dissipation that can persist in the steady state, W, depending on the load factor, λ.

whether the contact will shake down. For load factors greater than λ_2, shakedown is impossible, but the initial conditions still influence the steady-state solution and determine how much dissipation will be present. Finally, for load factors greater than λ_3, the amount of dissipation that is present in the steady state becomes unique.[1]

Note that some or all of these limits on the load factor can be zero. For example, for incomplete contacts, $\lambda_1 = \lambda_2 = \lambda_3 = 0$, which implies both that shakedown is impossible and that the amount of dissipation that is present in the steady state is always independent of the initial conditions. Moreover, even if $\lambda_2 > 0$, it is still possible that $\lambda_1 = 0$, implying that although shakedown is sometimes possible, whether it actually occurs always depends on initial conditions.

As previously mentioned, the amount of energy dissipated due to friction has a significant influence on the performance of engineering assemblies. Hence, to the extent that these limits on the cyclic load determine the character of the steady state response, they have practical relevance. However, some of these limits will be more relevant than others. For instance, Figure 9.1(b) illustrates that while zero steady-state energy dissipation is only guaranteed for load factors below λ_1, the amount of dissipation that can occur below λ_2 tends to be several orders of magnitude lower than that which is certain to occur for load factors just moderately greater than λ_2.

Thus, in many cases, the most practically relevant limit will be λ_2 because it corresponds to the boundary between there being the potential for minuscule to mild dissipation (for $\lambda < \lambda_2$) and the guarantee of significant dissipation (for $\lambda > \lambda_2$). In most cases, λ_1 is likely to be of somewhat less practical importance than λ_2 since knowledge of λ_2 in conjunction with a rough estimate of the worst-case dissipation that can occur below λ_2 is likely to be sufficient for design purposes. The exception is for cases in which all frictional energy dissipation must be avoided, and the initial conditions cannot be controlled

[1] An example situation that will cause the steady-state solution to become unique is if every point along the contact interface slips simultaneously at any point during the load cycle. Once this occurs, the tractions at every point are specified by the friction law, i.e. (9.2f). Thus, all information from the previous state (including the initial conditions) is removed from the system, and the next state will only depend on the loading regime.

with sufficient accuracy to ensure that they will lead to shake down.

Another reason that λ_1 is perhaps less relevant than λ_2 in practice is that all real contacts that are subjected to cyclic loads – even those that remain fully adhered – dissipate some energy due to friction. This is because all surfaces are actually rough, so a contact interface that appears to be flat is actually comprised of numerous microscopic contacts between surface asperities, some of which will undergo partial-slip even if the contact interface nominally appears to be stuck.[2] Thus, the existence of λ_1 is in some senses an artefact that arises from neglecting the influence of surface roughness. [3]

Similarly, λ_3 also probably has less practical relevance than λ_2. This is because it is quite unlikely that contacts that are modelled as being stationary will be loaded so severely that they reach λ_3 since this limit will probably tend to correspond to a sliding condition of some kind (i.e. when the loads are such that if they are increased any further, the system will be out of equilibrium). Conversely, if a contact is expected to be subjected to gross-sliding, it can be deduced that it will be loaded above λ_3 without performing any analysis at all. This is simply because for a contact that is subjected to gross-sliding, the tractions are defined by the friction law (i.e. (9.2f)), which will erase any prior history-dependence that the problem exhibited. Thus, λ_2 is probably the most relevant limit on the load factor for practical purposes.

Ahn et al. [2008] have presented a method for calculating both λ_1 and λ_2. Unfortunately, their approach involves solving a system of equations that becomes combinatorially more complex as the number of degrees of freedom in the system increases. For instance, Jang and Barber [2011] were able to apply this approach to determine both λ_1 and λ_2 for a system of 10 cracks, but applying this analysis to a significantly larger frictional system would be computationally prohibitive. As it is common for the finite element models that are used in practice to incorporate one or two orders of magnitude

[2] See §8.4.1 for a more detailed discussion of the influence of surface roughness.

[3] Of course, the same could also be said of λ_2 since the argument just presented simply asserts that shakedown, in a rigorous sense, cannot actually occur in practice. The reason this poses more of an issue for λ_1 than for λ_2 is that the amount of dissipation that this analysis predicts will occur for load factors that are moderately greater than λ_1 (but are less than λ_2) might end up being of the same order of magnitude as that which occurs in practice due to the effects of surface roughness for load factors below λ_1.

more nodes along the frictional interface than are considered in [Jang and Barber, 2011], it is of practical interest to develop a more computationally efficient approach.

Björkman and Klarbring [1987] have presented an efficient approach for calculating λ_2, which will be referred to as the *shakedown limit*. To do this, these authors formulate the contact problem in discrete form and show that the shakedown-limit calculation can be posed as a linear-programming problem for which several efficient solution algorithms exist. These authors then apply this calculation to two example problems and also compare the results to numerical results obtained using the finite element method.

In this chapter, this mathematical-programming approach for calculating the shakedown limit due to Björkman and Klarbring [1987] is examined. A simple approach for determining when the onset of slip or separation will occur for a contact subjected to monotonic loading is also presented. These calculations are then applied to the example (coupled) complete contact shown in Figure 7.4. Over 1,000 numerical simulations of the evolution of the contact interface with time are then performed under a wide range of cyclic loads and initial displacement states, and the influence of initial conditions on steady-state frictional energy dissipation is investigated.

9.2 Formulation

Consider some two-dimensional complete contact geometry, such as that shown in Figure 7.4, which has been discretized using the finite element method, such that there are N nodes in potential frictional contact along the interface. The reaction forces, $\boldsymbol{f}^{\mathrm{C}}$, and the relative displacements, $\boldsymbol{u}^{\mathrm{C}}$, at the contact nodes can be written as in (7.28) and (7.29), where q_i, p_i are the shear, normal reactions; v_i, w_i are the tangential, normal relative displacements; and $i \in \{1, ..., N\}$. Also recall that the sign convention for the contact forces and displacements is given in Figure 7.2, such that p_i is positive in compression, and w_i is positive for a positive gap.

An expression for the evolution of the contact reactions with time can be written as

$$\boldsymbol{f}^{\mathrm{C}}(t) = \boldsymbol{K}^{\mathrm{C}}\boldsymbol{u}^{\mathrm{C}}(t) + \boldsymbol{f}^{w}(t), \tag{9.1}$$

where $\boldsymbol{K}^{\mathrm{C}}$ is a symmetric $2N \times 2N$ *contact stiffness matrix*, and $\boldsymbol{f}^{w}$ accounts for the influence of the external loads on the contact reactions, and these can be obtained as described in §7.

9.2.1 The Coulomb friction law

Before defining the friction law, it is noted that the contacting bodies cannot inter-penetrate, and the normal contact reactions cannot be tensile:

$$w_i \geqslant 0, \quad p_i \geqslant 0. \tag{9.2a}$$

Furthermore, if there is a positive gap at node i, its normal and shear reactions must be zero. Conversely, if the normal reaction at node i is compressive, its gap must be zero:

$$w_i > 0 \Rightarrow p_i = q_i = 0 \tag{9.2b}$$

$$p_i > 0 \Rightarrow w_i = 0. \tag{9.2c}$$

Finally, it is assumed that the Coulomb friction condition applies in a point-wise sense at each node along the contact interface, and this requires that

$$|q_i| \leqslant f p_i, \quad f > 0 \tag{9.2d}$$

$$|q_i| < f p_i \Rightarrow \dot{v}_i = 0 \tag{9.2e}$$

$$0 < |q_i| = f p_i \Rightarrow \mathrm{sgn}(\dot{v}_i) = -\mathrm{sgn}(q_i), \tag{9.2f}$$

where f is the coefficient of friction, and a superposed dot denotes the derivative with respect to time, t. Thus, (9.2d) defines the admissible range of shear reactions; (9.2e)

states that if the magnitude of q_i is less than a critical value, node i is stationary; and (9.2f) states that if $|q_i|$ is equal to a critical value, slip occurs in the direction that opposes the shear reaction at that node.

9.2.2 Loading regime

In this chapter, a loading regime comprised of a static component, F_s, superposed with a quasi-statically applied, time-varying (cyclic) component, $F_c(t)$, is considered, where F_s, F_c are scalars that account for the strength of the applied loads. The effect of these applied loads on the contact reactions can be expressed as

$$\boldsymbol{f}^w(t) = F_s \boldsymbol{f}^{ws} + F_c(t) \boldsymbol{f}^{wc}, \tag{9.3}$$

where $\boldsymbol{f}^{ws}, \boldsymbol{f}^{wc}$ are the $\boldsymbol{f}^w$ vectors corresponding to a unit application of F_s, F_c, respectively. With this notation, (9.1) can be re-written as

$$\left\{ \begin{array}{c} \boldsymbol{q}(t) \\ \boldsymbol{p}(t) \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{v}(t) \\ \boldsymbol{w}(t) \end{array} \right\} + F_s \left\{ \begin{array}{c} \boldsymbol{q}^{ws} \\ \boldsymbol{p}^{ws} \end{array} \right\} + F_c(t) \left\{ \begin{array}{c} \boldsymbol{q}^{wc} \\ \boldsymbol{p}^{wc} \end{array} \right\}, \tag{9.4}$$

where $\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{v}, \boldsymbol{w}$ are defined in (7.29), and $\boldsymbol{K}$ is partitioned into sub-matrices corresponding to the following force-displacement relationships: tangential-tangential, $\boldsymbol{A}$; normal-tangential, $\boldsymbol{B}$; and normal-normal, $\boldsymbol{C}$.

In addition, the load factor, λ, is defined as

$$\lambda = \frac{F_c(t)|_{\max} - F_c(t)|_{\min}}{F_s} = \frac{F_c^{\max} - F_c^{\min}}{F_s}, \tag{9.5}$$

where $F_c^{\min}$, $F_c^{\max}$ are the minimum, maximum values of the cyclic load, respectively. Note that the load factor is only defined for cyclic loading regimes.

9.3 First violation of the stick condition

The load ratio, F_c/F_s, at which slip or separation first occurs under monotonic loading is always dependent on the initial residual slip displacement distribution, $\boldsymbol{v}(0)$. Nevertheless, this information may be helpful when planning or analysing experiments, particularly when the case of a null initial slip distribution, i.e. $\boldsymbol{v}(0) = \mathbf{0}$, is considered. Moreover, this calculation is trivially simple to carry out.

Consider a complete contact with $\boldsymbol{v}(0) = \mathbf{0}$ that is subjected to a loading regime in which the static load, F_s, is first applied and held constant, and the time-varying load, F_c, is subsequently applied monotonically. The load ratio at which slip or separation first occurs can be determined by checking when the stick condition is first violated. In other words, the lowest value of F_c/F_s that violates the condition $p_i \geqslant 0$ or $|q_i| < fp_i$ at any node, i, must be determined. Recall that for complete contacts, $\boldsymbol{w} = \mathbf{0}$ in the absence of external loads (by definition). Thus, (9.4) can be used to write these conditions as

$$p_i^{ws} + \left(\frac{F_c}{F_s}\right) p_i^{wc} \geqslant 0 \tag{9.6a}$$

$$q_i^{ws} + \left(\frac{F_c}{F_s}\right) q_i^{wc} < f(p_i^{ws} + \left(\frac{F_c}{F_s}\right) p_i^{wc}). \tag{9.6b}$$

Note that these conditions are only valid for predicting first violations, so once slip, for example, is predicted to initiate from (9.6b), separation is no longer accurately predicted from (9.6a).

9.4 The shakedown limit

It is perhaps most natural to think about frictional shakedown as it might arise in the context of an experiment: namely, there is some test configuration that is subjected to a particular combination of loads, and the coefficient of friction (which is assumed to remain constant) is determined by the material properties, surface treatment, etc. Thus, to determine the shakedown limit experimentally, one might apply cyclic loads of vari-

ous amplitudes, vary the initial displacement state, and use digital image correlation to determine the maximum cyclic load amplitude for which the contact shakes down. In contrast, the approach described here for calculating the shakedown limit is, in some sense, the reverse of this hypothetical experimental approach.

To pose the calculation of the shakedown limit as a mathematical-programming problem as in [Björkman and Klarbring, 1987], first consider a cyclic loading regime in which the cyclic load ranges from some minimum value, $F_c^{\min}$, to some maximum value, $F_c^{\max}$. Next, assume that the load factor (see (9.5)) corresponding to this load range *is* the shakedown limit if the coefficient of friction is some currently unknown value, denoted $f^{\min}$. The aim therefore becomes to determine $f^{\min}$: that is, the coefficient of friction below which it is impossible for the contact to shake down for the specified load range. This calculation can then be performed for a series of load ranges to determine the shakedown limit for a range of friction coefficients.

Note that if the load range is sufficiently large, it may be impossible for the contact to shake down irrespective of the coefficient of friction, making the initial assumption false. This can occur if there does not exist any residual slip displacement state that inhibits separation throughout the load cycle. In this case, the calculation will simply fail to converge, implying that shakedown is impossible for the specified load range.

9.4.1 Calculation of the shakedown limit

To determine $f^{\min}$, first recall that if a contact is to shake down, separation cannot occur at any time in the load cycle, i.e. $\boldsymbol{w}(t) = \boldsymbol{0}$. Therefore, for a contact that has shaken down, (9.4) can be used to write the ratio of the (element-wise) absolute value of the shear reactions to the normal reactions, $\boldsymbol{R}$, as

$$\boldsymbol{R}(\boldsymbol{v}; F_c) = \frac{|\boldsymbol{q}(F_c)|}{\boldsymbol{p}(F_c)} = \frac{|\boldsymbol{A}\boldsymbol{v} + F_s\boldsymbol{q}^{ws} + F_c\boldsymbol{q}^{wc}|}{\boldsymbol{B}\boldsymbol{v} + F_s\boldsymbol{p}^{ws} + F_c\boldsymbol{p}^{wc}}. \tag{9.7}$$

Notice that the contact reactions are written as a function of F_c as opposed to a strict function of time, t. This can be done because every load cycle for a shaken down contact is identical, and the reactions simply vary with the applied loads. Also notice that $\boldsymbol{v}$ is not a function of F_c because, by definition, slip cannot occur at any point during the load cycle for a contact that has shaken down.

According to the friction law defined in §9.2.1, slip will occur if any element of $\boldsymbol{R}$ becomes equal to the coefficient of friction. As the loads are applied quasi-statically, $\boldsymbol{R}$ will take on its maximum value at one (or both) of the *extreme* points (in time) of the load cycle, i.e. at $F_c^{\min}$ or $F_c^{\max}$. Hence, a vector, $\boldsymbol{S}(\boldsymbol{v})$, of length $2N$ can be constructed that is comprised of the contact reaction ratios, $\boldsymbol{R}$, at the *extreme* points in the load cycle as

$$\boldsymbol{S}(\boldsymbol{v}) = \left\{ \begin{array}{c} \boldsymbol{R}(\boldsymbol{v}; F_c^{\min}) \\ \boldsymbol{R}(\boldsymbol{v}; F_c^{\max}) \end{array} \right\}. \tag{9.8}$$

Thus, further slip will be inhibited throughout the load cycle if

$$f \geqslant \boldsymbol{S}(\boldsymbol{v})|_{\max}, \tag{9.9}$$

i.e. if the coefficient of friction is greater than the maximum element value of $\boldsymbol{S}$.

To determine $f^{\min}$, an optimal slip distribution, $\boldsymbol{v}^{opt}$, must be found. For the purposes of this calculation, an optimal slip distribution is one that: (i) prevents separation at all times in the load cycle and (ii) requires the lowest coefficient of friction to prevent slip from occurring. From (9.9), it can be seen that (ii) requires that $\boldsymbol{v}^{opt}$ minimizes the maximum value of $\boldsymbol{S}$, while (i) requires that the condition $\boldsymbol{p} \geqslant 0$ is simultaneously satisfied. Therefore, finding $\boldsymbol{v}^{opt}$ is a standard 'mini-max' optimization problem of the form

$$\min_{\boldsymbol{v}} \{ \max_{j} S_j(\boldsymbol{v}) \} \text{ such that } \boldsymbol{p} \geqslant 0, \tag{9.10}$$

where $j \in \{1, ..., 2N\}$. Several algorithms exist for the solution of such problems, e.g. [Brayton et al., 1979, Powell, 1978]. It is noted that the solution to this problem,

v^{opt}, generally is not unique. In addition, note that although the problem is formulated as a 'minimax' constrained optimization problem in this chapter, Björkman and Klarbring [1987] formulate it as a constrained maximization. However, both approaches are equivalent, and the reader is directed to the original paper by Björkman and Klarbring [1987] for a discussion of the mathematical structure of this optimization.

