ON THE FORMATION OF FINGERPRINTS

by

Michael U. Kuecken

Copyright © Michael U. Kuecken 2004

A Dissertation Submitted to the Faculty of the
GRADUATE INTERDISCIPLINARY PROGRAM
IN APPLIED MATHEMATICS
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2004
THE UNIVERSITY OF ARIZONA &
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Michael U. Kuecken entitled On the formation of fingerprints

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Alan Newell

Date 4/14/04
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Michael Wilson
ACKNOWLEDGEMENTS

It is hard to believe that the dissertation is finally written, and I am grateful to so many people who have helped me to make this work possible. First and foremost I would like to thank my advisor Alan Newell for the tremendous support he gave me. First, by suggesting a topic that was interesting and challenging, then, by numerous and intense discussion sessions, and finally, by still being enthusiastic even when it did not look that great. Many many thanks to Patrick Shipman. Sitting in the same boat as me it was much easier to get through all the pattern and elasticity theory together. Good luck for your work on cacti and plants!

Furthermore my thanks goes to Art Winfree who helped me getting started, and through his course "The art of scientific discovery" taught me that unsolved problems are there to be tackled. Unfortunately, his untimely death prevented him to see how the initial ideas finally developed. Many thanks to my committee members Alain Goriely and Tim Secomb for taking time to read and comment on this work. Also thanks to Darryl Holman, my outside reviewer, for his quick, encouraging and helpful feedback. This work could not have been written without the observations of biologists. Here I am especially indebted to the work of Kristine Bonnevie, Harold Cummins and William Babler whose contributions have been an inspiration.

Another big thanks to Heiko Gerdes. Definitely for the outstanding computer support (providing a computer, software, a terrific setup and computer bug counseling) which really speeded things up. His visualization program TRIVIZ is an awesome and extremely helpful tool for debugging and presentation purposes. Also I am indebted to the many discussions and encouragements when things did not work out.

Thanks also to Chris North, Emily Lane, Joaquin Fox, Taku Tachihara and Alexander Perlis for being so great roommates. I truly enjoyed living with you and will not forget the many flat dinners, parties, concerts and common outdoor activities. You guys rock. A special thanks goes to Emily for being such a picky proofreader. Her suggestions made the work much more readable. Thanks to Luis Garcia-Naranjo for making useful comments on another draft of the thesis. Thanks to Arlo Caine, Robert Lakatos and Steve Hayes for discussions.

Thanks to the Applied Maths Program’s head, Michael Tabor, and the secretaries, Stacey Wiley and Linda Silverman, for helping me through the university bureaucracy and keeping me out of trouble. Thanks to the March of Dimes Foundation and the Wistar Institute for letting me use some of their copyrighted material.

Thanks to Kacy Todirița for suggesting Silly-Putty for studying one’s own finger- and footprints. That was very useful for getting new ideas. Without it, I probably still would not know that there are double loops on both of my big toes.

Finally, and not least at all, thanks to my family for dealing with me being so far away an ocean across. It was worth it.
DEDICATION

Dedicated in memoriam to Harry "Rey" Tester.
# Table of Contents

List of Figures ........................................... 8  
List of Tables ........................................... 15  
Abstract .................................................. 16  

Chapter 1. Introduction ................................... 17  
1.1. Modeling philosophy .................................. 17  
1.2. What is dermatoglyphics? ............................ 20  

Chapter 2. Biological Background ......................... 25  
2.1. Embryology ........................................... 25  
2.1.1. Embryology of skin ............................... 25  
2.1.2. The volar pads .................................... 28  
2.1.3. The development of the epidermal ridge system .. 30  
2.2. The connection between volar pad geometry and ridge patterns .. 38  
2.2.1. Dermatoglyphics in mammals and marsupials .... 39  
2.2.2. Empirical studies .................................. 45  
2.2.3. Dermatoglyphics in malformed hands .......... 49  

Chapter 3. Theories of Ridge Development ................. 53  
3.1. The folding hypothesis and other mechanical hypotheses .... 53  
3.1.1. Kollmann's work .................................. 53  
3.1.2. Bonnevie's work .................................. 55  
3.1.3. Other theories based on mechanics and criticism of the folding hypothesis ......................... 61  
3.2. The nerve hypothesis .................................. 66  
3.3. The topological approach and the fibroblast hypothesis .... 70  
3.4. Miscellaneous theories ................................ 77  
3.4.1. Evolution as the creator of fingerprints ........ 77  
3.4.2. The Penrose hypothesis ......................... 79  
3.4.3. Reaction–diffusion models ....................... 82  

Chapter 4. A Buckling Model for Fingerprint Formation .... 84  
4.1. Biological arguments .................................. 86  
4.2. An overview of shell theory and derivation of the shallow shell equations 91  
4.3. Solutions and perturbed equations ................... 102  
4.4. Linear stability analysis ............................. 104
### Table of Contents—Continued

4.5. Nonlinear analysis .......................... 109
   4.5.1. Analysis of the displacement equations .... 109
   4.5.2. Analysis of the von Karman equations: weakly anisotropic case 113
   4.5.3. Analysis of the von Karman equations: strongly anisotropic case 128
4.6. Determination of parameters .......................... 137
4.7. Summary ................................ 141

CHAPTER 5. The Formation of Stress in the Epidermis ........ 145
5.1. Biological considerations .......................... 145
5.2. The role of hand geometry – a heuristic explanation 153

CHAPTER 6. Computer Experiments .......................... 160
6.1. Finding the stress field .......................... 160
6.2. Numerical Observations at the von Karman equations .... 165
6.3. Simulating the ridge pattern .......................... 171

CHAPTER 7. Conclusions ................................ 180
7.1. Summary ................................ 180
7.2. Achievements and critique .......................... 181
7.3. Open questions and suggestions for further work .......................... 185

APPENDIX A. Geometric Theory of Shallow Shells ........ 189
   A.1. Classical differential geometry of surfaces 189
   A.2. Deformation of a shallow shell 191

APPENDIX B. Numerical Algorithms .......................... 195
   B.1. Finite elements .......................... 195
   B.2. The spectral scheme .......................... 199
   B.3. shells.cpp .......................... 200
   B.4. Tests .......................... 204
   B.5. Visualization .......................... 207

REFERENCES ................................ 209
LIST OF FIGURES

FIGURE 1.1. The most frequently occurring fingertip patterns: whorl, loop and arch (from [28]). ................................. 21

FIGURE 1.2. A triradius (from [28]). ................................. 21

FIGURE 1.3. A menagerie of accidental patterns (from [92]), courtesy of March of Dimes. ........................................ 22

FIGURE 1.4. Examples of minutiae. The white dots represent sweat pores. ........................................ 23

FIGURE 1.5. Minutiae in a real fingerprint. Numerous ends and forks can be seen. There is a short ridge just below the triradius. Incipient ridges are present slightly above the loop core. The white line from the triradius to the loop core is used to find the configuration ridge count. This is the number of ridges that crosses the white line (from [28]). ................................. 23

FIGURE 2.1. The figure denotes the four important directions (ulnar, radial, distal, proximal), it establishes the enumeration system of the digits and shows the most important flexion creases. For the creases the nomenclature in [55] was used: WC - wrist crease, DIC – distal interphalangeal crease, PIC – proximal interphalangeal crease, MCPC – metacarpophalangeal crease, DTC – distal transverse crease, PTC – proximal transverse crease, TC – thenar crease. ................................. 26

FIGURE 2.2. The volar skin at the 10th week. The epidermis is located on top of the dermis and consists of three layers. The outer layer is the periderm. It is composed of elongated cells. Underneath the periderm is the intermediate layer. The innermost layer is the basal layer. The cells of this layer are darkly stained and have a columnar appearance. In comparison to the epidermis, the dermis appears amorphous. We can recognize a mesh of mesenchymal cells with no apparent structure (from [11], courtesy of March of Dimes). ........................................ 28

FIGURE 2.3. The locations of the volar pads. A1 to A5 denote the five apical pads on the fingertips. I1 to I4 denote the interdigital pads. II is usually not well developed. Th denotes the thenar pad and HTh the hypothenar pad. ................................. 29

FIGURE 2.4. The epidermis at about the 11th week. The basal layer is slightly undulated (from [46], courtesy of the Wistar Institute). ................................. 31

FIGURE 2.5. The undulations in the basal layer become more distinct and form the primary ridges (from [11], courtesy of March of Dimes). ................................. 31

FIGURE 2.6. A tangential section through a finger at about the 11th week. A whorl has formed (from [11], courtesy of March of Dimes). ................................. 34
LIST OF FIGURES—Continued

FIGURE 2.7. (a) A schematic view on the locations of the first primary ridges. They start to form at the nail furrow and in a localized area in the middle of the pad. (b) The center of a whorl forms. (c) A loop forms (from [19]).

FIGURE 2.8. (a) A whorl has formed, the three systems emanating from the ridge anlage, the nail furrow and the phalangeal crease start to merge and only the triradii are left unridged. (b) A tented arch forms, ridges form almost instantaneously on the volar pad (from [19]).