9.4.2 Improving convergence

One additional constraint that is often helpful for improving the convergence of the optimization in (9.10) is to eliminate the potential for rigid-body motion in the slip displacement distribution. This can be done by removing one node's slip displacement from the optimization and optimizing against a subset of the slip displacements, $\boldsymbol{v}^*$, where

$$\boldsymbol{v}^* = \{v_1, ..., v_{N-1}\}^T . \tag{9.11}$$

The slip displacement v_N, which is not optimized against, can then be set to

$$v_N = -\sum_{i=1}^{N-1} v_i \tag{9.12}$$

to force the sum of all the slip displacements to be zero. Note that it is not important which node's slip displacement is removed from the optimization.[4]

9.5 Application to example problem

The calculations described in this chapter are now applied to the example problem shown in Figure 7.4 to illustrate how they may be used in practice. Figure 7.4 shows the complete contact formed between two elastic continuum bodies: a square punch and an elastically similar half-plane. As these calculations use the discrete formulation, the first step in the analysis is to discretize the contact, which has been done using the commercial

[4]This was tested by removing nodes at various positions along the contact from the optimization and specifying their slip displacement according to (9.12); the solution was insensitive to which node was chosen.

finite element software ABAQUS/CAE. Of course, once these bodies have been discretized, the problem is not identical to the continuum problem in a rigorous sense. However, in practice, it is usually assumed that a continuum is well represented by a discrete system with a sufficiently refined mesh.

The details of the finite element model that was used are provided in Appendix §A.3.1, but here it suffices to state that both bodies have an elastic modulus of $E = 200GPa$, a Poission's ratio of $\nu = 0.3$, the model assumes plane-strain with unit depth, the contact interface comprises 128 uniformly spaced nodes, and the full model incorporating the punch and the half-plane comprises 50,040 degrees of freedom, where every node in the model is associated with two degrees of freedom. Note that the finite element model does not take advantage of the symmetry about the vertical centreline of the contact. This is simply because these results were generated from a set of models that were created to study several different loading regimes, and some of these loads do not maintain this horizontal symmetry, e.g. when a uniform shear traction is applied to the top of the punch.

To determine the sensitivity of the contact solution to the discretization, a mesh convergence study was performed in which both the number of contact nodes and the total number of nodes in the model were varied. The results of this study suggest that a model with 128 contact nodes provides a converged result at the interior of the contact but that the solution close to the corner is not fully converged. However, since obtaining a converged result at the corners with a uniform mesh along the contact would be computationally impractical, a more complex biased mesh would be required. As the primary object of this analysis is to demonstrate the technique, this simple mesh will be sufficient, but note that the technique presented in §9.4.1 can be applied to larger problems without modification.

The next step in this analysis is to apply the static-reduction technique described in §7, which reduces the finite element model to a discrete system of the N contact nodes alone and provides the contact stiffness matrix, $\boldsymbol{K}^{\mathrm{C}}$, and the $\boldsymbol{f}^{w}$ vectors corresponding to each applied load, $P, \sigma(t)$. Note that P corresponds to the mean contact pressure, and $\sigma(t)$ corresponds to the mean bulk load; hence, both loads have units of $[FL^{-2}]$, where

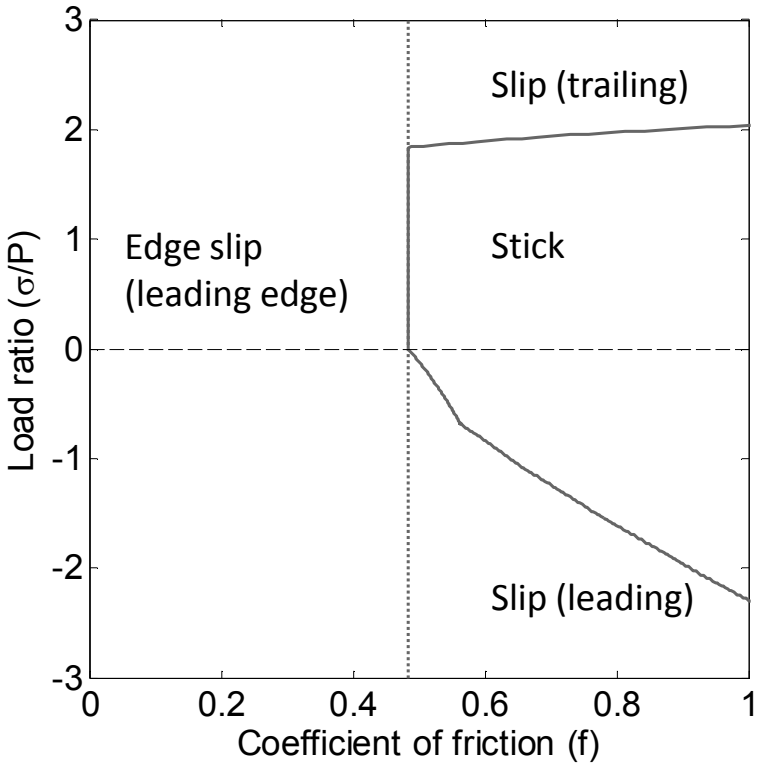


Figure 9.2: Slip-stick behaviour of the contact shown in Figure 7.4 when P is first applied to a contact with a null initial slip displacement distribution, and σ is then applied monotonically in tension (positive) or compression (negative).

F, L denote force, length, respectively. Thus, the applied loads for this example problem can be written in the form given by (9.3) as

$$\boldsymbol{f}^w(t) = P\boldsymbol{f}^P + \sigma(t)\boldsymbol{f}^\sigma, \tag{9.13}$$

where $\boldsymbol{f}^P$, $\boldsymbol{f}^\sigma$ have units of $[F/(FL^{-2})]$ or simply $[L^2]$ and correspond to the application of a unit load in the directions shown in Figure 7.4. Finally, the load factor is defined as

$$\lambda = \frac{\sigma(t)|_{\max} - \sigma(t)|_{\min}}{P} = \frac{\sigma_{\max} - \sigma_{\min}}{P}, \tag{9.14}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the minimum and maximum values of σ, respectively.

9.5.1 First violation of the stick condition

The load ratio, σ/P, at which slip or separation first occurs can now be determined for this contact when $\boldsymbol{v}(0) = \mathbf{0}$. To do this, it is assumed that P is first raised to some value and then held constant and that $\sigma(t)$ is subsequently applied monotonically (in tension or compression). The load ratio, σ/P, that results in the first violation of either of the conditions specified in (9.6) can then be found, and the results of this calculation are shown in Figure 9.2 on a plot of σ/P vs. f.

The way to interpret this figure is to pick a coefficient of friction, which corresponds to some vertical line, and to view the behaviour that is implied for monotonically increasing/decreasing σ, which corresponds to a bulk tension/compression, respectively. Note that here the convention is adopted that *leading-edge* slip refers to the edges of the punch slipping outwards, whereas *trailing-edge* slip refers to the edges of the punch slipping inwards (see Figure 4.1).

The first thing to notice from this figure is that if $f < 0.48$, the first violation of the stick condition occurs *before* σ is applied: that is, leading-edge slip initiates from the contact edge on the application of P alone. Recall that the asymptotic analysis of the contact edge (using the continuum formulation) presented in §4 predicts that edge slip will occur if $f < 0.543$ (since this is the value of $g^I_{r\theta}$ when $\phi = 90°$). The discrepancy between these results is due to the mesh being insufficiently fine near the contact edges. Hence, this provides a simple illustration of the improvement in accuracy that can be achieved by employing Williams' solution rather than a purely numerical approach.

In any case, when $f > 0.48$, the entirety of the contact interface remains stuck on the application of P. If σ is then monotonically applied in tension, slip initiates in the trailing-edge sense when $\sigma/P \sim 2$, depending on the coefficient of friction. Conversely, if σ is monotonically applied in compression, slip initiates in the leading-edge sense within the range $-2.3 < \sigma/P < 0$. Note that although slip initiates somewhat inwards from the contact edge when $f > 0.48$ (in both tension and compression), the point of first slip

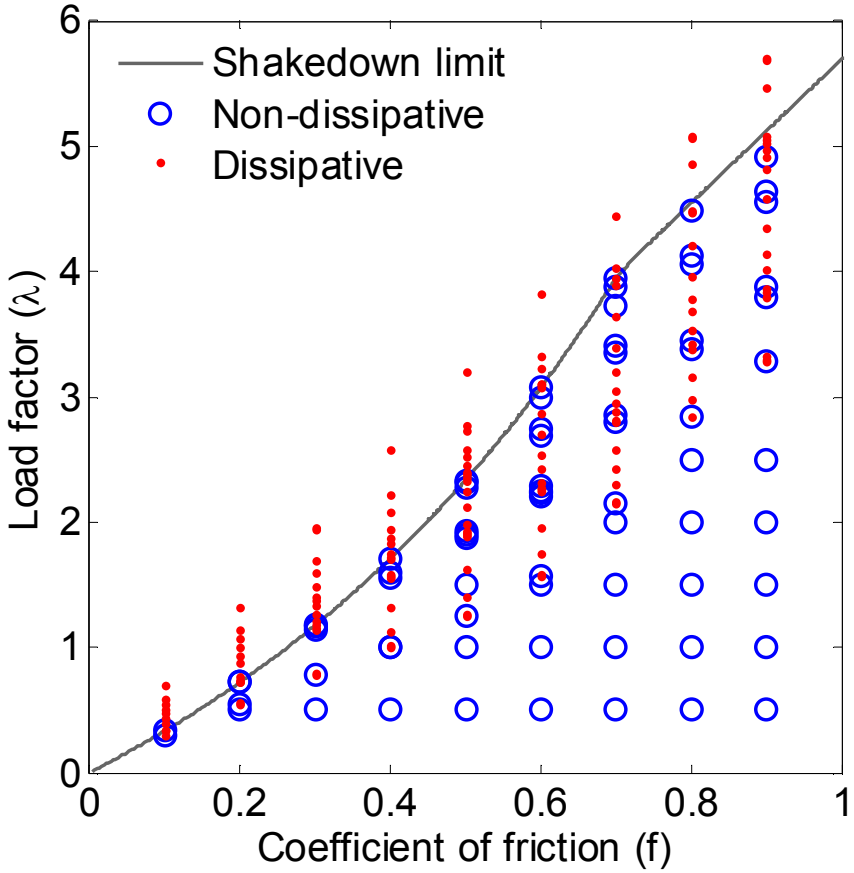


Figure 9.3: Plot of the (calculated) shakedown limit and the results of 347 transient simulations, which were run until a steady state was reached. Transient simulations that shook down are shown with a circle, and those that did not shake down are shown with a dot.

always remains quite close to the contact edge. Also notice that for $f < 1$, the first violation of the stick condition always results from slip, not separation.

9.5.2 The shakedown limit

In this section, the shakedown limit, λ_2, is calculated for the contact shown in Figure 7.4, and the results are compared to a series of transient simulations. The loading history under consideration is shown in Figure 7.5: namely, P is first applied and held constant, and $\sigma(t)$ is then varied within the range $0 \leqslant \sigma(t) \leqslant \sigma_{\max}$. In other words, this is a 'zero-to-peak' loading regime in cyclic tension, so the load factor is defined as $\lambda = \sigma_{\max}/P$ (see (9.5)). Note that the initial slip displacement condition, i.e. $\boldsymbol{v}(0)$, need not be null. Indeed, unlike the simulations performed in §7, the simulations presented in

this chapter used a wide range of $\boldsymbol{v}(0)$ distributions.[5]

The shakedown limit, λ_2, was calculated for the loading regime described above in MATLAB R2013a using the `fminimax` function, and the results are shown in Figure 9.3 on a plot of λ vs. f. Also shown in this figure are the results of 347 'marching-in-time' transient simulations,[6] which were carried out using the Gauss-Seidel algorithm described in §7.5. Simulations that eventually shook down are marked with a circle, and those that failed to shake down are marked with a dot. The main thing to recognize about this figure is that it demonstrates that the calculated values of λ_2 are consistent with the results of all the transient simulations that were performed: that is, every transient simulation for $\lambda > \lambda_2$ failed to shake down.

9.5.3 Frictional energy dissipation

All the transient simulations that were performed were subjected to the loading regime shown in Figure 7.5 until a steady state was reached, which was determined using a criterion based on the frictional energy dissipation per cycle, W. Specifically, the contact was determined to have reached a steady-state response if: (i) W was similar in size to the smallest meaningful value of W that can be represented with a double-precision floating-point number, or (ii) the change in dissipation from the previous load cycle was less than 0.01%. The frictional energy dissipation per cycle was calculated as

$$W = -\sum_{i=1}^{N} \oint_{cycle} q_i(t)\dot{v}_i(t)dt, \tag{9.15}$$

where the time-integration was performed numerically with a piecewise-linear approximation in each load increment. Note that W is a non-negative quantity because (9.2) states that the sign of q_i always opposes that of $\dot{v}_i$ for $\dot{v}_i \neq 0$. Also note that a convergence

[5]Note that $\boldsymbol{v}(0)$ was varied by first assuming a particular displacement distribution and then simply scaling this distribution. This approach enables any distribution to be used, but for the simulations presented here, $\boldsymbol{v}(0)$ was assumed either to be a linear or sinusoidal distribution or the optimal distribution obtained from the calculation of λ_2.

[6]It is difficult to identify some of these simulations in this and subsequent figures since multiple simulations were performed with the same λ but with different $\boldsymbol{v}(0)$.

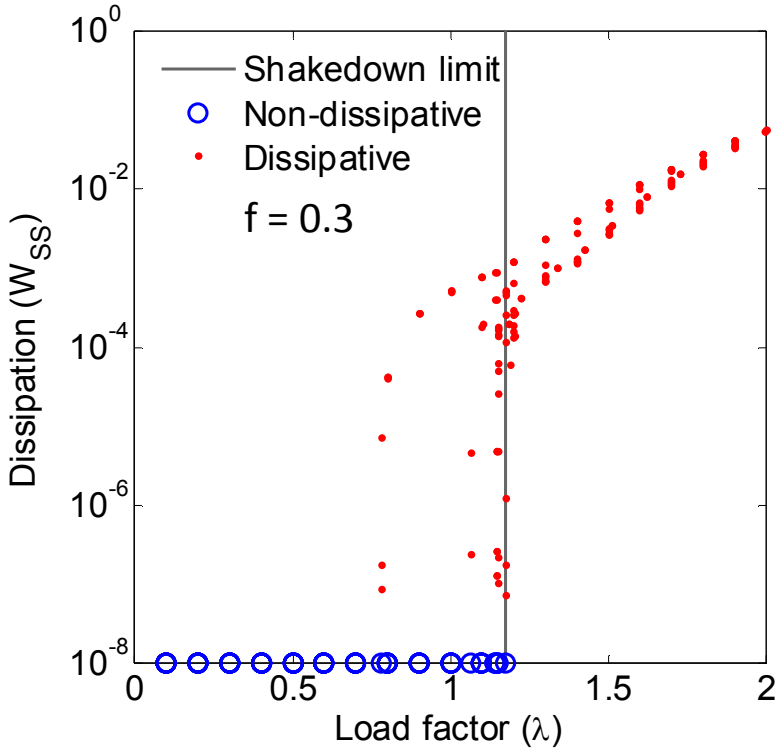
study was performed, and it was found that a load increment of 0.05 (with respect to σ) was sufficient to obtain a converged result. Hence, this load increment was used for all the transient simulations that were performed.

As this model is two-dimensional, W is the dissipation per load cycle per unit depth, so its units are $[(FL)/L]$ or simply $[F]$. Thus, the dissipation per cycle in the steady state can be written in dimensionless form, W_{SS}, which is given by

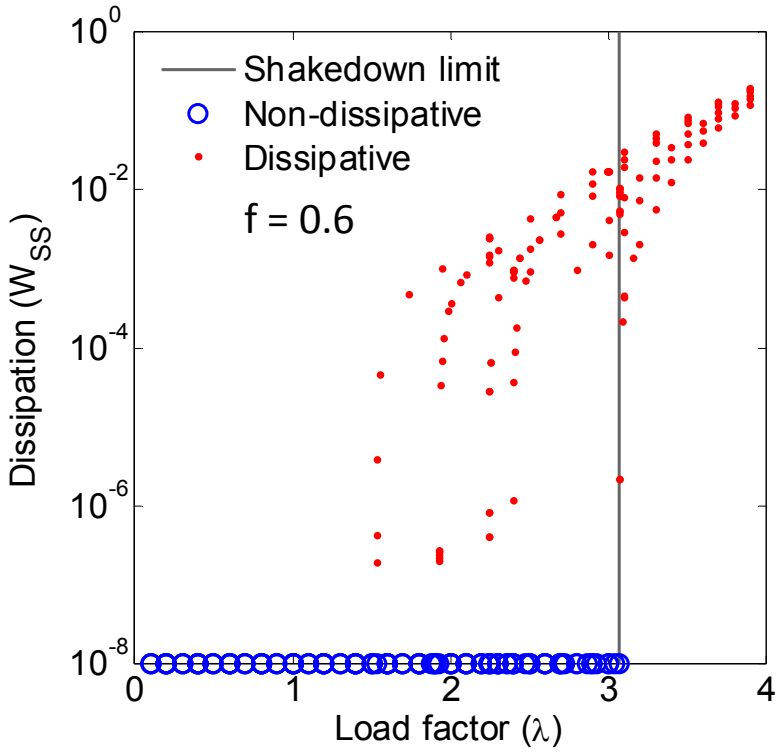
$$W_{SS} = \frac{E}{a^2 P^2 (1-\nu^2)} W, \tag{9.16}$$

where $2a$ is the contact-width. To illustrate the effect that initial conditions can have on W_{SS}, this quantity is plotted against λ in Figures 9.4-9.5 for 873 transient simulations that were carried out at three coefficients of friction: 0.3, 0.6, and 0.9. Each of these plots also includes a vertical line corresponding to λ_2 at the relevant coefficient of friction. Note that as the vertical scales on these plots are logarithmic, they cannot display a value of zero. Thus, simulations that shook down (for which $W_{SS} = 0$) are plotted as the arbitrarily small value of 10^{-8}.

Figures 9.3-9.5 demonstrate that the range of λ for which shakedown is *conditional* (i.e. where some simulations shook down while others did not) can be quite large. This is shown in Figures 9.4-9.5 since simulations failed to shake down when λ was as low as 66%, 50%, and 62% of the value of λ_2 when f was 0.3, 0.6, and 0.9, respectively. In addition, Figures 9.3-9.5 suggest that the size of the conditional region (as a percentage of λ_2) remains quite similar over the range of f that was considered. However, it is almost certain that the conditional region is somewhat larger than is suggested by these figures because it is unlikely that any of the $\boldsymbol{v}(0)$ distributions that were used were the absolute 'worst' $\boldsymbol{v}(0)$ possible (in terms of encouraging shakedown).



(a) $f = 0.3$



(b) $f = 0.6$

Figure 9.4: Dimensionless frictional energy dissipated per load cycle in the steady state, W_{SS}, plotted against load factor, λ, for $f = 0.3, 0.6$.

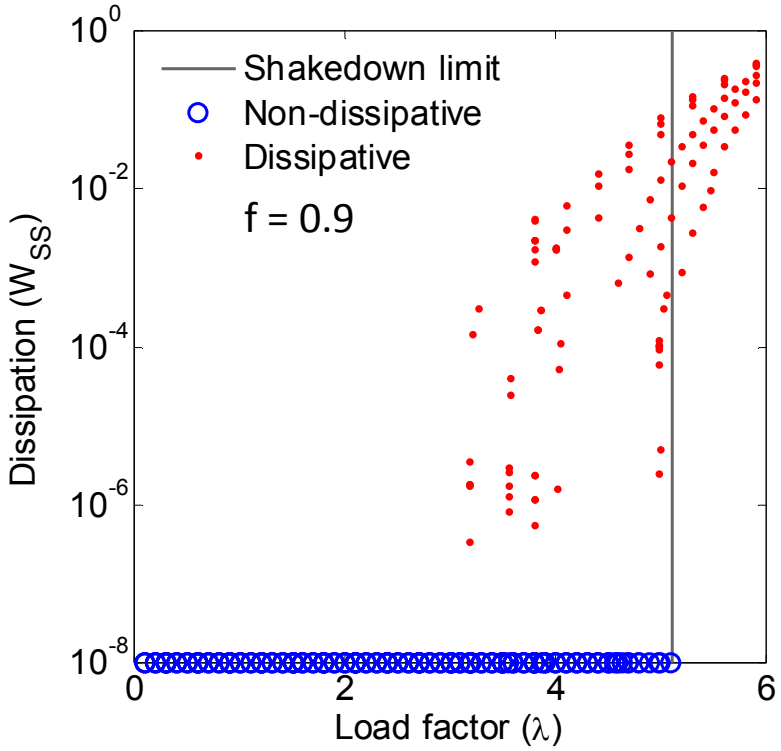


Figure 9.5: Dimensionless frictional energy dissipated per load cycle in the steady state, W_{SS}, plotted against load factor, λ, for $f = 0.9$.

9.5.4 Optimal initial residual slip displacement

Figures 9.4-9.5 illustrate that the frictional energy dissipated by a complete contact after several load cycles can be very sensitive to the initial residual slip displacement state, $\boldsymbol{v}(0)$. Hence, it is of practical interest to determine which sorts of initial displacement conditions result in a more/less dissipative contact. Conveniently, an optimal slip distribution for promoting shakedown is automatically obtained when calculating λ_2. Figure 9.6 shows three of these distributions in dimensionless form plotted against normalized position along the contact interface, x/a, for $f = 0.3, 0.6, 0.9$, where the dimensionless slip displacement is given by

$$\boldsymbol{v}_{dm} = \frac{E}{aP(1-\nu^2)}\boldsymbol{v}. \tag{9.17}$$

Although the information provided by Figure 9.6 is useful, it is clear that pre-loading a contact with initial residual displacement distributions such as these might be difficult

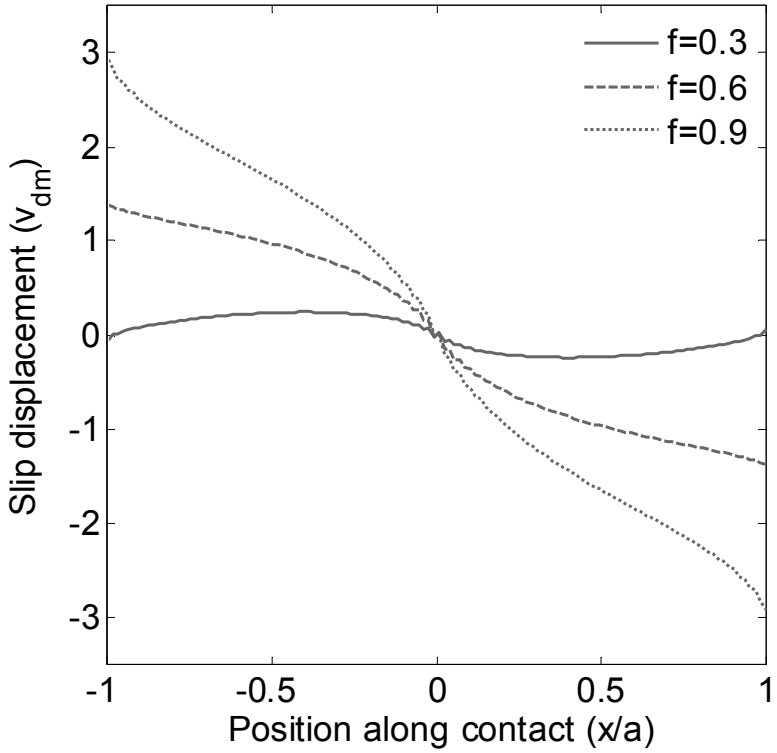


Figure 9.6: Optimal residual slip displacement distributions plotted in dimensionless form vs. normalized position along the contact, x/a.

in practice. However, a linear distribution of $\boldsymbol{v}(0)$ can be achieved simply by applying an appropriate bulk stress, σ, before the static normal load, P, is applied. Hence, this prompts the question: what is the best/worst initial condition for promoting shakedown that can be achieved by this procedure? To address this question, a series of simulations were performed with different bulk pre-loads, and the results of these simulations are shown in Figure 9.3. In addition, the influence of a (dimensionless) bulk pre-load, σ_{dm}, on steady-state dissipation, W_{SS}, is shown in Figure 9.7 for the example case of $f = 0.6$ and $\lambda = 3$, where the dimensionless bulk load is given by $\sigma_{dm} = \sigma/P$.

Figure 9.7 illustrates that shakedown can be encouraged by applying a tensile bulk pre-load prior to the formation of the contact, or it can be inhibited with a compressive bulk pre-load. In fact, even though $\lambda = 3$ is 97% of the value of λ_2 for $f = 0.6$, Figure 9.7 demonstrates that shakedown can still be achieved by applying a bulk pre-load of $\sigma_{dm} \geqslant 1.5$. Although it is perhaps surprising how large a range of W_{SS} can be excited

using a bulk pre-load alone, it should not be surprising that tensile bulk pre-loads promote shakedown, while compressive bulk pre-loads inhibit shakedown. This is because all the optimal slip distributions shown in Figure 9.6 consist mainly of trailing-edge slip, and trailing-edge slip is excited by tensile bulk loads, whereas leading-edge slip is excited by compressive bulk loads.

Also notice that the effect of the bulk pre-load on W_{SS} 'saturates' quite quickly, and the largest effect occurs near $\sigma_{dm} \sim 1$. This saturation effect is not surprising because if a very large bulk pre-load is applied, there is simply a larger amount of slip in the first cycle, but the steady state remains largely unchanged. Also, note that the reason the largest effect on W_{SS} is observed for tensile pre-loads (and not about zero pre-load) is probably because a zero-to-peak loading regime was used for these simulations in which $\sigma(t)$ cycles about the mean bulk load of $\sigma_{\max}/2$ as opposed to a fully reversing loading regime in which $\sigma(t)$ cycles about zero.

9.5.5 Comparison of results

In their work, Björkman and Klarbring [1987] apply the shakedown-limit calculation to two example problems. They also perform a series of finite element transient simulations (until a steady state is reached) with a null initial displacement condition and determine the cyclic load above which shakedown does not occur, which can be regarded as an estimate of the shakedown limit. These authors then find that this estimate of the shakedown limit gives similar results to the optimization calculation in most cases. This result could imply that $\lambda_1 \sim \lambda_2$ for their example problems, so initial conditions are unimportant. Conversely, it may imply that a null initial displacement condition is a near optimal displacement condition for promoting shakedown.

However, it seems unlikely for a null initial displacement distribution to be an optimal distribution for both example problems considered by Björkman and Klarbring [1987]. Thus, this may be an indication that $\lambda_1 \sim \lambda_2$ for the two examples considered by these authors, and this would be in sharp contrast to the results shown in Figures 9.3-9.5, which

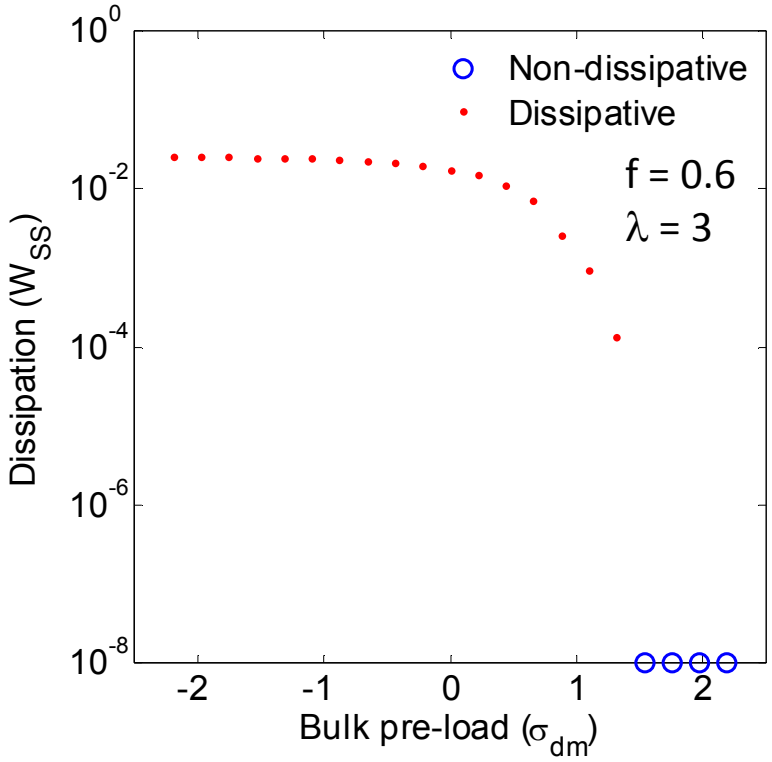


Figure 9.7: Influence of a (dimensionless) bulk pre-load, σ_{dm}, on the dimensionless energy dissipated per cycle in the steady state, W_{SS}, when $f = 0.6$ and $\lambda = 3$.

demonstrate that contacts may fail to shake down at load factors that are significantly below the shakedown limit. However, more transient simulations would need to be performed with different initial conditions to investigate the size of the conditional region for the contacts examined by these authors.

9.6 Discussion

The aim of the analysis presented in this chapter is to bring attention to a computationally efficient way for calculating the shakedown limit. This work has practical relevance because frictional contacts in engineering structures are very frequently subjected to a combination of static and cyclic loads resulting, at least transiently, in partial-slip conditions. Although it is well known that partial-slip plays a significant role in determining component performance [Barber, 2011, Farris et al., 2000], it is not always clear if a given configuration will result in these conditions after several load cycles. The mathematical-

programming approach due to Björkman and Klarbring [1987] provides an efficient way to determine the load range above which partial-slip is guaranteed.

The desirability of partial-slip is application specific and generally involves a trade-off between the positive effects of damping and the detrimental effects of fretting damage. Still, in many cases, the aim is to eliminate or minimize the extent of partial-slip and the fretting damage with which it is associated. For engineers with this aim, the results presented in Figures 9.3-9.5 may be a bit disheartening; these figures illustrate that the dissipative properties of practical complete contact geometries, which invariably exhibit significant levels of coupling even if both bodies are elastically similar, are highly sensitive to initial conditions. Moreover, partial-slip can persist in the steady state even for load factors that are well below the shakedown limit.

Although the calculation presented in §9.4 does not provide sufficient information to determine if a given load level is 'safe' (i.e. that shakedown is guaranteed), it does provide some very valuable information on how best to install components to encourage or discourage shakedown. This information is illustrated in Figure 9.6, which shows the optimal residual slip displacement distributions for three coefficients of friction. Notice that this information is automatically obtained from the calculation of λ_2. Furthermore, the nature of the optimal slip distribution for this particular configuration suggests that shakedown can be encouraged by pre-loading the contact with a bulk tension, and this is confirmed in Figure 9.7. This suggests that for some configurations there may be a simple procedure for encouraging or inhibiting shakedown in practice.

It is not clear how much can be inferred about other configurations from the small set of results presented in this chapter. This is meant both with regard to changes in the contact geometry and to changes in the types of loads that are applied, e.g. if a shear load were applied instead of a bulk load. However, one effect that is likely to apply more generally is that the size of the conditional region (see §9.5.3) will increase as the level of coupling increases. This is simply because shakedown is never conditional for the uncoupled case, so one might expect the size of the conditional region to be related

to how much coupling is present. This prompts the question as to whether the level of coupling can be quantified and used to predict the size of the conditional region. Some preliminary work has been done on this subject [Brake et al., 2013], but more work is required to determine the relationship between metrics of frictional coupling and the size of the conditional region.

9.7 Summary

In this chapter, a technique for efficiently calculating the shakedown limit of coupled, discrete frictional systems with a large number of degrees of freedom is examined. This calculation is then applied to an example complete contact problem between a square punch and an elastically similar half-plane, and the influence of initial conditions on steady-state energy dissipation is investigated. To do this, the contact is first discretized using the finite element method, and the static-reduction technique discussed in §7 is then employed. The calculated value of the shakedown limit is then compared to a series of transient simulations that were run until a steady state was reached. These results verify that the shakedown-limit calculation is correct. They also demonstrate that the dissipative properties of coupled complete contacts can be highly dependent on the initial residual slip displacement state. For instance, some transient simulations failed to shake down for load factors as low as 66%, 50%, and 62% of the shakedown limit when the coefficient of friction was 0.3, 0.6, and 0.9, respectively.

Chapter 10

Conclusion

In this chapter, a brief overview of this thesis is presented, and the main conclusions from each chapter are summarized. Some currently unsolved problems are then briefly discussed, and areas for future research are suggested.

10.1 Overview

This thesis has been concerned with analysing the fatigue performance of complete (i.e. sharp-edged) contacts. This has been done by obtaining an accurate description of the behaviour arising along the contact interface: that is, by quantifying the contact stresses and also the pattern of stick, partial-slip, and separation resulting from the applied loads. The philosophy underlying this approach is that a more accurate representation of contact behaviour will provide better parameters to correlate against fretting performance.

It is well known that fretting cracks almost invariably initiate at or near the edge of contact. However, even if a complete contact is fully adhered, a very large stress intensification arises near the contact edge, which results in a very complex and multi-axial stress state with very steep gradients. Thus, the region where fretting damage accumulates, which is the region that determines the fatigue life of the component, is very difficult to represent using a standard numerical approach, e.g. the finite element method. For this reason, in the earlier chapters of this thesis (i.e. §2-§6), an asymptotic approach is presented, which provides an accurate description of the contact edge.