FIGURE 2.9. Sometimes the primary ridges spread from two ridge anlagen. BONNEvie believed that accidental patterns arise in this case. However, it seems possible that in the above fingers regular loop and whorl configurations will form. (from [19])

FIGURE 2.10. The percentage increases of primary ridge breadth, hand length and number of minutiae are given as a function of the CRL. It seems that ridge breadth and hand length growth occur at almost the same rate. Further, there is an increase of minutiae until a CRL of 170 mm (from [45], courtesy of the Wistar Institute).

FIGURE 2.11. The percentage increases of primary ridge breadth, hand length and number of minutiae are given as a function of the CRL. The ridge breadth growth slightly lags behind the hand length growth at the beginning of ridge formation. Later on there is hardly any difference between these rates. Many minutiae are formed in the time between 75 mm and 150 mm CRL (from [45], courtesy of the Wistar Institute).

FIGURE 2.12. A basic scheme of the fully established ridge system.

FIGURE 2.13. Palm and sole of the night monkey, Aotus zonalis. The palmar pads are very pronounced (from [28]).

FIGURE 2.14. The dermatoglyphic patterns of the monkey in Figure 2.13. It is easy to see the correlation between palm topography and ridge alignment (from [28]).

FIGURE 2.15. The sole of Papio hamadryas. Summits are surrounded by concentric circles (whorls) and loop structures are found on the elongated elevations (from [97]).

FIGURE 2.16. The dermal surface of a koala. Spots are observed on the palm and ridges form on the finger apex along the nail furrow (from [79], courtesy of March of Dimes).

FIGURE 2.17. The dermal surface of a vulpine phalanger. Both spots and stripes are observed. In 20b the transition from stripes to dots can be observed (from [79], courtesy of March of Dimes).
LIST OF FIGURES—Continued

FIGURE 2.18. Cross-sections through the fingertips of the right hand of Bonnevie’s embryo No. 83. The thumb is on the right side. The separation of fingers has just occurred and the pads have started to appear. Note that digit I appears fairly symmetric, digit II is slanted slightly to the ulnar direction, digits III and IV are slightly slanted radially and digit V is strongly slanted to the radial side (from [19]). ........................................ 48

FIGURE 2.19. A scheme for the separation of the fingers. At first the thumb (digit I) separates. The remaining fingers form a mass with rounded ends. When these fingers separate fingers II and V retain the rounded outline on one side and become slanted toward the center of the hand mass. After separation has finished fingers II and V will exhibit stronger asymmetries than the other fingers. .................................................. 49

FIGURE 2.20. An example for hyperdactyly in two feet (the large toe and the small toe are doubled). The whorl patterns on the doubled hallux of the left foot are usual patterns on that digit, however the tibial loop on the hallux of the right foot is most unusual and could be explained by Bonnevie’s symmetry arguments. Also the arches on the doubled small toes on both feet are unusual pattern types. On the left foot a triradius at the fibular side is seen. This indicates that yet another small toe was originally formed at that place (from [26], courtesy of the Wistar Institute). .................................................. 50

FIGURE 3.1. A cross-section through an embryonic fingertip. The two papillary nerves (N.p.) converge towards the ridge anlage (Pap.) (from [19]). 57

FIGURE 3.2. Two cross-sections through an embryonic fingertip. The left figure reveals that two ridge anlagen are present. A more proximal cut shows incomplete convergence of the papillary nerves (N.p. I and N.p. II) that point to the centers of the ridge anlagen (Pap. I and Pap. II) (from [19]). .................................................. 59

FIGURE 3.3. Dissociation of ridges into spotty irregular patterns can be observed in this print (from [28]). .................................................. 60

FIGURE 3.4. On this section the epidermis is shown under the dermis. The basal layer is clearly seen darkly stained. Formation of primary ridges has just been initiated on the outside parts of the section and is more advanced toward the center. The very large number of basal layer cells in the center regions indicates that cell proliferations not only lead to compressive stress and folding, but also deepen the primary ridge once folding has occurred (from [19]). .................................................. 64

FIGURE 3.5. The dermis and epidermis of the opossum at the time of ridge initiation (from [68]). .................................................. 69
LIST OF FIGURES—Continued

FIGURE 3.6. (a) Surrounding a loop core in counterclockwise direction results in a twist of the wave director of 180°. (b) Similarly, surrounding the center of a triradius gives a -180° twist. 71

FIGURE 3.7. (a) A whorl has a twist of 360°. This is not surprising considering it is the composition of two loops (b). 71

FIGURE 3.8. Consider a ridge field on a disk that is always perpendicular to the boundary of the disk. Then the twist of the corresponding director field along the disk boundary is 360°. 72

FIGURE 3.9. At the fingertip the ridges run parallel to the finger margin. If we "straighten out" the finger margin it becomes evident that the wave director rotates by -180° along the margin. 73

FIGURE 3.10. This is figure 2.22 from [62]. According to Loesch pressure is built up in the lower layers of the epidermis and tension is formed in the upper layers. Note that the ridges follow the lines of greatest curvature. 81

FIGURE 4.1. The cytoskeleton of the basal layer cells are attached to each other by desmosomes and to the basal lamina by hemidesmosomes. 86

FIGURE 4.2. A schematic view of our model. For simplicity, an uncurved piece of fetal skin is shown. The basal layer is trapped between the upper epidermis layers and the dermis. We assume that the growth rate of the basal layer exceeds that of the upper layers. 87

FIGURE 4.3. The differential growth produces compressive stress in the basal layer. For the subsequent buckling process, the dermis and the intermediate layer act as beds of resisting springs. 89

FIGURE 4.4. After reaching a critical stress buckling of the basal layer takes place. 90

FIGURE 4.5. The strains \( \varepsilon_x \) and \( \varepsilon_y \) denote the (small) relative increase of a small piece of the shell in the \( x \)-direction and \( y \)-direction, respectively. The strain \( \varepsilon_{xy} \) is a measure for the geometric distortion. 93

FIGURE 4.6. In-plane stresses acting on a small area of the shell. As usual we have \( \sigma_{xy} = \sigma_{yx} \). 93

FIGURE 4.7. The three mechanisms responsible for the formation of strain, illustrated at \( \varepsilon_z \). The solid curves represent the undeformed state and the dashed lines the deformed state. See text for further explanation. 97

FIGURE 4.8. The \( y \)-axis is the direction of greatest curvature, the \( x \)-axis is the direction of smallest curvature. Further, the \( \hat{N}_z \)-axis is the one of largest stress and the \( \hat{N}_y \)-axis the one of smallest stress. 106
LIST OF FIGURES—Continued

FIGURE 4.9. Two different buckling scenarios. In the first case the stress on the system is maintained, the shell is squashed together. If the shell cannot move at the ends, the stress on the system decreases and the amplitude is reduced compared to the first case. ............................................... 111

FIGURE 4.10. A triad of wave vectors. The vectors all have the magnitude of the critical wave number and are spaced 120° away from each other. The sum of three cosine waves with these wave vectors results in hexagons. Note that we have $k_1 + k_2 + k_3 = 0$. This property is crucial for quadratic interactions that stabilize hexagon patterns. ............................................... 116

FIGURE 4.11. The stability domain in the $\sigma\eta$–plane for hexagons for $\phi = 2$. 125

FIGURE 4.12. The stability domain in the $\sigma\eta$–plane for rolls type I for $\phi = 2$. 126

FIGURE 4.13. The stability domain in the $\sigma\eta$–plane for rolls type II for $\phi = 2$. 127

FIGURE 4.14. The domain of the modes $k$ for which $A e^{ik\eta}$ with $A = e^{ik\eta}$ is stable. The stability domain (black) lies inside the domain of amplified wave vectors (grey). Parameters $\chi = 0.2$ and $\Gamma = 0.9$ were used for all examples. Further we have (a) $\Lambda = 0$, (b) $\Lambda = 1$, (c) $\Lambda = 5$, (d) $\Lambda = 30$. 135

FIGURE 4.15. The domain of the modes $k$ for which $y e^{ik\eta}$ with $A = e^{ik\eta}$ is stable. The stability domain (black) lies inside the domain of amplified wave vectors (grey). Parameters $\chi = 0.2$ and $\Gamma = 0.2$ were used for all examples. Further we have (a) $\Lambda = 0$, (b) $\Lambda = 1$, (c) $\Lambda = 5$, (d) $\Lambda = 30$. 136

FIGURE 5.1. The relation between folds and furrows (thick lines), the periphery of the palm and the epidermal ridges are shown. The ridges tend to run parallel to the thick lines and perpendicular to the margin of the palm. 147

FIGURE 5.2. (a) A cross-section of a surface with curvature $1/R$ (solid curve) is displaced by $w$ toward the center of the curvature circle (dashed curve). The cross-section becomes strained. Stress proportional to $w/R$ is induced. (b) No stress will be induced if the cross-section is free to expand in tangential direction. ................................. 148

FIGURE 5.3. (a) A cross-section through the apical volar pad at the beginning of volar pad digression. (b) If the fingertip grew exactly uniformly the fingertip would preserve its shape and only increase in size. (c) The fingertip changes its shape due to volar pad regression. The growth forces induced by the regression are obtained by morphing figure (b) into figure (c). To accomplish this task normal displacements toward the center of the pad are necessary. ................................. 150