A major benefit of this approach is that it provides parameters (i.e. the generalized

stress intensity factors, K_I, K_{II}) that incorporate a length scale in addition to a measure of the severity of the stress state. Since these parameters characterize the complex stress state near the contact edge, they should enable experimental data to be compared between different contacts, providing that the same contact angle is maintained. More experimental work is required to validate this approach (and this is discussed in §10.3), but the initial results are promising.

However, this asymptotic approach has some limitations: specifically, it cannot be applied when large amounts of slip or separation arise at the contact edge. Under some conditions, other asymptotic forms, e.g. those examined by Churchman [2006], Mugadu [2002], and Karuppanan [2007], may be able to be employed. However, in some cases, asymptotic forms simply cannot be applied, so it becomes necessary to take the behaviour of the entire contact interface into consideration. In these cases, the numerical approach that is presented in the later chapters (i.e. §7-§9) should be applied to obtain a description of contact behaviour.

The analysis in these later chapters is based on a sub-structuring technique, which improves the efficiency of many frictional analyses. For example, it enables transient frictional contact simulations to be run over 10 times more quickly than with commercial finite element software. It also enables such simulations to be solved without recourse to the contact solvers provided in commercial finite element codes by iteratively applying the friction law or by using linear-complementarity methods, which provides the user with greater control over the analysis. Perhaps most importantly, this approach greatly facilitates analyses of the steady-state behaviour of complete contacts that are subjected to cyclic loading. Specifically, it enables powerful optimization routines to be applied, which can determine the level of cyclic load above which frictional dissipation is guaranteed to persist in the steady state, and this has important implications for practical engineering components.

When designing real engineering components, there are a variety of distinct issues that must be addressed to ensure the success of the design. The primary concern is of course

safety, which is ensured by showing that all likely failure modes have been designed against. The asymptotic analysis presented in §2-6 can aid in this task by providing better parameters to use for predicting the fatigue performance of complete contacts. However, another important concern in many applications is the overall level of damping that is present in engineering assemblies, and this is very strongly influenced by the presence of frictional interfaces. The sub-structuring approach presented in §7-9 (and specifically the calculations in §9) provide very useful tools for investigating how much energy will be dissipated by frictional interfaces. Hence, these two approaches should be applied together to predict the behaviour and fatigue performance of complete contacts that are subject to fretting fatigue.

10.2 Conclusions

In §2-§6, an asymptotic approach for analysing complete contacts is presented. As this analysis is entirely based on Williams' solution for the stress state in a monolithic, semi-infinite notch, this solution is derived and discussed in detail in §2. An alternative set of parameters (i.e. d_0, G_0) due to Hills and Dini [2011] are also introduced, which take advantage of an internal length scale within Williams' (semi-infinite) solution. These parameters are very useful when analysing contacts subject to mixed-mode loading conditions, and they are used extensively in subsequent chapters.

The discussion of Williams' solution in §2 reveals that a singularity is implied in the elastic state of stress at the tip of sharp V-notches. Of course, singular stresses cannot actually arise in practice, so a zone of local plasticity will always be present in this region. In §3, it is assumed that this region of local plasticity results in a *process zone* where the non-linear damage processes that result in crack nucleation occur. Williams' solution is then used to characterize the mode-mixity *at the level of the process zone* (as this may be different from that suggested by the apparent mode-mixity of the applied loads) and also to assess whether the process zone is within the limits of small-scale yielding. The results that are obtained are universal in the sense that they can be applied to sharp V-notches of

any angle.

These calculations are then applied to experimental data reported in the literature on sharp V-notches that were subjected to monotonic, mixed-mode loading. The results reveal that most of these tests were carried out within 5% small-scale yielding, indicating that this asymptotic approach can be used to analyse these tests. In addition, it is found that although most of these tests were carried out in mixed-mode loading conditions, the process zone was usually mainly mode I or mixed-mode in character but was almost never mainly mode II in character. This is because the mode I and the mode II terms in Williams' solution are associated with different orders of singularity, so the resulting asymptotic stress field is not self-similar (unlike that for a crack). This suggests that it may be acceptable to neglect the mode II term in practice, especially if the notch angle is much less than $360°$.

In §4, it is shown that Williams' solution can be used to model the stress state at the edge of a complete contact when the interface is fully adhered. This follows from the fact that when there is no slip or separation along the contact interface, the displacement field across the interface is continuous, so the contact is equivalent to a notch from the point of view of linear elasticity. It is then shown that the minimum coefficient of friction required to prevent edge slip can be determined from Williams' solution and that this value depends only on the contact-edge angle. The combination of applied loads that results in the contact edges 'lifting off' (i.e. separating) is also determined. Note that this calculation only requires a calibration of the mode I generalized stress intensity factor to be obtained from a bilateral finite element model (i.e. a model in which a continuity of displacement condition is applied to the interface) and does not require the frictional interface to be taken into account.

These calculations are then applied to re-examine some fretting-fatigue data reported by Juoksukangas et al. [2013]. In their work, these authors plotted fretting results on a standard S-N plot, where the stress state at the contact edge was determined using elementary beam theory. However, the very high stresses that are present near the con-

tact edge are not well represented by those implied by beam theory, so these results are replotted using the amplitude of the mode I generalized stress intensity factor, ΔK_I. A discussion is then presented on how this alternative way of plotting fretting data enables inferences to be made about the fretting performance of other complete contacts of the same contact-edge angle, though more fretting tests must be carried out to validate this approach (e.g. those described in §10.3).

In §5, the three contact-edge responses that are most likely to arise in fretting tests on complete contacts are discussed. Each of these is then examined individually, and estimates of the extent of slip and/or separation are obtained in closed form for the full range of contact angles (i.e. for edge angles ranging from $0°$ to $180°$). These results remain valid provided that the slip/separation extent is not too large, such that it remains within the asymptotic region where Williams' solution controls the stress state. These calculations are then applied to the fretting-fatigue tests carried out by Juoksukangas et al. [2013] to characterize the contact-edge response during these fretting tests.

In §6, the results obtained in the previous chapters are combined, and plots in K_{II} vs. K_I space are developed, which enable fretting experiments to be quickly planned and analysed. This approach is then applied to three groups of fretting tests that were performed on complete contacts by Juoksukangas et al. [2013], Mugadu [2002], and Noraphaiphipaksa et al. [2013]. The results show that this asymptotic approach is most useful for the tests carried out by Juoksukangas et al. [2013] but is less helpful for the data reported by Mugadu [2002] and Noraphaiphipaksa et al. [2013]. This is because only a small to moderate amount of slip and separation occurred in Juoksukangas' tests, whereas very large amounts of slip and separation occurred during Mugadu's and Noraphaiphipaksa's tests. Therefore, the bilateral assumption cannot be applied to Mugadu's or Noraphaiphipaksa's tests, so only qualitative statements can be made about them based on this asymptotic approach. However, the loading regimes for which this asymptotic analysis could be applied can easily be found on these plots, and it is hoped that future tests will explore this load range.

In §7-§9, a different approach is presented for analysing complete contacts, which focuses on the contact interface as a whole. This analysis begins in §7 by presenting a numerical *sub-structuring* (i.e. *static-reduction*) procedure, which significantly reduces the number of degrees of freedom in two-dimensional finite element models of quasi-static frictional contact. This reduction is achieved by eliminating all the degrees of freedom from the stiffness matrix except those along the contact interface. The results of several numerical transient (i.e. 'marching-in-time') simulations are then presented, which reveal that the reduced model provides identical results to the commercial finite element software ABAQUS/CAE, but the solution is obtained over 10 times more quickly using the reduced model.

In §8, the transient evolution of the contact interface of a simple complete contact is examined using the static-reduction method discussed previously. The problem studied is that of a square elastic block pressed into an elastically similar half-plane and then subjected to a monotonically increasing shear load, which is raised until the entire contact interface starts to slide. This problem is examined in detail for a low-friction case and a high-friction case, and the results reveal that for moderate levels of shear force (i.e. less than 50% of the sliding condition), very different qualitative behaviour is observed for high and low coefficients of friction. However, for high shear forces (i.e. near to the sliding condition), the response for high and low friction coefficients is similar. Since fretting generally occurs during loading regimes that result in partial-slip and not full sliding, these results suggest that the fretting behaviour of complete contacts may undergo a qualitative change at a particular friction coefficient, which depends only on the contact-edge angle: namely, if the friction coefficient is above this value, partial-slip (and hence fretting damage) will not occur, whereas below this friction coefficient, partial-slip and fretting damage will occur.[1]

In §9, the steady-state behaviour of cyclically loaded complete contacts is examined. First, it is noted that complete contacts can *shake down*, such that the interface becomes

[1] However, this is only true if the applied loads do not result in edge separation; once separation occurs, slip will result irrespective of the coefficient of friction.

fully stuck after some number of load cycles (even if partial-slip occurs at the onset of loading). However, above a certain level of cyclic load, which is referred to as the *shakedown limit*, λ_2, it is impossible for shakedown to occur. An efficient numerical optimization approach is then presented for determining the shakedown limit for discrete elastic frictional contacts subjected to quasi-static loads. This technique is then applied to a large complete contact problem (i.e. the full model contained $\sim$50,000 degrees of freedom and $\sim$250 along the contact interface) to illustrate that this calculation can be applied to finite element models of realistic engineering components. The calculated value of the shakedown limit is then verified by comparing it to the results of over 1,000 numerical frictional contact simulations.

These numerical simulations are also used to examine the influence of initial conditions (i.e. the initial residual stress state) on the amount of frictional energy dissipation that continues to occur once a steady-state response has been achieved. The results demonstrate that steady-state frictional energy dissipation can be highly sensitive to the initial conditions. An optimal initial residual displacement state is also found, and it is shown that steady-state frictional dissipation (and hence fretting damage) can be significantly reduced by choosing a near optimal initial displacement condition. This result has important practical implications for the way in which components should be installed since this could significantly influence their fretting performance.

10.3 Future work

In this section, several topics that merit further investigation are briefly discussed.

Experimental work: As noted in §1, relatively few experimental studies of complete contacts have been reported in the literature. Some of these tests are examined in §6, but the analysis reveals that many of these were conducted using loading regimes that prohibit the use of the asymptotic approach developed here. Thus, the most important next step in the continuation of this work is to perform a series of experiments to validate the asymptotic approach.

The motive for developing the asymptotic approach presented in §2-6 is to be able to predict the fatigue performance of real engineering components based on fretting tests that can be performed with simplified test rigs. For this to be possible, it must be shown that the value of ΔK_I (or some other parameter from the asymptotic analysis) that results in material failure is geometry independent, so data obtained from simplified laboratory tests can be applied to real engineering components. Thus, experiments should be carried out to determine whether this is the case.

This can be determined by performing tests with identical (or at least very similar) trajectories in K_{II} vs. K_I space using test rigs that have very different contact geometries (but which use the same contact angle and material combination). This could be done simply by performing tests with the loading trajectory used by Juoksukangas et al. [2013], which is shown in Figures 5.1 and 6.2, but this time using the in-line test rig that is described in §6.3. If the same material is used for these tests, the data should fall along the same curve as the data obtained by Juoksukangas et al. [2013], which is shown in Figure 4.6.

A brief discussion of some other experiments that should be performed with the clamped-cantilever test rig is provided in §6.2.2. The aim of these experiments would be to determine whether partial-slip or closed-and-stuck conditions are more damaging for fatigue performance. In the experiments performed by Juoksukangas et al. [2013], the contact edge was subjected to fretting damage during one half of the load cycle, but it was closed and subjected to mainly mode I loading during the other half of the load cycle. Running the tests again with the slightly modified loading regime that is described in §6.2.2 would subject one of the contact edges to fretting conditions and the other to closed-and-stuck conditions. This would enable these tests to determine whether the damage due to fretting or 'notch fatigue' controls the fatigue life of the specimens simply by recording the corner from which failure initiates.

In addition to validating the asymptotic approach presented in §2-6, a series of experiments should also be performed to investigate frictional shakedown in general and

also to validate the calculation of the shakedown limit, λ_2, presented in §9. However, such experiments are far more challenging to carry out because they require the amount of partial-slip that occurs during the tests to be measured (and preferably for a full time history of the slip zones to be obtained). Thus, such tests would probably require the use of digital image correlation to estimate the contact slip displacements.

An example set of experiments might involve running several tests with the same initial condition but with different load factors and finding the load factor above which shakedown does not occur. The initial condition could then be varied for load factors near this value, e.g. by imposing an initial bulk tension or compression, and the sensitivity of the contact's steady-state response to initial conditions could be studied. The maximum load factor for which shakedown was achieved could then be compared to the shakedown limit obtained from the optimization calculation presented in §9. Note that the experimental data would be consistent with the calculations as long as shakedown always *fails to occur* for load factors greater than the calculated shakedown limit.

Dislocation solutions: While the calculations presented in §5 provide an estimate of the extent of slip and/or separation arising at the edge of complete contacts, the distributed dislocation technique can be employed to quantify these slip and separation zones exactly. Results have already been obtained by Churchman and Hills [2006c] and Paynter et al. [2010] for a 90° contact edge, but results for other angles could also be determined.

Dissimilar materials: All the analysis presented in this thesis is performed assuming that the materials in contact are elastically similar. Although this is a very common configuration, it is also very common for contacts to be formed between elastically dissimilar materials, e.g. brass on steel. As the analysis in this thesis is only rigorously applicable to elastically similar materials, future research should investigate how the results are affected by elastic dissimilarity. It is noted that all of the requisite tools for performing this analysis already exist: the asymptotic analysis that is based on Williams' solution must simply be performed again using the solution developed by Bogy [1968, 1971], and the

numerical analysis in §7-9 can already be performed for any configuration of interest.[2]

Although this analysis has not yet been performed, some remarks can be made regarding the extent to which the elastically similar results obtained here can be applied to contacts between dissimilar materials. The first thing to note about the current solution is that its most sensitive feature is the order of the elastic stress singularity at the contact edges. The order of the contact-edge singularity for the elastically dissimilar case can be determined from the characteristic equation of the solution due to Bogy [1968, 1971].[3] If the eigenvalues, λ_I, λ_{II}, remain similar, i.e. within about 10%, then the elastically similar results can probably be used to obtain an estimate of contact behaviour.

For example, if the cantilever tests examined in §4 were performed with steel specimens and brass contact pads with material properties of $E = 200GPa$ and $\nu = 0.3$ for the steel and of $E = 100GPa$ and $\nu = 0.34$ for the brass, Bogy's solution predicts that $(\lambda_I - 1, \lambda_{II} - 1) = (-0.4180, -0.0484)$, where E is Young's Modulus and ν is Poisson's ratio. Conversely, Williams' solution for the elastically similar case predicts that $(\lambda_I - 1, \lambda_{II} - 1) = (-0.4555, -0.0915)$. Hence, using brass (as opposed to steel) contact pads results in a change of 6.9% in λ_I and 4.7% in λ_{II} from the elastically similar case. Thus, the singularity becomes slightly weaker for the elastically dissimilar case, but the change is quite small. This suggests that the elastically similar results could probably be applied to the brass-on-steel case to get a rough sense of the contact behaviour that should be expected.

Also note that if two materials with different treatments (e.g. two steels with different values of surface roughness, fracture toughness, etc.) are being studied, the elastically similar results obtained here can be applied without introducing any error into the results. This is simply because these material properties do not enter the solution. Conversely, if

[2]This is a strength of the numerical approach presented in these chapters: it can be applied to any finite element model of contact between linear-elastic materials that are subjected to quasi-static loads. Indeed, the only way that elastic dissimilarity (and even component geometry) enters the solution is to the extent that it modifies the reduced contact stiffness matrix that is described in §7. All the solution techniques remain identical, regardless of the material combination that is chosen.

[3]Note that the characteristic equation will not be reproduced here since it is quite lengthy, but it can be found in [Churchman, 2006] §2 and [Karuppanan, 2007] §4.2 as well as in the original papers.

the two components have very different yield strengths, this will affect the shape of the plastic zone at the contact edge. However, this approach can still be used in most cases, and the yield strength of the component from which failure is expected to initiate should simply be used in the calculations.

Static reduction: The static-reduction technique presented in §7 is a very efficient tool for running frictional contact analyses, especially when long loading histories or multiple initial conditions are being examined. Currently this code can be applied to two-dimensional, linear-elastic contacts that are subjected to quasi-static loads. However, the current version of this code has some limitations, and it would be very beneficial to eliminate these in the future:

- The current code can only be applied to contacts whose interface lies approximately along a straight line, such that the normal and tangential directions are approximately the same for all the contact nodes. Thus, the code can only be applied to complete contacts, receding contacts, and incomplete contacts that can be modelled using half-plane theory. To relax this limitation, a transformation must simply be applied when calculating the $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ matrices in (7.30). This would enable more complex and realistic engineering components to be modelled using this approach, e.g. the dovetail and fir-tree joints in gas-turbine aero-engines.

- The current code can only handle two-dimensional contact problems, but it would be very beneficial to extend this approach to three-dimensional contact problems. The static-reduction procedure described in §7 can actually already be applied to three-dimensional contact problems with minimal modification. However, the Gauss-Seidel algorithm described in §7.5.1 would need to be significantly modified to handle the three-dimensional case. The main challenge would be to develop and implement the friction law in three dimensions since there are two shear directions in this case.

- The code could be extended to account for inertial effects.

- It may also be beneficial to develop a graphical user interface, so this solution procedure could be used more readily by the solid mechanics community. In addition, the code could be implemented in a faster coding environment than MATLAB, e.g. C++ or Fortran.

Frictional shakedown: An efficient optimization approach for calculating the shakedown limit, λ_2, is presented in §9. However, for coupled contacts (i.e. those for which relative tangential slip displacements modify the contact pressure distribution), when the amplitude of the cyclic load is less than λ_2, the initial conditions may determine whether dissipation persists in the steady state. For some problems, there exists another limit on the cyclic load, λ_1, below which dissipation is guaranteed to cease in the steady state.

No efficient techniques for calculating λ_1 have been found to date, so future research should attempt to develop such a calculation. Due to the difficulty of finding an efficient approach for determining the exact value of λ_1, the author has investigated methods for estimating λ_1 numerically. The only approach that has brought some degree of success thus far essentially involves performing a brute force search. This is done by searching for the 'worst' initial condition possible (in the sense that it prevents shakedown from occurring for the lowest load factor possible) by running numerous transient simulations for the specific contact under consideration while varying the load factor, λ, and the initial displacement condition, $\boldsymbol{v}(0)$.

The algorithm begins by running one transient simulation with values of λ and $\boldsymbol{v}(0)$ that are input by the user. The algorithm then proceeds to search for the load factor that marks the boundary between shakedown and cyclic slip *for this particular initial condition*, which will be denoted λ^*.[4] Note that this load factor is *not* λ_1 or λ_2, though

[4]This search for λ^* is performed with an algorithm that consists of a combination of the bisection method and the Newton-Raphson method. The reason a standard Newton-Raphson approach was not used can be understood by viewing Figure 9.1, which shows steady-state energy dissipation, W, as a function of load factor, λ. This figure illustrates that $W(\lambda)$ has a slope of zero for $\lambda < \lambda_1$ since all simulations shake down below this load factor. This causes the Newton-Raphson method to give unstable predictions (i.e. huge values) for the next estimate of λ^* since its prediction is based on the slope of $W(\lambda)$ near the current estimate of λ^*. Thus, if the previous simulation in the search for λ^* shook down, the bisection method is used to predict the next estimate of λ^* since its prediction does not depend on the derivatives of the function being searched and hence is more stable. Conversely, if the previous simulation did not shake down, the

it must lie within the interval $\lambda_1 \leqslant \lambda^* \leqslant \lambda_2$. This is because λ^* corresponds to a load factor at which shakedown has occurred, so $\lambda^* \leqslant \lambda_2$. Additionally, shakedown has failed to occur at $\lambda^* + \delta$, where δ is a small increment of the load factor, so $\lambda^* \geqslant \lambda_1$ (within a tolerance of δ).

Once the algorithm has found λ^*, this can be regarded as the first estimate of λ_1. In order to obtain a better estimate of λ_1, the initial condition needs to be optimized to search for the initial condition that results in the maximum amount of steady-state dissipation. To do this, the load factor is fixed at λ^*, and only $\boldsymbol{v}(0)$ is varied. In general, the search space for the worst $\boldsymbol{v}(0)$ involves varying N variables, where N is the number of contact nodes. However, optimizing against this many variables would be very computationally intensive, so in order to reduce the number of variables in this search, the initial displacement is assumed to be described by some function, $f(x)$, multiplied by a constant, C, such that $\boldsymbol{v}(0) = Cf(x)$, where x is a distance coordinate along the contact interface (as in Figure 8.1). The optimization is then performed by varying C within some bounds that are specified by the user.[5]

Notice that once the worst $\boldsymbol{v}(0)$ is found for the assumed displacement shape, shakedown will no longer occur at the current value of λ^* (if the simulation is run with $\boldsymbol{v}(0)$ specified as $Cf(x)$). Thus, the algorithm proceeds by performing a new search for an updated value of λ^*, which corresponds to the boundary between shakedown and cyclic slip *for this particular initial condition*, i.e. $\boldsymbol{v}(0) = Cf(x)$.

Once an updated value of λ^* is found, the algorithm searches for an even worse initial condition, which will cause the system to fail to shake down even when $\boldsymbol{v}(0) = Cf(x)$. This is done in a similar way as before except that the displacement shape is now assumed to be given by $\boldsymbol{v}(0) = Cf(x) + Dg(x)$, where $g(x)$ is a different displacement shape than $f(x)$. The term $Cf(x)$ is assumed to be given by the solution obtained from the previous search, and the algorithm searches for the value of D that maximizes W when $\lambda = \lambda^*$.

local slope will not be zero, so the Newton-Raphson prediction is used since it converges more quickly.

[5]The author performed this search using a fairly complex algorithm, which will not be described here, but simpler methods could also be used. For example, a scatter shot of 5-20 simulations could be performed within the search interval, and the $\boldsymbol{v}(0)$ that results in the largest value of W could then be selected.

This process of choosing a displacement shape and finding an updated value of λ^* can clearly be repeated as many times as the user desires. However, when using this technique, the author found that the law of diminishing returns tended to apply quite strongly, such that the estimate of λ_1 barely changed at all after the process had been applied with one or two displacement shapes. It is also noted that the author tended to use simple functions to describe the initial displacement condition, such as linear or sinusoidal distributions. As the displacement shape can be specified arbitrarily, the author also explored using random distributions of white noise, but these did not tend to have much of an effect on the solution.

Notice that this type of algorithm can also be applied to estimate λ_2 by searching for an initial condition that minimizes W for the current value of λ^*. Indeed, the author did just this to check that the optimization solution of λ_2 presented in §9 was correct, and much of the data shown in Figures 9.3-9.5 was obtained using this algorithm. Note that when searching for λ_2 as opposed to λ_1, the author sometimes used the optimal initial displacement distribution obtained from the optimization calculation of λ_2 (see §9.5.4) as the displacement shape for the $\boldsymbol{v}(0)$ search.

In addition, using this algorithm to search for λ_2 can provide a rough idea of the error associated with this approach in general (and hence in the estimate of λ_1). This is because the estimate of λ_2 can be compared with the exact result of the optimization calculation for λ_2 that is described in §9.4. However, it is important to note that if the user wishes to estimate the error in this algorithm's prediction of λ_1, the optimal $\boldsymbol{v}(0)$ from the optimization calculation of λ_2 should *not* be used as one of the displacement shapes. Otherwise, the error of this approach would be significantly underestimated (because this corresponds to assuming the optimal solution as the initial guess, which is quite unlikely to occur when searching for λ_1).

Of course, this approach still is quite computationally intensive since it involves running numerous transient simulations (typically 20-50) to obtain an approximate result. In fact, the time complexity of this approach is approximately proportional to the product

of running a single transient simulation and a constant factor to account for the number of simulations that must be performed. Even so, this approach still turns out to be more efficient than the exact solution technique due to Ahn et al. [2008] since the time complexity of their approach is exponential with the number of contact nodes. Thus, whereas the approach presented by Ahn et al. [2008] becomes infeasible for contacts with more than about 50 nodes, the author has applied this approximate approach to contacts with up to 250 nodes on a normal personal computer[6] and obtained an approximate solution within a couple of hours.

Coupling metric: Part of the motivation for being able to calculate λ_1 is that knowing both λ_1 and λ_2 would provide a means of quantifying how sensitive a given frictional system is to its initial conditions, i.e. how much the steady-state frictional energy dissipation can be influenced by changes to the initial residual stress state. This is important because it would provide information on whether the detailed way in which components are installed is important, for example. A different approach to this problem might be to attempt to quantify this sensitivity to initial conditions without reference to λ_1 or λ_2.

One way to do this is to recognize that contacts are only sensitive to initial conditions if they are coupled. A preliminary investigation by the author attempted to predict the sensitivity of frictional systems to their initial conditions by quantifying the level of coupling exhibited by the contact (see [Brake et al., 2013]), but this work has not been included in this thesis because conclusive results have not yet been obtained. It is noted that the coupling metrics proposed in [Brake et al., 2013] are:

$$\xi_{norm} = \frac{||\boldsymbol{B}||_2}{||\boldsymbol{A}||_2} \tag{10.1a}$$

$$\xi_{eig} = \frac{|\text{eig}(\boldsymbol{B})|_{\max}}{|\text{eig}(\boldsymbol{A})|_{\max}}, \tag{10.1b}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are sub-matrices of the reduced contact stiffness matrix, which are defined in (7.30). In other words, they are based on the ratio of some property of the sub-matrices

[6] These simulations were performed on the Dell computer that is described in §7.7.

of the reduced contact stiffness matrix: namely, the L2-norm, ξ_{norm}, and the eigenvalue with the largest modulus, ξ_{eig}.

Initial results suggest that these metrics are correlated with the sensitivity of steady-state frictional dissipation to initial conditions, but this requires deeper examination. In addition, it would be beneficial to develop coupling metrics that are based on the continuum formulation (as opposed to the discrete formulation, which the current metrics are based on). Most importantly, the relationship between any such metric and a contact's sensitivity to initial conditions must be investigated.

Inertial effects: All the analyses presented in this thesis have neglected inertial effects. Of course, such effects must be accounted for in many practical problems, so future research might investigate how to extend the analyses presented in this thesis to account for inertia. In addition, efforts should be made to determine how close the frequency of the cyclic load can be to the natural frequency of the component before the quasi-static analyses presented here should not be applied. For example, Brake and Hills [2014] report that, as a rule of thumb, a 10% error can be expected from a quasi-static contact solution when the loading frequency reaches 10% of the natural frequency of the system. These authors also report that once the loading frequency becomes much greater than this 10% figure, the error from a quasi-static solution becomes unacceptably large very quickly. However, it should be noted that these results were obtained for a particular geometry, and it is not clear how much can be inferred about other problems from these results.

Bibliography

J. Abad, J.M. Franco, R. Celorrio, and L. Lezáun. Design of experiments and energy dissipation analysis for a contact mechanics 3D model of frictional bolted lap joints. *Adv. Eng. Softw.*, 45(1):42–53, 2012. doi:10.1016/j.advengsoft.2011.09.021.

Y.J. Ahn and J.R. Barber. Response of frictional receding contact problems to cyclic loading. *Int. J. Mech. Sci.*, 50(10-11):1519–1525, October-November 2008. doi:10.1016/j.ijmecsci.2008.08.003.

Y.J. Ahn, E. Bertocchi, and J.R. Barber. Shakedown of coupled two-dimensional discrete frictional systems. *J. Mech. Phys. Solids*, 56(12):3433–3440, December 2008. doi:10.1016/j.jmps.2008.09.003.

J.E. Akin. The generation of elements with singularities. *Int. J. Numer. Meth. Eng.*, 10 (6):1249–1259, 1976. doi:10.1002/nme.1620100605.

G. Amontons. De la résistance causée dans les machines. *Mémoires de l'Académie Royale, A*, pages 257–282, 1699.

J.A. Araújo and D. Nowell. Mixed high low fretting fatigue of Ti6A14V: Tests and modelling. *Tribol. Int.*, 42:1276–1285, 2009. doi:10.1016/j.triboint.2009.04.024.

M.R. Ayatollahi, A.R. Torabi, and P. Azizi. Experimental and theoretical assessment of brittle fracture in engineering components containing a sharp V-notch. *Exp. Mech.*, 51 (6):919–932, 2011. doi:10.1007/s11340-010-9401-z.

N. Banerjee and D.A. Hills. Analysis of stick–slip and contact-edge behaviour in a simplified fretting fatigue test. *J. Strain Anal. Eng.*, 41(3):183–192, 2006. doi:10.1243/03093247JSA83.

N. Banerjee, D. Dini, and D.A. Hills. Frictional complete contacts subject to shear and bulk tension. *P. I. Mech. Eng. C-J. Mech.*, 222(12):2301–2309, 2008. doi:10.1243/09544062JMES1076.

J.R. Barber. *Elasticity*, volume 172. Springer, 2009.

J.R. Barber. Frictional systems subjected to oscillating loads. *Ann. Solid Struct. Mech.*, 2 (2-4):45–55, 2011. doi:10.1007/s12356-011-0017-5.

J.R. Barber, A. Klarbring, and M. Ciavarella. Shakedown in frictional contact problems for the continuum. *C. R. Mec.*, 336(1-2):34–41, January-February 2008. doi:10.1016/j.crme.2007.10.013.

J.R. Barber, M. Davies, and D.A. Hills. Frictional elastic contact with periodic loading. *Int. J. Solids. Struct.*, 48(13):2041–2047, 2011. doi:10.1016/j.ijsolstr.2011.03.008.

E. Bertocchi. *Selected topics on the plane elastic contact with friction*. PhD thesis, Department of Mechanical and Civil Engineering, University of Modena and Reggio Emilia, Italy, 2009. See Chapter 1.

G. Björkman and A. Klarbring. Shakedown and residual stresses in frictional systems. In *G M L Gladwell, H Ghonem and J Kalousek (Eds.), Contact Mechanics and Wear of Rail/Wheel Systems II: Proceedings of the 2nd International Symposium*, pages 27–39. 1987.

D.B. Bogy. Edge-bonded dissimilar orthogonal elastic wedges under normal and shear loading. *J. Appl. Mech.*, 35(3):460–466, 1968. doi:10.1115/1.3601236.

D.B. Bogy. Two edge-bonded elastic wedges of different materials and wedge angles under surface tractions. *J. Appl. Mech.*, 38(2):377–386, June 1971. doi:10.1115/1.3408786.

J. Boussinesq. Application des potentiels á l' étude de l' équilibre et du mouvement des solides élastiques. *Gauthier-Villars, Paris*, page 508, 1885.

M.R. Brake and D.A. Hills. Determination of the limits of quasi-static/rigid and dynamic solutions for problems with frictional interfaces. *Tribol. Int.*, 76:45–56, 2014. doi:10.1016/j.triboint.2013.09.008.

M.R. Brake, R.C. Flicek, and D.A. Hills. Development of a coupling metric to assess the shakedown limits for a contact interface. In *5th World Tribology Congress*, Torino, Italy, September 2013.

R. Bramhall. *Studies in Fretting Fatigue*. DPhil thesis, Department of Engineering Science, University of Oxford, UK, 1973.

R.K. Brayton, S.W. Director, G.D. Hachtel, and L. Vidigal. A new algorithm for statistical circuit design based on quasi-newton methods and function splitting. *IEEE T. Circuits Syst.*, 26(9):784–794, 1979. doi:10.1109/TCS.1979.1084701.

C. Cattaneo. Sul contatto di due corpi elastici: distribuzione locale degli sforzi. *Rendiconti dell'Accademia Nazionale dei Lincei*, 27:342–348,434–436,474–478, 1938.

C.Z. Cheng, Z.R. Niu, and N. Recho. Effect of non-singular stress on the brittle fracture of V-notched structure. *Int. J. Fracture.*, 174(2):127–138, 2012. doi:10.1007/s10704-012-9680-8.

C.M. Churchman. *Asymptotic analysis of complete contacts*. DPhil thesis, Department of Engineering Science, University of Oxford, UK, 2006.

C.M. Churchman and D.A. Hills. The edge dislocation in a three-quarter plane. Part II: Application to an edge crack. *Eur. J. Mech. A-Solid*, 25(3):389–396, 2006a. doi:10.1016/j.euromechsol.2005.12.001.

C.M. Churchman and D.A. Hills. General results for complete contacts subject to oscillatory shear. *J. Mech. Phys. Solids*, 54(6):1186–1205, June 2006b. doi:10.1016/j.jmps.2005.12.005.

C.M. Churchman and D.A. Hills. Slip zone length at the edge of a complete contact. *Int. J. Solids. Struct.*, 43(7-8):2037–2049, April 2006c. doi:10.1016/j.ijsolstr.2005.06.099.

C.M. Churchman, A.M. Korsunsky, and D.A. Hills. The edge dislocation in a three-quarter plane. Part I: Influence functions. *Eur. J. Mech. A-Solid*, 25(1):42–50, 2006. doi:10.1016/j.euromechsol.2005.06.010.

M. Ciavarella. The generalized Cattaneo partial slip plane contact problem. I - Theory. *Int. J. Solids. Struct.*, 35(18):2349–2362, 1998a. doi:10.1016/S0020-7683(97)00154-6.