FIGURE 5.4. (a) The fingertip, the location of the ridge anlage and certain directions radiating from it. (b) The interdigital region with the center of the interdigital pad and two directions radiating from it. .................. 155
LIST OF FIGURES—Continued

FIGURE 5.5. The forces shaping a whorl. (a) The volar pad is high, rounded and normal pressure in the area produces radial stress away from the ridge anlage. (b) The volar pad is less pronounced and stress along A–C dominates in the center. 156

FIGURE 5.6. The forces shaping tented arches and arches. (a) For a flatter pad little stress is generated in direction D–F. Boundary effects dominate and a tented arch arises. (b) If the pad is even flatter and little normal pressure is applied the boundary forces dominate the stress pattern and arches form. 157

FIGURE 5.7. The forces shaping loops. (a) If we have an asymmetric pad (the line of the location of greatest radius is indicated by the dashed line) the pattern becomes asymmetric as well. For moderate pad elevation stresses in direction A–C dominate. The direction of the ridges in the center of the pad is influenced by the boundary forces. (b) If the ridge anlage is asymmetric the pattern becomes asymmetric as well due to interaction of the stress arising at the ridge anlage and the stress arising at the boundary. 158

FIGURE 5.8. A configuration not found on fingerprints. See text for explanation. 159

FIGURE 6.1. The direction of the forces acting at the boundary. 161

FIGURE 6.2. (a) Normal load is applied on the summit of a pronounced pad. (b) A whorl arises, largest stress is found at the center of the whorl and at the periphery close to the nail and the DIC. 162

FIGURE 6.3. (a) A pad slanted to the left side is subjected to a normal pressure centered at the ridge anlage which is shifted to the right side. (b) A loop forms. 163

FIGURE 6.4. (a) The pad is flat and little normal load is applied. (b) An arch arises. 163

FIGURE 6.5. Buckling patterns for the stress field SI using $\chi = 1.05$ and $\tau = 0.95$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$ 166

FIGURE 6.6. Buckling patterns for the stress field SI using $\chi = 1.2$ and $\tau = 1.1$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$ 167

FIGURE 6.7. Buckling patterns for the stress field SI using $\chi = 1.6$ and $\tau = 1.4$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$ 168

FIGURE 6.8. A target according to stress field SIII. Here we used $\chi = 0.2$ and $\Lambda = 0.57$. 169

FIGURE 6.9. A loop consisting of a convex disclination and many dislocations according to stress field SIII. Here we used $\chi = 0.2$ and $\Lambda = 0.57$. 169
LIST OF FIGURES—Continued

FIGURE 6.10. A loop consisting of a convex disclination and many dislocations according to stress field SIV. Here we used \( \chi = 0.2 \) and \( \Lambda = 0.57 \). 170

FIGURE 6.11. A whorl pattern obtained using approach 1. Ridge formation takes place simultaneously over the surface. The characteristic features of the whorl including the dislocations are well-captured. However, note the sharp change in the wave vector above the triradii. 173

FIGURE 6.12. A loop pattern using approach 1. The pattern exhibits the characteristic flow of loops. The triradii and the loop core are well-represented. 173

FIGURE 6.13. An arch pattern using approach 1 exhibiting realistic flow of the pattern. 174

FIGURE 6.14. A whorl pattern forms. Ridge formation takes place at the periphery and the whorl core and spreads over the remainder of the skin surfaces. Finally the triradii are filled in. 176

FIGURE 6.15. A loop forms. The ridges start to form at the periphery and the loop core, finally the triradius is filled in. 177

FIGURE 7.1. Real and simulated fingerprint patterns: (a) whorl, (b) loop and (c) arch. 183

FIGURE B.1. An example face indicating the local and global coordinates. 196

FIGURE B.2. A cantilever of length \( L \) with rectangular cross-section with dimensions \( b \) and \( c \) is bent by a force \( P \) at its end. 205
LIST OF TABLES

TABLE 2.1. Configuration percentages of 24,518 Norwegians. Values are given for each individual finger and for all fingers together (from [18]). .... 46

TABLE B.1. Computed and exact thin plate solution for the center deflection of a uniformly loaded plate. ......................... 205

TABLE B.2. Computed and exact solutions for deflections and stresses of a loaded cantilever. ................................. 206

TABLE B.3. Computed and exact solutions for deflections and stresses of a loaded cylinder. ................................. 206
ABSTRACT

The fingerprint pattern (epidermal ridge pattern) becomes established at about the 10th to 16th week of pregnancy, when the lowest layer of the epidermis, the basal layer, becomes undulated. The pattern established by these undulations becomes visible on the skin surface in subsequent weeks of pregnancy. We argue that the undulation process is initiated by buckling of the basal layer. The necessary compressive stress is generated by differential growth.

The instability is investigated using the classic von Karman equations for curved surfaces. The analysis reveals that ridges (rolls) are the most common pattern type and that the local ridge direction of the pattern is perpendicular to the direction of largest stress. For certain parameter regimes dot patterns (hexagons) are a stable solution of the equations. Such dot patterns are, in fact, observed on the palms of certain marsupials.

The stress in the basal layer is induced by two mechanisms. First, the basal layer expands faster than the other skin structures. Such expansion is resisted at the major flexion creases and the nail furrow. Second, there is a change in fingertip geometry at the time of pattern generation that provides a different source of growth stress.

The combination of the two processes predicts the correct sequence of pattern spread over the fingertip. It also explains the observation that fingerprint configurations are related to the fingertip geometry at the time of ridge formation. Computer simulations for the most important configurations exhibit many features of actual fingerprints and suggest directions for future work.
1.1 Modeling philosophy

Modeling embryological processes is notoriously difficult due to the immense complexities of morphological change. Cells migrate, proliferate, align with each other, exert forces on each other, stick to each other, excrete hormones; genes switch on or off depending on various environmental factors; hormones, morphogens and nutrients diffuse through the tissue and modulate all kinds of cell responses. Looking at this bewildering process it seems at first sight that everything potentially acts on everything in an embryo.

It seems hopeless to understand this full process in its entirety and one has to separate specific phenomena from this entangled web of dependencies that are easy enough to study and still tell us something about how the full system works. Another challenge is a purely practical one. Experiments on embryos are difficult to conduct in vivo and not always conclusive in vitro.

Mathematical modeling can be of enormous help in this situation. Nature is full of patterns that have a mathematical connection. Well-known examples are the logarithmic spiral of certain sea shells, the Fibonacci series in the spiral number of sunflower seeds or the stripes and dots on tropical fish. Once basic mechanisms are known, models can study the interactions between certain biological phenomena and, to a certain degree, replace experiments that would be difficult or impossible to conduct. Mathematics can relate empirical observations to each other and suggest theories that would not have been obvious purely from biological observations.

However, modeling morphogenetical processes presents enormous challenges. Unlike many theories in the physical sciences (such as elasticity theory or fluid dynamics)
the microscopic equations for embryological processes are not well-established. Even if equations can be established there are conflicting requirements. Taking into account too much biological complexity will lead to 'messy' models that are intractable analytically and often even numerically. However, simplifying or ignoring too much of the biology threatens the credibility of the model and renders it useless for understanding morphological change.

The difficulty in conducting experiments leads to a paucity of information needed to develop a successful model. For example, the parameters modelers are interested in to validate their models are often unavailable.

The question arises as to what makes a good mathematical model of morphological processes. This is a difficult question and it is actually somewhat easier to argue what does not make a good model. Apparently, a successful model should incorporate some biological knowledge into a mathematical framework and — after some analysis — use the results to address some of the biological questions and ask new ones.

Often it is argued that a model is reasonable because the patterns produced by it 'resemble' the ones seen in nature. However, this kind of reasoning is dangerous. Many differential equations with certain symmetries exhibit the same kind of solution behavior. Therefore, a completely different mechanism could still produce the same pattern type.

Another danger is the focus on mathematical techniques that considers the chosen equations as the primary study object. While this approach may produce valuable facts about the equations it does not necessarily tell us anything interesting about the biological processes. The equation one uses are not sacred and may have to be modified in the course of the investigations. It is important to relate the analysis results back to the biological observations and check whether these results are consistent with what we observe or are mere mathematical artefacts.

The more credible models succeed in explaining certain nongeneric peculiarities of the pattern or reducing the possible explanations of the observations to an abstract
principle. A successful model should obtain more from the analysis than was invested in it in the beginning. Therefore it is not so important whether or not the model is "right", but how much we can learn about the process that we are interested in in the first place. Obviously, consistencies between model and reality are desirable but one can and should also gain insight by focusing on model shortcomings.

It is important to formulate the analysis results in a language that biologists can understand. Hiding behind intimidating equations may make it difficult for biologists to criticize a certain model, but they will not be able to use it either. We have to realize that mathematical models are not developed for applied mathematicians but for the biologists to use as a tool for understanding nature.

Ultimately, models are judged by the facts that are collected by biologists. Perfectly reasonable models have been proven wrong by experimental observations. Even if a mathematical model cannot ultimately "prove" that it adequately describes a certain process, it can change the way people think about biology. It can help us to identify new directions of inquiry to deepen our understanding how a full-grown animal emerges from a single cell.