M. Ciavarella. The generalized Cattaneo partial slip plane contact problem. II - Examples. *Int. J. Solids. Struct.*, 35(18):2363–2378, 1998b. doi:10.1016/S0020-7683(97)00155-8.

M. Ciavarella and G. Meneghetti. On fatigue limit in the presence of notches: classical vs. recent unified formulations. *Int. J. Fatigue*, 26(3):289–298, 2004. doi:10.1016/S0142-1123(03)00106-3.

M. Ciavarella, D.A. Hills, and G. Monno. The influence of rounded edges on indentation by a flat punch. *P. I. Mech. Eng. C-J. Mech.*, 212(4):319–327, 1998. doi:10.1243/0954406981521259.

M. Comninou. Stress singularity at a sharp edge in contact problems with friction. *J. Appl. Math. Phys.*, 27(4):493–499, 1976. doi:10.1007/BF01594906.

C.A. Coulomb. Théorie des machines simples, en ayant regard au frottement de leurs parties, et a la roideur des cordages. *Mem. Math. Phys., X, Paris*, pages 161–342, 1785.

B.A. Cowles. High cycle fatigue in aircraft gas turbines – an industry perspective. *Int. J. Fracture.*, 80(2-3):147–163, 1989. doi:10.1007/BF00012667.

J. Ding, W.S. Sum, R. Sabesan, S.B. Leen, I.R. McColl, and E.J. Williams. Fretting fatigue predictions in a complex coupling. *Int. J. Fatigue*, 29(7):1229–1244, 2007. doi:10.1016/j.ijfatigue.2006.10.017.

D. Dini and D.A. Hills. Frictional energy dissipation in a rough hertzian contact. *J. Tribol.*, 131(2):021401, 2009. doi:10.1115/1.3063697.

D. Dini, C. Churchman, R. Rajasekaran, and D.A. Hills. A correlation of the process zone properties in complete, incomplete and almost complete fretting contacts. *Int. J. Mech. Sci.*, 46(3):491–508, 2004. doi:10.1016/j.ijmecsci.2004.03.004.

J. Dundurs. Discussion of "Edge-bonded dissimilar or orthogonal elastic wedges under normal and shear loading". *J. Appl. Mech.*, 36:650–652, 1969. doi:10.1115/1.3564739.

J. Dundurs and M. Stippes. Role of elastic constants in certain contact problems. *J. Appl. Mech.*, 37(4):965–970, 1970. doi:10.1115/1.3408725.

M.L. Dunn, W. Suwito, and S. Cunningham. Fracture initiation at sharp notches: Correlation using critical stress intensities. *Int. J. Solids. Struct.*, 34(29):3873–3883, 1997a. doi:10.1016/S0020-7683(96)00236-3.

M.L. Dunn, W. Suwito, S. Cunningham, and C.W. May. Fracture initiation at sharp notches under mode I, mode II, and mild mixed mode loading. *Int. J. Fracture.*, 84 (4):367–381, 1997b. doi:10.1023/A:1007346203407.

F.M. Eden, W.N. Rose, and F.L. Cunningham. The endurance of metals: Experiments on rotating beams at University College, London. *Proc. Inst. Mech. Eng.*, 4:839–974, 1911.

M.H. El Haddad, K.N. Smith, and T.H. Topper. Fatigue crack propagation of short cracks. *J. Eng. Mater.-T ASME*, 101(1):42–46, 1979a. doi:10.1115/1.3443647.

M.H. El Haddad, T.H. Topper, and K.N. Smith. Prediction of non propagating cracks. *Eng. Fract. Mech.*, 11(3):573–584, 1979b. doi:10.1016/0013-7944(79)90081-X.

M. Elices, G.V. Guinea, J. Gomez, and J. Planas. The cohesive zone model: Advantages, limitations and challenges. *Eng. Fract. Mech.*, 69(2):137–163, 2002. doi:10.1016/S0013-7944(01)00083-2.

K. Endo and H. Goto. Initiation and propagation of fretting fatigue cracks. *Wear*, 38(2): 311–324, 1976. doi:10.1016/0043-1648(76)90079-X.

K. Endo and H. Goto. Reply to comments on "Initiation and propagation of fretting fatigue cracks". *Wear*, 43(2):269–270, 1977. doi:10.1016/0043-1648(77)90122-3.

T.N. Farris, M.P. Szolwinski, and G. Harish. Fretting in aerospace structures and materials. In *D.W. Hoeppner, V. Chandrasekaran, C.B. Elliott (Eds.), Fretting Fatigue: Current Technology and Practice, ASTM STP*, volume 1367, pages 523–537. ASTM, West Conshohocken, PA, 2000.

A. Fatemi and D.F. Socie. A critical plane approach to multiaxial fatigue damage including out-of-phase loading. *Fatigue. Fract. Eng. M.*, 11(3):149–165, 1988. doi:10.1111/j.1460-2695.1988.tb01169.x.

A.J. Fenner and J.E. Field. La fatigue dans les conditions de frottement. *Rev. Metall*, 55: 478–485, 1958.

B. Fine and G. Rosenberger. *The Fundamental Theorem of Algebra*. Springer Undergraduate Texts in Mathematics and Technology. Springer New York, 1997. ISBN 9780387946573.

A. Flamant. Sur la répartition des pressions dans un solide rectangulaire chargé transversalement. *Compte. Rendu. Acad. Sci. Paris*, 114:1465, 1892.

R. Flicek, D.A. Hills, and D. Dini. Progress in the application of notch asymptotics to the understanding of complete contacts subject to fretting fatigue. *Fatigue. Fract. Eng. M.*, 36(1):56–64, January 2013. doi:10.1111/j.1460-2695.2012.01694.x.

R.C. Flicek. Static reduction of finite element models for increased computation efficiency. UTC Report 340, University of Oxford, 2013.

R.C. Flicek, D.A. Hills, and D. Dini. Refinements in the characterisation of mode-mixity and small scale yielding at sharp notch roots. *Eng. Fract. Mech.*, 126:73–86, August 2014. doi:10.1016/j.engfracmech.2014.04.022.

R.C. Flicek, D.A. Hills, J.R. Barber, and D. Dini. Determination of the shakedown limit for large, discrete frictional systems. *Eur. J. Mech. A-Solid*, 49:242–250, January-February 2015a. doi:10.1016/j.euromechsol.2014.08.001.

R.C. Flicek, D.A. Hills, and D. Dini. Sharp edged contacts subject to fretting: A description of corner behaviour. *Int. J. Fatigue*, 71:26–34, February 2015b. doi:10.1016/j.ijfatigue.2014.02.015.

R.C. Flicek, R. Ramesh, and D.A. Hills. A complete frictional contact: The transition from normal load to sliding. *Int. J. Eng. Sci.*, 92:18–27, July 2015c. doi:10.1016/j.ijengsci.2015.03.006.

E.E. Gdoutos and P.S. Theocaris. Stress concentrations at the apex of a plane indenter acting on an elastic half plane. *J. Appl. Mech.*, 42(3):688–692, September 1975. doi:10.1115/1.3423663.

M.C. Gean and T.N. Farris. Elevated temperature fretting fatigue of Ti-17 with surface treatments. *Tribol. Int.*, 42(9):1340–1345, 2009. doi:10.1016/j.triboint.2009.04.027.

L. Gorbatikh, B. Nuller, and M. Kachanov. Sliding on cracks with non-uniform frictional characteristics. *Int. J. Solids. Struct.*, 38(42):7501–7524, 2001. doi:10.1016/S0020-7683(01)00011-7.

J.A. Greenwood and J.B.P. Williamson. Contact of nominally flat surfaces. *Proc. Roy. Soc. A.*, 295(1442):300–319, 1966. doi:10.1098/rspa.1966.0242.

Z. Handzel-Powierza, T. Klimczak, and A. Polijaniuk. On the experimental verification of the Greenwood-Williamson model for the contact of rough surfaces. *Wear*, 154(1): 115–124, 1992. doi:10.1016/0043-1648(92)90247-6.

H.R. Hertz. Über die Beruehrung fester, elastischer Körper. *Journal für die reine und angewandte Mathematik*, 92:156–171, 1882.

D.A. Hills. Mechanics of fretting fatigue. *Wear*, 175(1):107–113, 1994. doi:10.1016/0043-1648(94)90173-2.

D.A. Hills and D. Dini. What level of friction guarantees adhesion in a complete contact? *J. Strain Anal. Eng.*, 39(5):549–551, January 2004. doi:10.1243/0309324041896425.

D.A. Hills and D. Dini. Characteristics of the process zone at sharp notch roots. *Int. J. Solids. Struct.*, 48(14-15):2177–2183, July 2011. doi:10.1016/j.ijsolstr.2011.03.023.

D.A. Hills and R.C. Flicek. A discussion of "Numerical and experimental investigations on fretting fatigue: Relative slip, crack path, and fatigue life" by N. Noraphaiphipaksa, C. Kanchanomai, and Y. Mutoh. *Eng. Fract. Mech.*, 113:52–55, January 2015. doi:10.1016/j.engfracmech.2014.10.024.

D.A. Hills, D. Nowell, and J.J. O'Connor. On the mechanics of fretting fatigue. *Wear*, 125(1):129–146, 1988. doi:10.1016/0043-1648(88)90198-6.

D.A. Hills, P.A. Kelly, D.N. Dai, and A.M. Korsunsky. *Solution of crack problems: The distributed dislocation technique*, volume 44. Springer, 1996.

D.A. Hills, R.J.H. Paynter, and D. Dini. An overview of the quantification of fretting fatigue lives of complete contacts. *Eng. Fract. Mech.*, 80:3–12, January 2012a. doi:10.1016/j.engfracmech.2011.06.008.

D.A. Hills, A. Thaitirarot, J.R. Barber, and D. Dini. Correlation of fretting fatigue experimental results using an asymptotic approach. *Int. J. Fatigue*, 43:62–75, 2012b. doi:10.1016/j.ijfatigue.2012.02.006.

D.A. Hills, R.C. Flicek, and D. Dini. Sharp contact corners, fretting and cracks. *Frattura ed Integritá Strutturale (Fracture and Structural Integrity)*, 25:27–35, July 2013. doi:10.3221/igf-esis.25.05.

D.A. Hills, R.C. Flicek, and D. Dini. A discussion of: Development of a complete contact fretting test device by J Juoksukangas et al. *P. I. Mech. Eng. J-J. Eng.*, 228(1):123–126, January 2014. doi:10.1177/1350650113500946.

D.W. Hoeppner. Comments on "Initiation and propagation of fretting fatigue cracks". *Wear*, 43(2):267–268, 1977. doi:10.1016/0043-1648(77)90121-1.

D.W. Hoeppner and G.L. Goss. A fretting-fatigue damage threshold concept. *Wear*, 27 (1):61–70, 1974. doi:10.1016/0043-1648(74)90084-2.

J. Jäger. A new principle in contact mechanics. *J. Tribol.*, 120(4):677–684, 1998. doi:10.1115/1.2833765.

Y.H. Jang and J.R. Barber. Frictional energy dissipation in materials containing cracks. *J. Mech. Phys. Solids*, 59(3):583–594, 2011. doi:10.1016/j.jmps.2010.12.010.

J. Juoksukangas, A. Lehtovaara, and A. Mäntylä. Development of a complete contact fretting test device. *P. I. Mech. Eng. J-J. Eng.*, 227:570–578, 2013. doi:10.1177/1350650112466162.

S. Karuppanan. *Analysis of incomplete and complete contacts in sliding and partial slip*. DPhil thesis, Department of Engineering Science, University of Oxford, UK, 2007.

S. Karuppanan and D.A. Hills. Frictional complete contacts between elastically similar bodies subject to normal and shear load. *Int. J. Solids. Struct.*, 45(17):4662–4675, August 2008a. doi:10.1016/j.ijsolstr.2008.04.006.

S. Karuppanan and D.A. Hills. Sliding of general frictional complete contacts. *Int. J. Mech. Sci.*, 50(9):1404–1410, September 2008b. doi:10.1016/j.ijmecsci.2008.07.001.

S. Karuppanan, C.M. Churchman, D.A. Hills, and E. Giner. Sliding frictional contact between a square block and an elastically similar half-plane. *Eur. J. Mech. A-Solid*, 27 (3):443–459, 2008. doi:10.1016/j.euromechsol.2007.09.001.

R. Khadem and J.J. O'Connor. Adhesive or frictionless compression of an elastic rectangle between two identical elastic half-spaces. *Int. J. Eng. Sci.*, 7(2):153–168, 1969a. doi:10.1016/0020-7225(69)90054-8.

R. Khadem and J.J. O'Connor. Axial compression of an elastic circular cylinder in contact with two identical elastic half spaces. *Int. J. Eng. Sci.*, 7(8):785–800, 1969b. doi:10.1016/0020-7225(69)90061-5.

J.K. Kim and S.B. Cho. A unified brittle fracture criterion for structures with sharp V-notches under mixed mode loading. *J. Mech. Sci. Technol.*, 22(7):1269–1278, 2008. doi:10.1007/s12206-008-0315-y.

H. Kitagawa and S. Takahashi. Applicability of fracture mechanics to very small cracks or the cracks in the early stage. In *Second International Conference on Mechanical Behavior of Materials. ASM, Metals Park, Ohio. 1976, 627-631*, 1976.

A. Klarbring, M. Ciavarella, and J.R. Barber. Shakedown in elastic contact problems with Coulomb friction. *Int. J. Solids. Struct.*, 44(25-26):8355–8365, December 2007. doi:10.1016/j.ijsolstr.2007.06.013.

K. Knothe. History of wheel/rail contact mechanics: From Redtenbacher to Kalker. *Vehicle Syst. Dyn.*, 46(1-2):9–26, 2008. doi:10.1080/00423110701586469.

W.T. Koiter. General theorems for elastic-plastic solids. *Progress in Solid Mechanics (Editors: I.N. Sneddon and R. Hill)*, 4:167–221, 1960.

Y. Kondo and M. Bodai. The fretting fatigue limit based on local stress at the contact edge. *Fatigue. Fract. Eng. M.*, 24(12):791–801, 2001. doi:10.1046/j.1460-2695.2001.00449.x.

Y. Kondo, C. Sakae, M. Kubota, T. Nagasue, and S. Sato. Fretting fatigue limit as a short crack problem at the edge of contact. *Fatigue. Fract. Eng. M.*, 27(5):361–368, 2004. doi:10.1111/j.1460-2695.2004.00750.x.

Y. Kondo, C. Sakae, M. Kubota, and K. Yanagihara. Non-propagating crack behaviour at giga-cycle fretting fatigue limit. *Fatigue. Fract. Eng. M.*, 28(6):501–506, 2005. doi:10.1111/j.1460-2695.2005.00896.x.

S.S. Law, Z.M. Wu, and S.L. Chan. Analytical model of a slotted bolted connection element and its behaviour under dynamic load. *J. Sound. Vib.*, 292(3):777–787, 2006. doi:10.1016/j.jsv.2005.09.028.

P. Lazzarin and R. Zambardi. A finite-volume-energy based approach to predict the static and fatigue behavior of components with sharp V-shaped notches. *Int. J. Fracture.*, 112 (3):275–298, 2001. doi:10.1023/A:1013595930617.

P. Lazzarin, F. Berto, M. Elices, and J. Gómez. Brittle failures from U-and V-notches in mode I and mixed, I+ II, mode: A synthesis based on the strain energy density averaged on finite-size volumes. *Fatigue. Fract. Eng. M.*, 32(8):671–684, 2009. doi:10.1111/j.1460-2695.2009.01373.x.

P. Lazzarin, F. Berto, and M. Zappalorto. Rapid calculations of notch stress intensity factors based on averaged strain energy density from coarse meshes: Theoretical bases and applications. *Int. J. Fatigue*, 32(10):1559–1567, October 2010. doi:10.1016/j.ijfatigue.2010.02.017.

D. Leguillon. Strength or toughness? A criterion for crack onset at a notch. *Eur. J. Mech. A-Solid*, 21(1):61–72, 2002. doi:10.1016/S0997-7538(01)01184-6.

L. Limmer, D. Nowell, and D.A. Hills. A combined testing and modelling approach to the prediction of the fretting fatigue performance of splined shafts. *P. I. Mech. Eng. G-J. Aer.*, 215(2):105–112, 2001. doi:10.1243/0954410011531808.

Y. Liu, Y.H. Jang, and J.R. Barber. Finite element implementation of an eigenfunction solution for the contact pressure variation due to wear. *Wear*, 309(1):134–138, 2014. doi:10.1016/j.wear.2013.11.004.

J.I. McCool. Comparison of models for the contact of rough surfaces. *Wear*, 107(1): 37–60, 1986. doi:10.1016/0043-1648(86)90045-1.