In this sense, the mathematician interested in biology is in the same situation as a jigsaw puzzle player. His or her training in formal and abstract thinking and the command over mathematical techniques can help to connect the pieces together and to suggest ways how new pieces can be found. However, mathematics cannot "create" biology and we have to stick with the pieces of information which are provided to us by observation and experiment.

In this work we will ask the question of how fingerprint patterns are formed in the embryo. There are several reasons why fingerprints are a suitable object for mathematical attention. Many researchers have described the embryology of fingerprints and have accumulated a lot of empirical knowledge about fingerprint development. Further, there is an extensive literature about fingerprint changes in individuals with certain hand or finger malformations or with genetic diseases. A lot of information is
available describing the ridge pattern in other mammals.

Many theories about fingerprint development have been published, mostly by biologists. The few mathematical attempts to describe fingerprints all suffer from severe shortcomings and none of them is a useful base for further research.

It is the purpose of this work to provide such a base. We will argue for a certain mode of fingerprint development that aims to integrate as many biological ideas as possible. We will see how mathematics helps us to understand this picture and how computer simulations suggest a way the larger scale patterns form.

This work should not be seen as an attempt to explain everything. We will be able to connect some jigsaw pieces but many holes remain in the picture. We hope that the pieces in place make it easier to attach new ones and make the underlying picture visible.

1.2 What is dermatoglyphics?

The term dermatoglyphics (Greek: δηρυγμα – Skin, γαλυφω – to carve) was proposed in 1926 by HAROLD CUMMINS to denote the study of epidermal ridges on fingers, palms and soles. The pattern of these ridges is what we colloquially call a fingerprint.

In the scientific literature, fingerprints are referred to as epidermal or dermal ridges because the ridges also occur on palms and soles as well. In this work we will use the term epidermal ridges or just call them ridges.

Traditionally, the ridge configurations on the fingertips (apical patterns) have met with special interest. Three main configurations are discriminated: whorls, loops and arches (see Figure 1.1). For this classification a certain structure is used that is characterized by three ridge bundles meeting at an angle of approximately 120°. This structure is called a triradius (see Figure 1.2). A triradius is sometimes referred to as a delta. An apical pattern having two triradii is classified as a whorl, a pattern with one triradius is called a loop and one with no triradius is called an arch. A whorl
Figure 1.1. The most frequently occurring fingertip patterns: whorl, loop and arch (from [28]).

Figure 1.2. A triradius (from [28]).

is characterized by a target or spiral structure on the fingertip center, a loop by the Roman arch structure. If such structures are present triradii form as a topological necessity (as we will argue later in more detail).

A loop on the fingertips that opens in the direction of the thumb is called a radial loop, otherwise it is called an ulnar loop. Similarly, a loop on toes that opens in the direction of the hallux is called a tibial loop whereas a loop in the opposite direction is referred to as a fibular loop. There are various apical configurations that represent intermediate forms between these three main types. Further there are a class of configurations, so-called accidentals, that cannot be classified this way. Examples of accidentals are double loops or more complicated whorls (see Figure 1.3).

Another important feature of fingerprints are the numerous small defects in the ridge pattern such as forks, ends, incipient ridges, enclosures, short and island ridges (see Figure 1.4 and Figure 1.5). These defects are called minutiae in fingerprint literature.

Epidermal ridge patterns have been of interest for humans for a long time. They can be seen in prehistoric petroglyphs and were probably already used for identification by the Chinese 2000 years ago, long before Europeans introduced them for the same purpose at the end of the 19th century. Scientific work on the morphology
of fingerprints started in the late 17th century and was summarized in a paper by PURKINJE in 1823 [89]. He developed a classification scheme which already included the three basic fingerprint configurations: whorls, ulnar and radial loops and arches. The classic book by GALTON in 1892 [36] was a further landmark in the evolution of the field. It broadened the scope of fingerprint research to include problems such as inheritance and ethnic variation. The most influential researcher in the field of the 20th century was HAROLD CUMMINS. His wonderful book “Fingerprints, Palms and Soles” [28] is still the most comprehensive source of information on the subject.

Today a huge literature is available that classifies fingerprint patterns, lists the distribution of the most common configuration types among certain populations and attempts to find inheritable features of fingerprints [51]. According to these studies, fingerprints are statistically linked to ethnicity, gender, genetic defects and even more obscure features like profession, high blood pressure or sexual orientation [47]. After reviewing this literature one may wonder if there is in fact some truth in “palm reading”.

A well-known and important application of fingerprints is identification. This is possible since all humans have a unique ridge pattern that does not change in life. This uniqueness is due to the individual position of the minutiae that allows identification even if only a small patch of skin is considered. Today fingerprints are routinely used by police departments, and data banks contain millions of samples. In spite of some recent discussion concerning the scientific foundation of fingerprint identification [1] it seems that fingerprint evidence will continue to be used extensively in court and
FIGURE 1.4. Examples of minutiae. The white dots represent sweat pores.

FIGURE 1.5. Minutiae in a real fingerprint. Numerous ends and forks can be seen. There is a short ridge just below the triradius. Incipient ridges are present slightly above the loop core. The white line from the triradius to the loop core is used to find the configuration ridge count. This is the number of ridges that crosses the white line (from [28]).
will not be completely replaced by DNA identification methods. Fingerprints are also increasingly used as an identification tool in biometric devices, for instance in car keys [53].

Fingerprints are also useful in other areas, such as genetic, medical or ethnic studies. It is well-established that genetic defects frequently lead to typical changes in fingerprint patterns. In the cases of Down syndrome and Turner syndrome these changes are well-documented [95]. Thus, fingerprints can be used as an easy tool for the diagnosis of genetic defects. However, more reliable DNA methods have recently diminished this application of fingerprints.

The importance of fingerprints for ethnic studies stems from the fact that different groups of people have statistical differences in their fingerprint features. An extensive summary of such data can be found in the appendix of BABLER’s PhD thesis [8]. This data can be used to help establish the degree of relatedness between different peoples. Again, the importance of this application has faded after the spread of DNA analysis.
CHAPTER 2

BIOLOGICAL BACKGROUND

2.1 Embryology

2.1.1 Embryology of skin

Before we describe the process of fingerprint initiation we will cover some facts about embryonic skin that will be useful for the understanding of this work. In this subsection, we will talk about the embryonic development of skin in general. In the following we will look specifically at the volar pads.

In this work we will use some common anatomical concepts (ulnar, radial, distal, proximal) that are explained in Figure 2.1. This figure also contains an overview of important anatomical landmarks on the palm. Note that in medical literature the skin on the palm is referred to as volar, whereas the skin on the back side of the hand is called dorsal.

Adult skin consists of two layers with very different properties and different embryological origin. The upper layer is called the epidermis and represents a typical epithelial tissue. It originates from the embryonic ectoderm. The epidermal cells are organized in ordered layers. New cells arise from cell divisions in the innermost layer and, in the course of a few weeks, migrate to the skin surface. During this migration they produce a protein, keratin, that is largely responsible for the mechanical properties of skin. If cells have achieved a certain degree of keratinization they die off and form the outer corneous layer of the epidermis.

In about the third week of fetal life the epidermis consists of a single layer of undifferentiated cells. After six weeks of pregnancy two layers can be recognized, an outer and an inner layer. The outer layer is called periderm and is a purely embryological
FIGURE 2.1. The figure denotes the four important directions (ulnar, radial, distal, proximal), it establishes the enumeration system of the digits and shows the most important flexion creases. For the creases the nomenclature in [55] was used: WC – wrist crease, DIC – distal interphalangeal crease, PIC – proximal interphalangeal crease, MCPC – metacarpophalangeal crease, DTC – distal transverse crease, PTC – proximal transverse crease, TC – thenar crease.
structure that is lost once the epidermal cells produce keratin. The inner layer is called the stratum basalis (or basal layer) or stratum germinativum. Between 8 and 10 weeks another layer is formed between periderm and stratum basalis, the stratum intermedium (or intermediate layer). Desmosomes between the cells of the basal layer and hemidesmosomes, that attach the basal layer cells to the basal lamina, are first seen at about the 8th week [49, 86]. The basal lamina is a protein sheet that seals the epidermis off from the dermis. In the following weeks more layers in the epidermis can be distinguished and keratinization starts in about the 14th week.

The epidermis sits on top of the lower skin layer which is called the dermis. The dermis forms from the embryonic mesenchyme and is innervated at about 6 to 8 weeks of pregnancy. Axonal growth cones project to the epidermis and contact Merkel cells in the epidermis at about 8 to 10 weeks of pregnancy [66]. Often the axons are located close to blood vessels forming what is called in the literature a nerve-vessel pair [49]. The embryonic dermis of the 10th week consists mainly of fibroblasts, the precursor of cells as diverse as fat cells, muscle cells, cartilage cells and connective tissue. There is still a lot of empty intercellular space between the fibroblasts. In the subsequent weeks, as the fibroblasts proliferate and produce fibers such as collagen and glycosaminoglycans the dermis becomes much denser.

Changes in the volar skin, critical for fingerprint development, start at the 10th week. Interestingly, and maybe not accidentally, this is also the time of hair follicle initiation.