J.R. McDowell. Fretting corrosion tendencies of several combinations of materials. In *Symposium on Fretting Corrosion, ASTM STP 144*, pages 24–39, Philadelphia, 1953.

J.R. McDowell, H.H. Uhlig, W.D. Tierney, A. McClellan, and O.J. Horger. Fretting corrosion. *Anti-Corrosion Methods and Materials*, 1(7):247–252, 1954.

E. Melan. Theorie statisch unbestimmter systeme aus ideal-plastischem baustoff. *Sitzungsber. Akad. Wiss. Wien, Abt. 2A*, 145:195–218, 1936.

R.D. Mindlin. Compliance of elastic bodies in contact. *J. Appl. Mech.*, 16:259–268, 1949.

R.D. Mindlin and H. Deresiewicz. Elastic spheres in contact under varying oblique forces. *J. Appl. Mech.*, 21:327–344, 1953.

A. Mugadu and D.A. Hills. A generalised stress intensity approach to characterising the process zone in complete fretting contacts. *Int. J. Solids. Struct.*, 39(5):1327–1335, 2002. doi:10.1016/S0020-7683(01)00210-4.

A. Mugadu and D.A. Hills. The evolution of the process zone when a complete contact is subject to cyclically varying loads. *Int. J. Solids. Struct.*, 40(17):4417–4435, 2003. doi:10.1016/S0020-7683(03)00184-7.

A.E. Mugadu. *Studies in Fretting Fatigue of Complete Contacts*. DPhil thesis, Department of Engineering Science, University of Oxford, UK, 2002.

Tristan Needham. *Visual Complex Analysis*. Clarendon Press Oxford, 1997.

T. Nicholas. Critical issues in high cycle fatigue. *Int. J. Fatigue*, 21(1):S221–S231, September 1999. doi:10.1016/S0142-1123(99)00074-2.

K. Nishioka and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 2, Fretting fatigue testing machine and some test results. *Bulletin of the JSME*, 12(50): 180–187, 1969a. doi:10.1299/jsme1958.12.180.

K. Nishioka and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 3, Some phenomena and mechanisms of surface cracks. *Bulletin of the JSME*, 12(51): 397–407, 1969b. doi:10.1299/jsme1958.12.397.

K. Nishioka and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 4, The effect of mean stress. *Bulletin of JSME*, 12(51):408–414, 1969c. doi:10.1299/jsme1958.12.408.

K. Nishioka and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 5, The effect of relative slip amplitude. *Bulletin of the JSME*, 12(52):692–697, 1969d. doi:10.1299/jsme1958.12.692.

K. Nishioka and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 6, Effects of contact pressure and hardness of materials. *Bulletin of the JSME*, 15(80): 135–144, 1972. doi:10.1299/jsme1958.15.135.

K. Nishioka, S. Nishimura, and K. Hirakawa. Fundamental investigations of fretting fatigue: Part 1, On the relative slip amplitude of press-fitted axle assemblies. *Bulletin of the JSME*, 11(45):437–445, 1968. doi:10.1299/jsme1958.11.437.

N. Noraphaiphipaksa, C. Kanchanomai, and Y. Mutoh. Numerical and experimental investigations on fretting fatigue: Relative slip, crack path, and fatigue life. *Eng. Fract. Mech.*, 112:58–71, 2013. doi:10.1016/j.engfracmech.2013.10.007.

D. Nowell. *An analysis of fretting fatigue*. DPhil thesis, Department of Engineering Science, University of Oxford, UK, 1988.

D. Nowell and D.A. Hills. Mechanics of fretting fatigue tests. *Int. J. Mech. Sci.*, 29(5): 355–365, 1987. doi:10.1016/0020-7403(87)90117-2.

R.J.H. Paynter, D.A. Hills, and D. Dini. Separation and slip at the edge of a complete contact: An asymptotic solution. *Int. J. Solids. Struct.*, 47(18):2613–2619, September 2010. doi:10.1016/j.ijsolstr.2010.05.026.

M.J.D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis*, pages 144–157. Springer, 1978. doi:10.1007/BFb0067703.

E. Priel, A. Bussiba, I. Gilad, and Z. Yosibash. Mixed mode failure criteria for brittle elastic V-notched structures. *Int. J. Fracture.*, 144(4):247–265, 2007. doi:10.1007/s10704-007-9098-x.

N. Pugno, M. Ciavarella, P. Cornetti, and A. Carpinteri. A generalized paris law for fatigue crack growth. *J. Mech. Phys. Solids*, 54(7):1333–1349, 2006. doi:10.1016/j.jmps.2006.01.007.

R. Rajasekaran and D. Nowell. Fretting fatigue in dovetail blade roots: Experiment and analysis. *Tribol. Int.*, 39:1277–1285, 2006. doi:10.1016/j.triboint.2006.02.044.

F.J. Redtenbacher. *Die Gesetze des Locomotiv-Baues*. F. Basermann, Mannheim, 1855.

C. Ruiz, P.H.B. Boddington, and K.C. Chen. An investigation of fatigue and fretting in a dovetail joint. *Exp. Mech.*, 24(3):208–217, 1984. doi:10.1007/BF02323167.

M. Sabsabi, E. Giner, and F.J. Fuenmayor. Experimental fatigue testing of a fretting complete contact and numerical life correlation using X-FEM. *Int. J. Fatigue*, 33(6): 811–822, June 2011. doi:10.1016/j.ijfatigue.2010.12.012.

A. Sackfield, A. Mugadu, and D.A. Hills. The influence of an edge radius on the local stress field at the edge of a complete fretting contact. *Int. J. Solids. Struct.*, 39(17): 4407–4420, 2002. doi:10.1016/S0020-7683(02)00337-2.

W. Schütz. A history of fatigue. *Eng. Fract. Mech.*, 54(2):263–300, 1996. doi:10.1016/0013-7944(95)00178-6.

A. Seweryn. Brittle fracture criterion for structures with sharp notches. *Eng. Fract. Mech.*, 47(5):673–681, 1994. doi:10.1016/0013-7944(94)90158-9.

A. Seweryn and A. Łukaszewicz. Verification of brittle fracture criteria for elements with V-shaped notches. *Eng. Fract. Mech.*, 69(13):1487–1510, 2002. doi:10.1016/S0013-7944(01)00138-2.

A. Seweryn and K. Molski. Elastic stress singularities and corresponding generalized stress intensity factors for angular corners under various boundary conditions. *Eng. Fract. Mech.*, 55(4):529–556, 1996. doi:10.1016/S0013-7944(96)00035-5.

A. Seweryn, S. Poskrobko, and Z. Mróz. Brittle fracture in plane elements with sharp notches under mixed-mode loading. *J. Eng. Mech.*, 123(6):535–543, 1997. doi:10.1061/(ASCE)0733-9399(1997)123:6(535).

K.N. Smith, T.H. Topper, and P. Watson. A stress-strain function for the fatigue of metals. *J. Mater.*, 5:767–778, 1970.

W. Suwito, M.L. Dunn, and S.J. Cunningham. Fracture initiation at sharp notches in single crystal silicon. *J. Appl. Phys.*, 83(7):3574–3582, 1998. doi:10.1063/1.366574.

W. Suwito, M.L. Dunn, S.J. Cunningham, and D.T. Read. Elastic moduli, strength, and fracture initiation at sharp notches in etched single crystal silicon microstructures. *J. Appl. Phys.*, 85(7):3519–3534, 1999. doi:10.1063/1.369711.

P.S. Symonds. Shakedown in continuous media. *J. Appl. Mech.*, 18:85–89, 1951.

M.P. Szolwinski and T.N. Farris. Mechanics of fretting fatigue crack formation. *Wear*, 198(1):93–107, 1996. doi:10.1016/0043-1648(96)06937-2.

A. Thaitirarot, R.C. Flicek, D.A. Hills, and J.R. Barber. The use of static reduction in the finite element solution of two-dimensional frictional contact problems. *P. I. Mech. Eng. C-J. Mech.*, 228(9):1474–1487, June 2014. doi:10.1177/0954406213509086.

G.A. Tomlinson. The rusting of steel surfaces in contact. *Proc. R. Soc. Lond. A*, 115(771): 472–483, 1927. doi:10.1098/rspa.1927.0104.

G.A. Tomlinson, P.L. Thorpe, and H.J. Gough. An investigation of the fretting corrosion of closely fitting surfaces. *Proc. Inst. Mech. Eng.*, 141(1):223–249, 1939. doi:10.1243/PIME_PROC_1939_141_034_02.

E.J. Warlow-Davies. Fretting corrosion and fatigue strength: Brief results of preliminary experiments. *Prc. I. Mech. Eng.*, 146(1):32–38, 1941. doi:10.1243/PIME_PROC_1941_146_012_02.

R.B. Waterhouse. A historical introduction to fretting fatigue. In M. Helmi Attia and R. B. Waterhouse, editors, *Standardization of fretting fatigue test methods and equipment, ASTM STP 1159*, pages 8–9. American society for testing and materials, Philadelphia, 1992.

M.L. Williams. Stress singularities resulting from various boundary conditions in angular plates in extension. *J. Appl. Mech.*, 19(4):526–528, 1952.

N.S. Xi, P.D. Zhong, H.Q. Huang, H. Yan, and C.H. Tao. Failure investigation of blade and disk in first stage compressor. *Eng. Fail. Anal.*, 7(6):385–392, 2000. doi:10.1016/S1350-6307(99)00045-X.

Appendix A

Additional material

A.1 Additional material for §3

A.1.1 Mode-mixity calculation

The process zone at the tip of a notch subjected to mixed-mode loading, $r_p(\theta)$, can be estimated by equating (3.1) and (3.2) and substituting in stresses from (3.3) and (2.27). Simplifying and manipulating the resulting expression gives

$$3\left(\frac{k}{G_0}\right)^2 = \left(\frac{r_p}{d_0}\right)^{2(\lambda_I-1)} p_I(\theta) + \left(\frac{r_p}{d_0}\right)^{2(\lambda_{II}-1)} p_{II}(\theta) + \left(\frac{r_p}{d_0}\right)^{(\lambda_I+\lambda_{II}-2)} p_M(\theta). \tag{A.1}$$

The left side of this equation accounts for the strength of the applied load, and the right side specifies the size and shape of the process zone based on the contribution of a pure mode I, a pure mode II, and a mixed-mode term. The $p_\xi(\theta)$ functions in this expression, where $\xi \in \{I, II, M\}$, account for the θ-variation of the mode I, mode II, and mixed-mode terms, respectively. Note that the $p_\xi(\theta)$ functions are comprised of combinations of the angular functions given in (2.22) and (2.23), and they account for the influence of Poisson's ratio, ν.

Similarly, the process zone at the tip of a notch subjected to pure mode n loading can

be obtained as above except by using stresses from (3.4) instead of from (2.27). Simplifying and manipulating the resulting expression as before gives

$$\frac{3}{(mS_n)^2} \equiv 3\left(\frac{k}{K_n a^{\lambda_n - 1}}\right)^2 = \left(\frac{r_p^n}{a}\right)^{2(\lambda_n - 1)} p_n(\theta), \tag{A.2}$$

where $r_p^n(\theta)$ defines the shape of the pure mode n process zone, where $n \in \{I, II\}$.

To calculate the correlation coefficient between a mixed-mode process zone and a pure mode I process zone, ρ_I, and a pure mode II process zone, ρ_{II}, the shape of the pure mode I, II process zones, $r_p^I(\theta), r_p^{II}(\theta)$, must first be calculated using (A.2). As these are self-similar, this can be done simply by fixing $mS_n = 1$. Similarly, the shape of the mixed-mode process zone, $r_p(\theta)$, can be obtained by fixing G_0/k and using (A.1). A convenient way to perform this calculation is first to discretize $r_p(\theta), r_p^I(\theta), r_p^{II}(\theta)$ with respect to θ and then to calculate ρ_I, ρ_{II} with these discretized process zones. The values of G_0/k corresponding to particular values of ρ_I, ρ_{II} must then be determined, so a numerical search algorithm must be employed. For the results in §3, an algorithm that used a combination of the bisection method and the Newton-Raphson method was used for this purpose.

The typical convergence of this technique, i.e. the variation in the value of G_0/k corresponding to any given value of ρ_n with the number of points used in the discretization, N, is shown in Figure A.1 for the example case of $\rho_{II} = 0.9$ for a 330° notch with $\nu = 0.3$. In this figure, N is varied from 8 to 512 (in multiples of 2), and it is clear that a converged result is obtained for $N = 256$ or greater. Hence, all the results in §3 were computed at this level of discretization. Note that this calculation was performed for a range of combinations of α, ν, and ρ_n, and the rate of convergence was found to be insensitive to ν but to be slightly slower for large values of α and ρ. Thus, the example shown in Figure A.1 shows close to the slowest convergence within the range of α and ρ considered in §3.

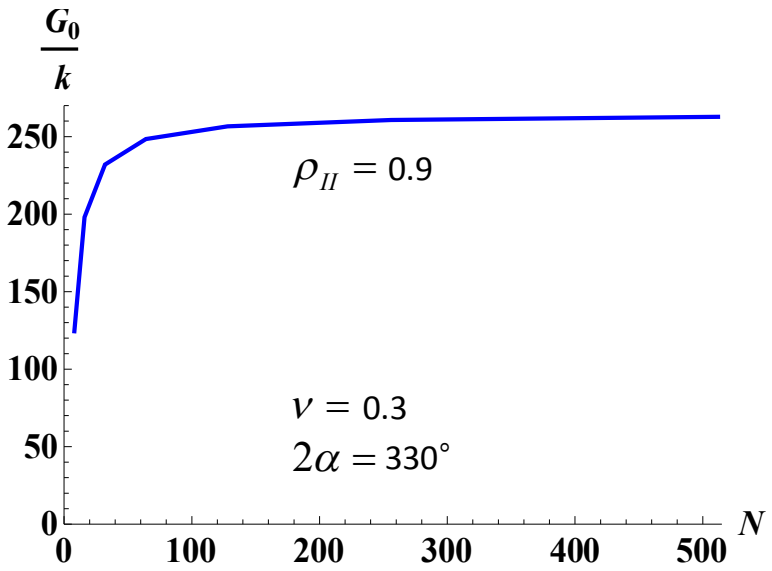


Figure A.1: Typical convergence of ρ with the number of points used in the discretization, N, for the example case of $\rho_{II} = 0.9$, $\nu = 0.3$, and $2\alpha = 330°$.

A.1.2 Small-scale-yielding calculation

To calculate the small-scale-yielding limit, the size of the process zone, $r_p(\theta)$, must be estimated. This can be done by equating (3.1) with (3.2), substituting in stresses from (3.3) and (2.20), using (2.28), and manipulating and simplifying the resulting expression, which gives

$$\frac{3}{(mS_I)^2} = \left(\frac{r_p}{a}\right)^{2(\lambda_I-1)} p_I(\theta) + \left(\frac{S_{II}}{S_I}\right)^2 \left(\frac{r_p}{a}\right)^{2(\lambda_{II}-1)} p_{II}(\theta) + \left(\frac{S_{II}}{S_I}\right) \left(\frac{r_p}{a}\right)^{(\lambda_I+\lambda_{II}-2)} p_M(\theta) \tag{A.3a}$$

$$\frac{3}{(mS_{II})^2} = \left(\frac{S_I}{S_{II}}\right)^2 \left(\frac{r_p}{a}\right)^{2(\lambda_I-1)} p_I(\theta) + \left(\frac{r_p}{a}\right)^{2(\lambda_{II}-1)} p_{II}(\theta) + \left(\frac{S_I}{S_{II}}\right) \left(\frac{r_p}{a}\right)^{(\lambda_I+\lambda_{II}-2)} p_M(\theta). \tag{A.3b}$$

The two expressions in (A.3) are equivalent and can both be used to calculate the size of the process zone in comparison to the nearest geometric feature from the finite problem under consideration, a.

To perform this calculation, a particular small-scale-yielding limit must be selected,

which fixes $r_p^{\max}/a$, and the ratio S_{II}/S_I must also be fixed to some value. A numerical optimization can then be used to search (A.3a) for the value of θ that minimizes mS_I, which occurs when $\theta = \theta^{\max}$, where $\theta^{\max}$ is the angle corresponding to the maximum plastic radius, $r_p^{\max}$, i.e. $r_p^{\max} = r_p(\theta^{\max})$. Thus, the value of mS_I that corresponds to the specified small-scale-yielding limit is obtained. Since S_{II}/S_I is known, the corresponding value of mS_{II} can be calculated easily. This calculation must then be performed for multiple values of S_{II}/S_I. As this ratio tends to infinity as pure mode II loading conditions are approached, this calculation can be performed as described with (A.3a) for $0 \leq S_{II}/S_I \leq 1$ and with (A.3b) for $0 \leq S_I/S_{II} \leq 1$ to account for the full range of mode-mixities.