The processes during skin morphogenesis are still a topic of current research. A good summary with much more details than the synopsis presented here can be found in [35, 50].
FIGURE 2.2. The volar skin at the 10th week. The epidermis is located on top of the dermis and consists of three layers. The outer layer is the periderm. It is composed of elongated cells. Underneath the periderm is the intermediate layer. The innermost layer is the basal layer. The cells of this layer are darkly stained and have a columnar appearance. In comparison to the epidermis, the dermis appears amorphous. We can recognize a mesh of mesenchymal cells with no apparent structure (from [11], courtesy of March of Dimes).

2.1.2 The volar pads

The volar pads are temporary eminences of the embryonic volar skin. They should not be confused with muscular eminences but rather consist of subcutaneous tissue and fat. The volar pads are believed to play a crucial role in ridge pattern determination.

The pads form at distinct places on the palmar and plantar surface of many mammals and marsupials. Three series of pads have been established. The first series consists of pads at the volar side of the fingertips (apical pads). The second one consists of the pads on the distal part of the palm on the space between the bases of the fingers (interdigital pads). In some species these pads are somewhat relocated and are rather found below the base of the fingers. The third series includes the pads in the thenar and hypothenar areas (thenar and hypothenar pad) (see Figure 2.3).

The pads of these series are referred to as primary pads because they can also be seen on many other mammal and marsupial species. There are also less distinct, so-called secondary pads, that are not generally observed in other species. For example, in humans there are secondary pads at the base of the digits. They are located in
FIGURE 2.3. The locations of the volar pads. A1 to A5 denote the five apical pads on the fingertips. I1 to I4 denote the interdigital pads. I1 is usually not well developed. Th denotes the thenar pad and HTh the hypothenar pad.
pairs in the area where the finger meets the palmar surface. Some monkeys have secondary pads at the lower phalanx of the thumb or other digits. Secondary pads are important because they help us to understand dermatoglyphic changes in certain species.

In the human embryo, volar pads start to form at the 7th week of pregnancy. They continue to grow until about the 9th week and finally appear as high, rounded hillocks with a clearly defined base. Later on they digress, appear less pronounced and their base merges with the surrounding tissue. Although apical and interdigital pads are still present in higher term embryos, and sometimes even at birth, the hand geometry approaches that of an adult. In human hands the hypothenar, the thenar and first interdigital pad are not well-developed and digress early. On the foot, however, the thenar pad merges with the first interdigital pad and forms what is called the hallucal pad, a large eminence below the base of the big toe.

In most primate species, volar pads persist until adulthood and are believed to serve as cushions for walking. Because of the evolutionary changes in humans that freed hands for locomotion purposes and relocated the mass center backwards, the pads are no longer needed for their original purpose and digress in the embryo. However, as we will see later, they leave their traces in the dermatoglyphic patterns.

More information on volar pads can be found in [27, 96, 97, 106].

2.1.3 The development of the epidermal ridge system

Epidermal ridge development has been described by various researchers [11, 19, 38, 43, 46, 49, 77, 80, 86, 93]. Their main results mostly coincide regarding the most important events and their timing, although discussion and interpretation of these observations varies.

The times given in the literature for the initiation of ridge formation varies from the 10th to the 13th week of pregnancy. Often no times are given but, instead, the
Crown-Rump-Length (CRL) of the embryo is used to denote embryonic development. Usually a CRL of 80 mm is given for the onset of ridge formation.

It is observed, at this time, that the basal layer of the epidermis begins to appear slightly undulated (see Figure 2.4). These undulations become quickly more pronounced and are called primary ridges (see Figure 2.5). Their formation process is described by different researchers as follows.

**Penrose [86]:** "... there is undulation in the basal layer of the epidermis"

**Hirsch and Schweichel [49]:** "In the fourth month ... the first distinct, sharply delineated, foldlike proliferations of the epidermis make their way in to the mesenchyme."

**Hale [46]:** "The earliest primary ridges appear as localized condensations..."
in the basal cell layer. (They, the author) penetrate the superficial layer of the dermis . . .”

BONNEVIE [19]: “Solche Faltenbildungen . . . werden vorgefunden. . . Es sind dies die ersten Papillarleisten. “ (”Such foldings . . . are found. These are the first primary ridges.”, translation by the author)

SCHAEUBLE [93]: “Diese Faltenbildungen . . . sind die ersten im Entstehen begriffenen Papillarleisten der Ballen.” (“These foldings . . . are the first developing primary ridges of the pads.”, translation by the author)

The key words in these descriptions of this first and, as it turns out, crucial step in ridge formation are “proliferation” and “foldings”. It indicates the viewpoint the respective researcher has regarding to the formation of the observed undulations.

Almost no study indicates that the undulation pattern is established by a prepattern formed in the dermis. The only exception is the work by MOROHUNFOLA et al. [67, 68] who mention mesenchymal condensations in the dermis prior to primary ridge formation. However, we have considerable doubt regarding the validity of this interpretation (see discussion in Section 3.3). Therefore, the situation in fingerprint formation seems to be quite different than in hair follicle or feather formation where such condensations are reliably observed and are deemed necessary for the formation process.

Primary ridges do not form simultaneously at the palmar and plantar surface. The timing of their formation was investigated by BONNEVIE [19], SCHAEUBLE [93] and GOULD [38]. As a rule of thumb, the development on the planta (sole and toes) is retarded by a week compared to the development on the palma (palm and fingers). On the palma, primary ridges first form at the summits of the apical pads of the fingers and along the nail furrow. Ridge formation is then initiated in the interdigital palm areas and in the midpalm along the flexion creases. Later on, thenar and hypothenar
areas become ridged. Ridge development concludes with the proximal and then the middle phalanges of the fingers. On the sole the sequence of ridge development is similar.

BONNEVIE and SCHAEUBLE were interested in the place on the volar pads where the ridges appear for the very first time. They called this place "Papillaranlage". There does not seem to be an English version of this concept, so we will introduce the name ridge anlage for it. According to BONNEVIE and SCHAEUBLE the ridge anlage is a small patch of volar skin characterized by intense cell proliferation at the time of ridge initiation and an overall increased thickness of the epidermis. In whorls the ridge anlage often coincides with the center of the whorl (see Figure 2.7 (a), (b)); in a loop the ridge anlage coincides with the ridges making up the core of the loop. (see Figure 2.7 (c)).

The pattern on the apical pads is usually established by three converging ridge systems (see Figure 2.8 (a)). The first one is the ridge system established by the ridge anlage (BONNEVIE calls them pattern ridges), the second one is formed along the nail furrow (called mantel ridges), the third one forms just distally of the DIC crease (called basal ridges). When these ridge systems contact each other, both triradii and dislocations are formed. The triradii on the distal part of the palm arise in a similar fashion. However, the primary ridge pattern on the fingertip is sometimes established almost instantaneously as shown in Figure 2.8 (b).

It is very important to know that, in humans, the ridges follow the major flexion creases and ridge formation on the palm away from the pads starts along these creases [93].

GOULD’s thesis [38] provides another source of interesting facts concerning ridge spread over the palmar and plantar surfaces. Although few figures and no actual photographs are included, it still provides plentiful information that cannot be found elsewhere. It would be desirable to implement a similar study using more modern techniques that could shed more light on the important question how ridges spread
A tangential section through a finger at about the 11th week. A whorl has formed (from [11], courtesy of March of Dimes).

(a) A schematic view on the locations of the first primary ridges. They start to form at the nail furrow and in a localized area in the middle of the pad. (b) The center of a whorl forms. (c) A loop forms (from [19]).

in time over the palmar and plantar surfaces.

Gould confirmed most observations by Bonnevie and Schaeuble. Interestingly, she noticed that loops in the interdigital areas first form at the distal margin of the palm and as outlines of the interdigital pads whereas the pattern core is inserted later. If true, this would be a different mode of pattern generation than the one acting at the distal phalanges, where it is generally agreed on that ridges start to form at the ridge anlage. Her observations on the interdigital areas of the palm are somewhat in conflict to the ones of Schaeuble who emphasized that ridge formation starts at the summits of the interdigital pads.
FIGURE 2.8. (a) A whorl has formed, the three systems emanating from the ridge anlage, the nail furrow and the phalangeal crease start to merge and only the triradii are left unridged. (b) A tented arch forms, ridges form almost instantaneously on the volar pad (from [19]).

FIGURE 2.9. Sometimes the primary ridges spread from two ridge anlagen. Bonnevie believed that accidental patterns arise in this case. However, it seems possible that in the above fingers regular loop and whorl configurations will form. (from [19])

Generally, there seems to be a lot of variation in the timing of ridge development. Gould found a 97 mm CRL embryo with complete ridging of the distal phalanges whereas a 118 mm CRL embryo exhibited incomplete ridge development at the triradii in this area.

It is frequently mentioned [11, 46] that during fetal development of the hand the number of ridges increases to keep up with the hand's growth. This process of ridge multiplication is believed to be responsible for the formation of the many small defects (minutiae) in fingerprint patterns.
FIGURE 2.10. The percentage increases of primary ridge breadth, hand length and number of minutiae are given as a function of the CRL. It seems that ridge breadth and hand length growth occur at almost the same rate. Further, there is an increase of minutiae until a CRL of 170 mm (from [45], courtesy of the Wistar Institute).

However, the only available quantitative study [45] presents numbers that are puzzling. HALE measured the length of embryo hands and feet and the breadth of the primary ridges. One should expect that the relative increase of hand or foot length exceeds the relative increase in primary ridge breadth. But according to HALE's graphs the increase in foot length is only slightly larger than the increase in ridge breadth. In the case of the hand this difference is virtually nonexistent (see Figures 2.10 and 2.11). This is the more confusing because HALE's graphs are consistent with each other (so no printing error is likely) but his conclusions very clearly indicate that

"...the percent increment in growth of the hand and foot ... in the fetus is observed to exceed the growth of the ridge in breadth."

Overall it does not seem likely that there is a dramatic increase in the number of ridges. From embryo pictures we have estimations of a ridge separation of about 35 \( \mu m \) that are spread on a halfcylinder with a radius of about 600 \( \mu m \). On this
FIGURE 2.11. The percentage increases of primary ridge breadth, hand length and number of minutiae are given as a function of the CRL. The ridge breadth growth slightly lags behind the hand length growth at the beginning of ridge formation. Later on there is hardly any difference between these rates. Many minutiae are formed in the time between 75 mm and 150 mm CRL (from [45], courtesy of the Wistar Institute).

halfcylinder about 50 ridges could form, not much less than are usually found on a finger.

There are strong clues that the ridge system significantly changes from the 13th to the 19th week because the number of minutiae increases [45] in this time (see Figures 2.10 and 2.11). Studying these changes in more detail would be a worthwhile research project. This period of primary ridge development ends at a fetal age of 19 weeks or a CRL of 150mm. Although the ridge system just begins to become visible on the outer surface at this time, the geometry of the ridge system is now established for life and will not change anymore.

At the 14th week, sweat gland ducts start to project from the bottom of the primary ridges into the dermis. Together with the increased proliferation pressure of the cells in the primary ridges, they are believed to establish the ridge pattern on the
After primary ridge formation ceases at the 19th week, *secondary ridges* appear as folds between the primary ridges. Their shape is similar to the primary ridges but they are shallower and do not contain sweat glands. Secondary ridges can be found between all primary ridges by 24 weeks. Now dermal papillae, peg-like protrusions of the dermis, invade the epidermis in the space between primary and secondary ridges, thus forming double rows [78]. A schematic view on the fully developed ridge system is provided in Figure 2.12.

Although the appearance of the dermal papillae changes throughout life and sometimes obscures the secondary ridges, the geometry of the primary ridges remains the same.

### 2.2 The connection between volar pad geometry and ridge patterns

There is considerable evidence that the shape of the volar pads influences the ridge patterns. Most of the evidence comes from empirical studies, observations of dermatoglyphics of mammals and marsupials and studies of dermatoglyphics of malformed hands.
2.2.1 Dermatoglyphics in mammals and marsupials

The first researchers that suggested a connection between volar pads and ridge configurations were Whipple [106] and Schlaginhaufen [96, 97]. They studied the volar surface of primates and some other mammals. In many primate species, the volar pads do not regress and can still be observed in adults. The palms of such primates (compare Figures 2.13, 2.14 and 2.15) display a close correlation between pads and ridge configurations that is hard to overlook. High, rounded pads display a whorl, with the center of the whorl sitting on the top of the pad. Here the ridges almost appear as elevation lines of the pads. Less prominent elliptic pads show elliptic configurations surrounding the line of highest elevation. Also loop and double-loop configurations may arise in this case.

When pads do not appear, likewise ridge configurations do not form and the epidermal ridges align in parallel fields. For instance, in humans the thenar pad and the first interdigital pad digress early in the embryo and rarely give rise to configurations. Similarly, secondary pads give rise to configurations in places where they are not usually found, for instance on the lower phalanges of the fingers of certain primates such as Cebus fatuellus.

It is important to point out that an eminence itself does not necessarily lead to configurations like whorls and loops. As an example we refer to the thenar area, which is already covered by a large eminence in the embryonic stage, whereas the thenar pad digresses early. Another example is the calcar area (heel) of the foot which exhibits large eminences but which is not covered by volar pads. Pattern configurations are rarely found in these areas.

On the lower phalanges of the digits of different monkey species a variety of ridge directions can be observed. The ridges most frequently run parallel to the phalangeal flexion creases as in humans and all other anthropomorphae, but they can also form wedges pointed distally like in Nycticebus tardigradus. Other, rarer, forms are also
FIGURE 2.13. Palm and sole of the night monkey, *Aotus zonalis*. The palmar pads are very pronounced (from [28]).

FIGURE 2.14. The dermatoglyphic patterns of the monkey in Figure 2.13. It is easy to see the correlation between palm topography and ridge alignment (from [28]).
WHIPPLE and SCHLAGINHAUFEN developed interesting theories that relate the volar pads and the patterns on them to the way of life of the respective animal [96, 97, 106]. The volar pads seem to serve as cushions for walking. The ridges are thought to increase friction, especially important for animals in arboreal habitats, and increase nerve sensibility. WHIPPLE conjectured that the ridges are aligned in a way to best prevent slipping which nicely explains why the ridges surround the summit of the pads in nearly concentric circles.

Some primates, such as Lemur brunnens, Galago garnetti, Nycticebus tardigradus and Midas rosalia, have volar skin that exhibits more complex structures than just ridges. In Lemur brunnens and Galago garnetti we find flat areas and depressions between small patches of ridges. Sometimes these patches are quite small and exhibit only a single circle of a ridge or even a single sweat–pore. The morphology of these patches was examined in more detail by SCHLAGINHAUFEN who prepared histologic sections of the dermis–epidermis interface underneath them. This interface proved to be much more complex than in humans. Often there is a greater space between two
primary ridges which indicates the boundary between two ridge patches.

Unfortunately SCHLAGINHAUFEN and WHIPPLE did not investigate the dermis-epidermis of Nycticebus tardigradus and Midas rosalia, because it is possible that the situation is somewhat different in these two species. The ridges disappear in depressed areas, however, we do not see ridge patches but so-called warts, circular structures that surround a single sweat-pore. These structures are clearly homologous to the ridges. First, they have the same morphology as the ridges surrounding a single sweat gland duct. Second, they often gradually lengthen and become regular ridges. Because of the missing histologic evidence it is not clear if the wart pattern is due to a dermis-epidermis interface that displays a dot pattern or if the underlying primary ridges are only brought to the surface close to the sweat gland duct. It seems that the evidence supports the latter hypothesis. For one, the warts are usually arranged in lines. Further, they often do not appear circular but are deformed, presumably following the primary ridges. In fact, it has been suspected that the sweat gland ducts play an important role in establishing the surface pattern [49]. This idea is supported by the sequence of events that establishes the ridges on the skin surface. At first small eminences around the sweat gland ducts appear forming a dot pattern. This dot patterns is consequently replaced by ridges as the embryo matures.

However, there are examples of true dot patterns on the dermis-epidermis interface in some species. They can be found in marsupial animals such as the vulpine phalanger (Trichosorus vulpecula) and the koala (Phascolarctos cinereus). DANKMEIJER [29] observed that the volar surfaces of various marsupials exhibit areas covered by warts and areas covered by ridges. Proof that these surface phenomena correspond to changes in the dermal-epidermal interface was established by OKAJIMA [79], who removed the epidermis of these specimen and stained the ridges using toluidine blue. The palm of the koala (see Figure 2.16) exhibits no visible pads and the whole surface shows warts which seem to be arranged roughly hexagonally. Only on the apical phalanx of the digits, ridges, which follow the outline of the finger, can be seen. Even
FIGURE 2.16. The dermal surface of a koala. Spots are observed on the palm and ridges form on the finger apex along the nail furrow (from [79], courtesy of March of Dimes).
FIGURE 2.17. The dermal surface of a vulpine phalanger. Both spots and stripes are observed. In 20b the transition from stripes to dots can be observed (from [79], courtesy of March of Dimes).
more interesting is the situation in the vulpine phalanger (see Figure 2.17). Here interdigital, thenar and hypothenar pads are well-developed and covered with ridges. However, contrary to the situation in most primates, the ridges do not appear as contour lines, but seem to run across the pads following the greatest curvature. The area between the pads is again filled with warts. Every wart consists of a few dermal papillae surrounding a small polygonal area that corresponds to the primary ridges. The space between the warts corresponds to the secondary ridges. Note that in these marsupials the sweat glands appear on the bottom of the secondary ridges, not on the bottom of the primary ridges, as usual in mammals. Overall, it seems that, again, geometry influences the volar pattern (not only pattern direction but also pattern type), although it is likely that ridges are formed in a somewhat different mode than in other mammals.

Hexagonal dot patterns with remarkable regularity are found on the dermis-epidermis interface of the dog, unfortunately we have to rely for this observation on only one photo which shows a small part of the palm [79].

Finally we want to mention the usage of rats (Rattus norvegicus) for studying dermatoglyphics experimentally [79]. Although rats do not show dermatoglyphic features on the volar skin, the dermis-epidermis interface displays ridges and forms distinguishable patterns. So far, to our knowledge, the results of this approach are modest, but it enables the much-needed possibility of testing hypotheses experimentally using a common laboratory animal.