A.1.3 Processing of experimental data

To process the experimental data in [Ayatollahi et al., 2011, Dunn et al., 1997b, Kim and Cho, 2008, Priel et al., 2007, Seweryn and Łukaszewicz, 2002, Seweryn et al., 1997], the following data are required: the mode I, II generalized stress intensity factors, K_I, K_{II}, at failure; the characteristic length dimension, a, (see Figure 3.8); and the shear yield strength of the material, k. Once K_n, a, k are known, (2.28) can be used to obtain mS_I, mS_{II}, and these values can then be plotted in Figures 3.10-3.13. To calculate G_0/k, the values of mS_I, mS_{II} must simply be substituted into (2.29a), and these can be used to plot Figure 3.14.

In all the papers considered, the tensile strength, σ_y, of the tested PMMA is reported, but the shear yield strength, k, is not. Thus, for the analysis in §3, the shear yield strength is assumed to be given by

$$k = \frac{\sigma_y}{\sqrt{3}} \tag{A.4}$$

in the absence of material data. In Table A.1, the values of a and σ_y used in this analysis are reported, which have been obtained from [Ayatollahi et al., 2011, Dunn et al., 1997b, Kim and Cho, 2008, Priel et al., 2007, Seweryn and Łukaszewicz, 2002].

The values of the generalized stress intensity factors at failure were obtained in dif-

Author	Test group	$a\ (mm)$	$\sigma_y\ (MPa)$
Ayatollahi et al. [2011]	270°	14.142	70.5
	300°	11.547	70.5
	330°	10.353	70.5
Dunn et al. [1997b]	270°	2.54	69.8
	270°	5.08	69.8
	270°	7.62	69.8
	270°	10.16	69.8
Kim and Cho [2008]	270°	20	61.7
	285°	20	61.7
	300°	20	61.7
	330°	20	61.7
Priel et al. [2007]	315° at $198K$	3.266	179.5
	315° at $296K$	3.788	111.8
Seweryn and Łukaszewicz [2002]	280°	25	102.8
	300°	25	102.8
	320°	25	102.8

Table A.1: Values of the characteristic length, a, and tensile strength, σ_y, used in this analysis.

ferent ways depending on the information reported in each paper. For the data reported by Ayatollahi et al. [2011], image extrapolation software was used to infer the generalized stress intensity factors at failure directly from Figure 8 in [Ayatollahi et al., 2011]. Both mode I, II generalized stress intensity factors in this figure are normalized by the *critical mode I generalized stress intensity factor*, which is denoted K_{Ic}^V in [Ayatollahi et al., 2011] (and which is distinct from the *fracture toughness*, which is denoted K_{Ic} in [Ayatollahi et al., 2011]). To obtain the actual values of the generalized stress intensity factors at failure, K_{Ic}^V was calculated from equations (8),(11) in [Ayatollahi et al., 2011] using the values of fracture toughness ($K_{Ic} = 1.96MPa\sqrt{m}$) and tensile strength (see Table A.1) reported in [Ayatollahi et al., 2011]. From equations (4),(A1),(A2) in [Ayatollahi et al., 2011], it can be seen that the authors use the same normalization of the angular functions as in the formulation used here, so no rescaling is required.

For the data reported by Dunn et al. [1997b], image extrapolation software was used to obtain calibrations of the generalized stress intensity factors from Figure 5 in [Dunn et al., 1997b]. Equation (5) in [Dunn et al., 1997b] was then used with the failure loads reported in Table 3 in [Dunn et al., 1997b] to determine the generalized stress intensity factors at failure. Unfortunately, explicit mention of the scaling that the authors applied

Test Group	$a = 2.54$			$a = 5.08$		
mS_I	0.301	0.299	0.295	0.225	0.219	0.203
mS_{II}	0.176	0.174	0.172	0.153	0.149	0.138

(a) Tests on specimens of $a = 2.54$ and $a = 5.08$.

Test Group	$a = 7.62$			$a = 10.16$		
mS_I	0.183	0.189	0.183	0.149	0.157	0.149
mS_{II}	0.144	0.148	0.144	0.132	0.139	0.132

(b) Tests on specimens of $a = 7.62$ and $a = 10.16$.

Table A.2: Dimensionless generalized stress intensity factors at failure for the tests by Dunn et al. [1997a].

Temp $= 296K$		Temp $= 198K$	
mS_I	mS_{II}	mS_I	mS_{II}
0.146	0.013	0.159	0.021
0.153	0.029	0.149	0.034
0.145	0.039	0.141	0.049
0.138	0.056	0.120	0.077
0.129	0.098	0.115	0.095

Table A.3: Dimensionless generalized stress intensity factors at failure for the tests by Priel et al. [2007].

to the angular functions is not made in [Dunn et al., 1997b]. For the analysis presented here, it has been assumed that these authors used the typical normalization of $f^I_{\theta\theta}(0) = 1$ and $f^{II}_{r\theta}(0) = 1$ because this normalization is used in several other papers by the authors around the same period, e.g. [Dunn et al., 1997a, Suwito et al., 1998, 1999]. As this is the same scaling used in the formulation used here, no rescaling is required.

For the tests conducted by Kim and Cho [2008], image extrapolation software was used to infer the mode I, II generalized stress intensity factors at failure (denoted K^n_{Icr}, K^n_{IIcr} in [Kim and Cho, 2008]) directly from Figure 10 in [Kim and Cho, 2008]. However, the formulation used by Kim and Cho [2008] does not normalize the angular functions in the same way as in the formulation used here and also pulls out a factor of $1/\sqrt{2\pi}$ as shown in equation (4) in [Kim and Cho, 2008]. Thus, the generalized stress intensity factors inferred from Figure 10 in [Kim and Cho, 2008] (i.e. K^n_{Icr}, K^n_{IIcr} in [Kim and Cho, 2008]) must be rescaled to match those in the formulation used here. To do this, the obtained values of K^n_{Icr} and K^n_{IIcr} have been multipled by $f^I_{\theta\theta n}(\theta = 0)/\sqrt{2\pi}$ and $f^{II}_{r\theta n}(\theta = 0)/\sqrt{2\pi}$, respectively, where $f^I_{\theta\theta n}(\theta = 0)$ and $f^I_{r\theta n}(\theta = 0)$ are the angular functions used in [Kim

Loading Angle	$2\alpha = 320^\circ$		$2\alpha = 300^\circ$		$2\alpha = 280^\circ$	
	mS_I	mS_{II}	mS_I	mS_{II}	mS_I	mS_{II}
0°	0.049	0.000	0.058	0.000	0.080	0.000
0°	0.055	0.000	0.045	0.000	0.070	0.000
0°	0.056	0.000	0.054	0.000	0.077	0.000
15°	0.050	0.016	0.046	0.018	0.075	0.035
15°	0.054	0.018	0.050	0.020	0.072	0.034
15°	0.052	0.017	0.058	0.023	0.070	0.033
30°	0.057	0.041	0.056	0.048	0.062	0.063
30°	0.047	0.034	0.047	0.040	0.061	0.062
30°	0.044	0.032	0.048	0.041	0.067	0.068
45°	0.045	0.056	0.053	0.078	0.063	0.112
45°	0.049	0.060	0.051	0.076	0.062	0.110
45°	0.051	0.062	0.049	0.072	0.067	0.118
60°	0.034	0.073	0.044	0.114	0.055	0.168
60°	0.037	0.078	0.035	0.091	0.052	0.160
60°	0.036	0.076	0.044	0.111	0.050	0.153
75°	0.023	0.107	0.024	0.130	0.044	0.291
75°	0.020	0.094	0.022	0.119	0.038	0.249
75°	0.023	0.103	0.025	0.140	0.029	0.188
90°	0.000	0.137	0.000	0.217	0.000	0.492
90°	0.000	0.125	0.000	0.188	0.000	0.531
90°	0.000	0.135	0.000	0.201	0.000	0.505

Table A.4: Dimensionless generalized stress intensity factors at failure for the tests by Seweryn et al. [1997] based on the calibrations of the generalized stress intensity factors reported in [Seweryn and Łukaszewicz, 2002].

and Cho, 2008] evaluated along the notch bisector. These angular functions are written out explicitly below equation (2) in [Kim and Cho, 2008].

For the tests reported by Priel et al. [2007], the ratio of the generalized stress intensity factors, K_{II}/K_I, (denoted A_2/A_1 in [Priel et al., 2007]) is given in Table 5 in [Priel et al., 2007]. The value of the mode I generalized stress intensity factor at failure is given in Table 10 in [Priel et al., 2007] (and is denoted A_1 in [Priel et al., 2007]). This information enables the value of the mode II generalized stress intensity factor at failure to be calculated easily. From equation (5) in [Priel et al., 2007], it is clear that Priel et al. [2007] normalize the angular eigenfunctions in the same way as in the formulation used here, so no rescaling is required.

For the data reported by Seweryn et al. [1997], slightly different calibrations of the generalized stress intensity factors are reported in [Seweryn et al., 1997] and [Seweryn and Łukaszewicz, 2002]. The latter set of calibrations were obtained using a more refined model and are more accurate. Hence, this latter set of calibrations have been used, which

Loading Angle	$2\alpha = 330^\circ$		$2\alpha = 300^\circ$	
	mS_I	mS_{II}	mS_I	mS_{II}
0°	0.099	0.000	0.055	0.000
0°	0.106	0.000	0.057	0.000
0°	0.103	0.000	0.059	0.000
30°	0.134	0.053	0.082	0.086
30°	0.162	0.060	0.086	0.090
30°	0.182	0.071	0.086	0.092
45°	0.120	0.080	0.056	0.100
45°	0.120	0.081	0.060	0.107
45°	0.140	0.096	0.066	0.119
60°	0.066	0.080	0.045	0.136
60°	0.076	0.090	0.058	0.178
60°	0.080	0.092	0.066	0.205
75°	0.051	0.131	0.033	0.213
75°	0.051	0.131	0.044	0.291
75°	0.061	0.152	0.057	0.377
90°	0.000	0.095	0.000	0.364
90°	0.000	0.138	0.000	0.416
90°	0.000	0.168	0.000	0.412

Table A.5: Dimensionless generalized stress intensity factors at failure for the tests by Kim and Cho [2008] for $2\alpha = 330^\circ, 300^\circ$.

are given in Table 1 in [Seweryn and Łukaszewicz, 2002], along with the failure loads reported in Table 2 in [Seweryn et al., 1997] to determine the generalized stress intensity factors at failure. From [Seweryn and Łukaszewicz, 2002, Seweryn and Molski, 1996, Seweryn et al., 1997], it is clear that the authors use the same normalization of the angular eigenfunctions as in the formulation used here. However, from equation (1) in [Seweryn and Łukaszewicz, 2002], it is clear that their formulation differs from that used here by a factor of $(2\pi)^{\lambda-1}$. Thus, the obtained values of the generalized stress intensity factors must be multiplied by $(2\pi)^{\lambda-1}$ to rescale them for the formulation used here.

A.1.4 Experimental data

In this section, the calculated values of the dimensionless generalized stress intensity factors at failure are reported in Tables A.2, A.3, A.4, A.5, A.6, and A.7 for the data presented in §3. The standard form of the generalized stress intensity factors can be calculated from these values using the data in Table A.1 along with (A.4) and (2.28).

Loading Angle	$2\alpha = 285°$		$2\alpha = 270°$	
	mS_I	mS_{II}	mS_I	mS_{II}
0°	0.116	0.000	0.133	0.000
0°	0.129	0.000	0.139	0.000
0°	0.129	0.000	0.139	0.000
30°	0.175	0.104	0.188	0.135
30°	0.187	0.113	0.205	0.144
30°	0.195	0.121	0.217	0.149
45°	0.152	0.155	0.222	0.261
45°	0.163	0.172	0.227	0.266
45°	0.167	0.172	0.234	0.287
60°	0.121	0.211	0.172	0.346
60°	0.126	0.220	0.173	0.359
60°	0.146	0.254	0.177	0.363
75°	0.082	0.313	0.165	0.713
75°	0.120	0.451	0.172	0.743
75°	0.136	0.521	0.172	0.743
90°	0.000	0.577	0.000	0.855
90°	0.000	0.603	0.000	0.855
90°	0.000	0.611	0.000	0.855

Table A.6: Dimensionless generalized stress intensity factors at failure for the tests by Kim and Cho [2008] for $2\alpha = 285°, 270°$.

A.2 Additional material for §6

A.2.1 Juoksukangas' experimental data

Juoksukangas et al. [2013] reported that the clamped-cantilever test rig (see Figure 4.2) that they used for their fretting tests had a specimen half-width of $a = 5mm$ and a cantilever length of $L = 200mm$. In addition, they reported that the 'quenched and tempered 34CrNiMo6 steel' used for the tests had a 0.2% yield strength of $\sigma_y = 927MPa$. For the fretting tests they performed, a clamping load was applied that resulted in a mean contact pressure of $p_0 = 100MPa$. Since the shear yield stress, k, was not reported by these authors, it has been calculated using (A.4).

The dimensionless generalized stress intensity factors resulting from $p_0 = 100MPa$ can be calculated using (4.12), which gives $mS_I = -0.045$ and $mS_{II} = 0.077$. The implied bending stress at the extreme points in the load cycle, σ_a, has been inferred from Figure 8 of [Juoksukangas et al., 2013] using image extrapolation software. The corresponding applied tip load, P, can be calculated using (4.14), and (4.12) can then be used

$2\alpha = 330^\circ$		$2\alpha = 300^\circ$		$2\alpha = 270^\circ$	
mS_I	mS_{II}	mS_I	mS_{II}	mS_I	mS_{II}
0.000	0.266	0.000	0.459	0.000	0.969
0.000	0.268	0.000	0.464	0.000	1.009
0.000	0.267	0.000	0.470	0.000	1.081
0.059	0.171	0.065	0.355	0.066	0.740
0.069	0.202	0.070	0.382	0.073	0.812
0.073	0.211	0.070	0.384	0.073	0.826
0.121	0.156	0.124	0.183	0.132	0.429
0.122	0.159	0.139	0.204	0.135	0.434
0.131	0.170	0.155	0.228	0.152	0.484
0.159	0.068	0.185	0.062	0.158	0.193
0.163	0.070	0.187	0.062	0.171	0.211
0.179	0.077	0.224	0.075	0.174	0.211
0.184	0.000	0.174	0.000	0.187	0.000
0.195	0.000	0.204	0.000	0.202	0.000
0.196	0.000	0.213	0.000	0.213	0.000

Table A.7: Dimensionless generalized stress intensity factors at failure for the tests by Ayatollahi et al. [2011].

mS_I	0.110	0.116	0.110	0.064	0.064	0.064
mS_{II}	0.209	0.214	0.209	0.170	0.170	0.170
mS_I	0.054	0.054	0.054	0.038	0.033	0.017
mS_{II}	0.161	0.161	0.161	0.148	0.143	0.130

(a) Clamping load, p_0, and a positive cantilever tip load, P.

mS_I	-0.200	-0.205	-0.200	-0.153	-0.153	-0.153
mS_{II}	-0.055	-0.059	-0.055	-0.015	-0.015	-0.015
mS_I	-0.143	-0.143	-0.143	-0.127	-0.122	-0.107
mS_{II}	-0.006	-0.006	-0.006	0.007	0.011	0.024

(b) Clamping load, p_0, and a negative cantilever tip load, P.

Table A.8: Dimensionless generalized stress intensity factors for the tests by Juoksukangas et al. [2013].

to determine mS_I, mS_{II}. These values are reported in Table A.8, which correspond to the combined contribution of the clamping load, p_0, and the cantilever tip load, P.

A.3 Additional material for §7 and §9

A.3.1 The finite element model

The punch and the half-plane were modelled using ABAQUS/CAE version 6.12-3 as two-dimensional, plane-strain geometries, which incorporated a combination of `CPE4R` and `CPE3` elements. The square punch was modelled with no external boundary conditions, and each of its edges were of length $2a$. The model of the punch was comprised of

5,316 nodes, and it had 128 uniformly spaced nodes along the contact interface. Because a true half-plane cannot be modelled using the finite element method, the half-plane was approximated as a large rectangular block of width $20a$ and height $10a$. This block had two boundary conditions applied to it: (i) the displacement in the y-direction was constrained to be zero along the entirety of its bottom edge, and (ii) the displacement in the x-direction was constrained to be zero at one node at the centre of the bottom edge of the block shown in Figure 7.4. The model of the block was comprised of 19,704 nodes. Like the punch, the block also had 128 uniformly spaced nodes along the contact interface, which was centred about its top edge. Thus, the full model incorporated a total of 50,040 degrees of freedom with 256 degrees of freedom along the contact interface.