2.2.2 Empirical studies

From the primate studies it was learned that high and rounded pads are related to the formation of whorl patterns. This hypothesis is supported by the work of BABLER, who studied the formation of fingerprints in embryos [8, 9, 10]. Because the volar pads digress starting from the 10th week, we should expect to see more whorls if
ridge formation is initiated early. Babler studied embryos which already exhibited recognizable patterns. He divided them in two groups, a group of embryos of 55–85 mm CRL and a group of embryos of 86–115 mm CRL. The first group can be thought of as embryos where ridge formation took place early and the second group as a control.

Babler observed that, in the early ridge differentiation group, 95.2% of all fingers displayed whorls, which is far more than the observed frequency in human populations, 4.8% displayed loops and no arches were found (compare to the first row of Table 2.1). In the second group, the frequency of whorls was 30.4%, the frequency of loops was 52.2% and the frequency of arches 17.4%. These frequencies for whorls and loops are comparable to the ones observed postnatally. However the arch frequency is significantly higher (compare to the first row of Table 2.1). Babler included in his study spontaneous abortions that show an unusually high frequency of arches. He speculated that some of these abortuses had undetected congenital defects that could have influenced the pattern type. It has been known for a long time that unusual dermatoglyphics is associated with chromosomal defects and harmful developmental events.

Further, there is strong evidence that asymmetries in the volar pad produce asymmetries in the ridge pattern. This evidence was obtained by Bonnevie [18] who observed that the pattern frequency is not the same on all digits (see Table 2.1). For

<table>
<thead>
<tr>
<th>Finger</th>
<th>Whorls (%)</th>
<th>Radial Loops (%)</th>
<th>Ulnar Loops (%)</th>
<th>Arches (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>25.65</td>
<td>5.81</td>
<td>61.14</td>
<td>7.40</td>
</tr>
<tr>
<td>I</td>
<td>35.04</td>
<td>0.34</td>
<td>60.71</td>
<td>3.91</td>
</tr>
<tr>
<td>II</td>
<td>28.89</td>
<td>23.98</td>
<td>30.66</td>
<td>16.47</td>
</tr>
<tr>
<td>III</td>
<td>16.22</td>
<td>2.31</td>
<td>70.44</td>
<td>11.03</td>
</tr>
<tr>
<td>IV</td>
<td>37.10</td>
<td>0.78</td>
<td>58.71</td>
<td>3.41</td>
</tr>
<tr>
<td>V</td>
<td>11.01</td>
<td>1.64</td>
<td>85.18</td>
<td>2.17</td>
</tr>
</tbody>
</table>

Table 2.1. Configuration percentages of 24,518 Norwegians. Values are given for each individual finger and for all fingers together (from [18]).
instance digits I and IV often display whorls, and arches are more often found on digits II and III. Digit V has almost always ulnar loops, whereas radial loops are almost exclusively found on digit II. Although different human races have different pattern frequencies and the frequencies differ between left and right hands, these observations are still true.

BONNEVIE [21] offered a beautiful explanation for these observations. She studied how symmetric or asymmetric the pads on the different digits are. Digit I and IV proved to be the most symmetric, and digit II and V proved to be the most asymmetric. BONNEVIE explained these symmetry relations by the way the fingers form from the embryonic handplate. The thumb separates early from the other digits and preserves a fairly symmetrical appearance. Digits III and IV are still fairly symmetric because they are cut out from the middle of the handmass. The end fingers of the handmass (fingers II and V), however, display the greatest asymmetry because they are cut out at the end of the handmass (see Figure 2.18 and 2.19). They are slanted toward the center of the handmass. Therefore they have opposite asymmetry which is reflected by the high occurrence of radial loops on digit II and of ulnar loops on digit V.

These considerations were supported by measurements of 130 embryo fingers. BONNEVIE considered transverse cuts through the finger at the position of the ridge anlage and determined whether the position of the ridge anlage was ulnar, radial or symmetric. To determine the asymmetry of the finger, BONNEVIE determined the location on these transverse cuts which is farthest away from the developing bones. She called this location the one of largest radius. Then she determined whether the location of largest radius was ulnar, radial or symmetric. This qualitative measure of asymmetry was augmented by a quantitative measure, called index of symmetry.

According to BONNEVIE's theories a symmetric pattern like a whorl should arise if the ridge anlage coincides with the position of greatest radius. Ulnar loops arise if the ridge anlage is found ulnar to the greatest radius and, similarly, a radial loop will
FIGURE 2.18. Cross-sections through the fingertips of the right hand of DONNEVIE's embryo No. 83. The thumb is on the right side. The separation of fingers has just occurred and the pads have started to appear. Note that digit I appears fairly symmetric, digit II is slanted slightly to the ulnar direction, digits III and IV are slightly slanted radially and digit V is strongly slanted to the radial side (from [19]).

arise if the ridge anlage is radial to the position of the greatest radius.

Even a casual look over the numbers reveals clear differences between the fingers. For instance, there is no example of a radial greatest radius on digit I or II and no example for an ulnar greatest radius on digit IV and V.

DONNEVIE calculated the percentages of these three cases for every digit and compared them with the known pattern statistics. Although the numbers are not correct for all digits, they are still quite interesting. The high number of radial loops on digit II is reproduced, whereas no radial loops are predicted for digits IV and V, where they are in fact very rare. Further, it is confirmed that digit IV has the highest degree of symmetry, which is reflected in the high number of whorls that can be found on this digit.

Differences in the predicted percentages (sometimes quite large ones) may be accounted for by the fact that the number of fingers observed was too small to estimate reliable numbers, that the scheme does not account for arches, and that the symmetry of the finger cannot solely be judged from one cut alone.
FIGURE 2.19. A scheme for the separation of the fingers. At first the thumb (digit I) separates. The remaining fingers form a mass with rounded ends. When these fingers separate fingers II and V retain the rounded outline on one side and become slanted toward the center of the handmass. After separation has finished fingers II and V will exhibit stronger asymmetries than the other fingers.

Apart from these considerations there is another strong hint that the symmetry relations of a finger play a significant role in determining the overall pattern. In BONNEVIE’s material there was not a single instance that an ulnar ridge anlage developed on a finger with a radial greatest radius or vice versa. This proximity of the ridge anlage to the greatest radius is just another example of the observation made on the palms and soles of monkeys: centers of patterns like whorls and loops are found on or very close to the summits of the pads.

However, there is a caveat that does not seem to be appreciated enough in fingerprint literature. Although there are many examples that confirm whorl patterns on high rounded pads, there are also counterexamples to this rule. The most striking one are the human interdigital pads that often appear as fairly pronounced elevations close to the distal outline of the palm. However, whorls in the interdigital areas are extremely rare. Therefore not only the geometry, but also other factors, are important here.

2.2.3 Dermatoglyphics in malformed hands

The literature on the effects of malformations of fingers, like syndactyly (webbing of fingers), hyperdactyly (excess number of fingers) or brachydactyly (reduced length of fingers), on their dermatoglyphics is very comprehensive. Here we will only describe
FIGURE 2.20. An example for hyperdactyly in two feet (the large toe and the small toe are doubled). The whorl patterns on the doubled hallux of the left foot are usual patterns on that digit, however the tibial loop on the hallux of the right foot is most unusual and could be explained by Bonnevie's symmetry arguments. Also the arches on the doubled small toes on both feet are unusual pattern types. On the left foot a triradius at the fibular side is seen. This indicates that yet another small toe was originally formed at that place (from [26], courtesy of the Wistar Institute).

the classic work of CUMMINS and some special topics.

In a paper from 1926 [26] CUMMINS, presented a summary of his observations on hands with various malformations and used them to confirm the hypotheses mentioned earlier in this chapter.

In cases of hyperdactyly where a digit is doubled and the separation of the doubled digit is complete one often observes patterns that look normal in the sense that their type can also be observed on the unaffected fingers. Often the pattern on the fingers that are doubled is the same. Sometimes a rarer pattern type (especially radial/tibial loops or arches) is observed on one of the fingers/toes, which indicates that the conditions for ridge alignment have changed. This corresponds to an observation by Bonnevie that the apical pads of affected fingers in hyperdactyly rarely develop
normally (see Figure 2.20).

If the separation is incomplete (for instance if the digits share one or more phalanges), unusual and irregular patterns that defy an obvious classification appear. If the surface of the two fingers is flat, the ridges run, as usual, transverse to the axis of the finger. However, in cases where the two fingers are only connected with a skinband, forming a trough between the fingers, the ridges sometimes turn in a longitudinal direction as they approach this trough.

Accessory fingers often have an accessory triradius at their base. This triradius is then all that is left when the finger is amputated in utero or by surgery (see Figure 2.20 for an example).

In cases of syndactyly somewhat regular patterns occur if the fingers are not too intimately fused. One can find two patterns on the mass that makes up the fused fingers. If the fusion is more complete the patterns become more irregular. Complicated loops and whorls with extra triradii are formed that, again, cannot be classified easily. The corresponding interdigital triradii often fail to appear and the distal portion of the palm is then covered by parallel ridges that do not form configurations.

In his conclusions, CUMMINS presented results that remind us of the observations presented in the previous sections. Portions of the volar surface that are covered by large eminences that have a well-defined circular border are covered by whorls with the center of the whorl being on the summit of the elevation. CUMMINS pointed out that the height of an eminence is not as important for a whorl to form as the fact that it is distinctly circumscribed over the surrounding area. Apical elevations with a distinct asymmetry develop loops, where the loops are oriented in the direction of the less steep descent. Flat areas exhibit ridge fields that do not form distinct patterns.

An interesting aberration of normal fingerprints is the rare so-called “ridges-off-the-end” syndrome [30]. Whereas in normal fingerprints the ridges tend to align themselves parallel to the nail furrow, this syndrome is characterized by ridges that
run vertically off the end of the fingers and toes. Very unusual patterns like cusps or patterns with one triradius but without a loop arise. Other dermatoglyphic anomalies are present as well.

Finally, we are going to describe some dermatoglyphic changes in cases of anonychia (absence of nails). BATTLE et al. [13] described dermatoglyphics in a family where Mackinder’s hereditary brachydactyly occurs. In affected individuals upper phalanges are missing or malformed. Often nails do not develop or form only partially. The authors observed that, in these cases, the triradii of loops and whorls are removed far distally and often only concentric circles of ridges surrounding the finger stump are found. Not even the slight curvature which a typical arch pattern exhibits is found in these cases.

PENROSE [84] presented another case of anonychia where the ridges also extend to the region which is normally covered by the nail. He observed whorl and loop structures in the region of the nail. Unfortunately, these ridges are only pictured as a diagram and no reference is given to the original source.
CHAPTER 3
THEORIES OF RIDGE DEVELOPMENT

In this chapter we will review the existing theories that attempt to explain fingerprint formation. We will refer to the theory that explains fingerprint formation by folding of the basal layer as the folding hypothesis. The idea that fingerprints are formed by a prepattern of either nerves or capillaries will be called the nerve hypothesis. Finally we will refer to the theory that claims that fingerprint patterns are first outlined in a prepattern of fibroblasts in the dermis as the fibroblast hypothesis.

3.1 The folding hypothesis and other mechanical hypotheses

3.1.1 Kollmann's work

The pioneer of the folding hypothesis was Arthur Kollmann. His 1883 published study [56] "Der Tastapparat der menschlichen Rassen und der Affen in seiner Entwicklung und Gliederung" offered the first attempt to explain how the epidermal ridges form. Some of his main ideas coincide quite closely with those of this thesis, which we will discuss later.

Kollmann's central concept is the "Seitendruck" (side or lateral pressure) that conditions ridge alignment. He observed that the shape of epidermal cells varies depending on whether they are subject to lateral pressure (leading to cylindrical or pyramidal cells) or not (leading to more varied cell forms such as spheres, cones or more irregular shapes). As an example he mentioned the primitive streak of the chick, which consists of tightly packed pyramidal cells. As cells migrate away from the primitive streak, they lose this shape. If they again organize in epithelial membranes, their pyramidal shape is recovered.
In the basal layer of the epidermis, KOLLMANN found cylindrical, lengthy cells whose long axis is perpendicular to the skin surface. He interpreted these shapes as an expression of lateral pressure acting on the cells. This lateral pressure is built up by intense cell proliferation in the basal layer, which is exhibited by many mitoses.

The only way this lateral pressure can be released is by invasion of the epidermis into the underlying dermis, which offers less resistance than the overlying epidermal layers. Or, in other words, buckling of the basal layer takes place. He argued that then the ridge direction will be perpendicular to the largest lateral pressure. The possibility that vice versa the dermis could invade the epidermis and be the pattern forming agent is dismissed by the remark that, in this case, a vertical pressure should occur, which produces basal layer cells that are flat.

Up to this point we agree with the main points of KOLLMANN's interpretations and, indeed, will offer very similar conclusions later on. However, we do not agree with everything he hypothesized.

For one, he not only explained the formation of the epidermal ridges using lateral pressure, but went on to relate the formation of virtually all skin appendages (like hairs, feathers, sweat glands, sebaceous glands) to lateral pressure and implicitly buckling. He did not offer an explanation how all these different structures are supposed to arise from a single mechanism. Also, he claimed that sweat glands are formed before the primary ridges, which is not true.

Further, he stated that the volar pads are not formed by accumulation of subcutaneous tissue, but are yet another expression of lateral pressure, an interpretation that has been largely dismissed in the literature [96, 106]. Another point of critique is his confusion of relative growth (which is the only one that is of importance here) and absolute growth and the erroneous conclusions he draws from it.

KOLLMANN interpreted the apical patterns as a superposition of longitudinal and transverse pressure. If longitudinal pressure dominates arches will arise. Otherwise a combination of the two will lead to the variety of patterns we know. This inter-
pretation is not convincing because too many questions remain open. What makes the difference between a loop and a whorl? How exactly does the superposition of longitudinal and transverse pressures act? We will attempt to answer some of these questions subsequently, using the concept of the stress tensor instead of what Kollmann called lateral pressure.

In spite of these shortcomings, this early work provided valuable insights in the biology of ridge formation and paved the way for the more sophisticated theories by Bonnevie.

3.1.2 Bonnevie’s work

Kristine Bonnevie must have been a remarkable person. In 1912 she became the first female professor in Norway and even today one can find her name on websites on the internet [2]. Her articles, mostly written in beautiful German, are examples for sophisticated and lucid scientific work. It is no exaggeration that her observations were the main inspiration for this work.

We have already mentioned that Bonnevie observed that primary ridges usually form first in a small localized area that she named “Papillaranlage”, or as we call it, ridge anlage. In a 1927 paper [19], she described the mechanism of ridge formation at the apex of the finger as follows:

"Die Keimschichtzellen sind hier höher, ihre Kerne länger; sie liegen auch sehr dicht gedrängt, ... Die äußeren 2–3 Lagen bilden hier ein aus mehr oder weniger stark abgeplatteten Zellen bestehendes Periderm. Diese Streckung der Peridermzellen ist desto mehr auffallend, je mehr man sich dem Zentrum der Papillaranlage nähert. ...

Das ganze Bild der Papillaranlage gibt auf diesem ersten Stadium den Eindruck einer lokalisierten, regen Zellproliferation in Str. germinativum der Epidermis, ein Eindruck, der auch durch zahlreiche Mitosen gestützt
wird. . . . Dieser Ausbreitung scheint jedoch durch den Widerstand der Peridermzellen eine Grenze gesetzt zu werden, während nach innen gegen das noch undifferenzierte welche Coriumgewebe hin, eine Faltung der heranwachsenden Keimschicht wohl möglich ist. . . . Es sind dies die ersten Papillarleisten. Die unter der Papillaranlage sichtbare Verdichtung des Coriumgewebes scheint nur als Druckwirkung von seiten der Epidermis hervorgebracht zu sein; . . . ”

("The basal layer cells are higher here, the nuclei are longer; they appear very crammed, . . . The outer 2–3 layers form the periderm consisting of more or less flattened cells. This extension of the peridermal cells is the more striking, the more one approaches the center of the ridge anlage. . . . The entire picture of the ridge anlage at this stadium gives the impression of a localized, lively cell proliferation in the basal layer, an impression that is also supported by many mitoses. . . . This expansion seems to be resisted by the peridermal cells, whereas a folding toward the inner, still undifferentiated and soft, dermal tissue seems possible. . . . These are the first primary ridges. The densified dermal tissue underneath the ridge anlage seems only to be caused by the pressure from the epidermis.", translation by the author)

This passage makes it clear that BONNEVIE considered folding or buckling to be the chief mechanism for ridge formation. After folding has taken place at the ridge anlage, more primary ridges are formed along the existing ridges of the ridge anlage, the nail furrow and the finger flexion creases. In this way, ridge systems arise that spread over the finger surface until they meet and form triradii. BONNEVIE indicated that the curvature of the volar pad influences the way these systems spread, although she was not very explicit as to how this would happen.

BONNEVIE did not observe the increased cell proliferation in areas away from
the ridge anlage. However, she explained the formation of stress by tension of the periderm, which puts a compressive stress on the basal layer. To release the pressure, folding takes place. She noted that the epidermis flattens out in areas with strong ridge development and argued that the energy stored in the periderm is now released. Later on, we will offer a different interpretation for this phenomenon.

The tension of the periderm cannot explain a folding process in the palm because there is little or no curvature in these regions. But still a buckling process could operate here due to cell enlargement and cell proliferation that puts pressure on the basal layer. The fact that mitoses are not observed in areas away from the ridge anlage does not necessarily mean that they do not take place. Gould [38] pointed out that, even after fetal death, the mitotic process tends to become completed so that the typical spindle patterns disappear.

It is clear from Bonnevie’s investigations that the position of the ridge anlage influences the pattern type. Apparently the ridge anlage is always close to the greatest radius of the finger (the position on the pad farthest away from the phalangeal bone). However, deviations do occur and the position of the ridge anlage apparently varies

Figure 3.1. A cross-section through an embryonic fingertip. The two papillary nerves (N.p.) converge towards the ridge anlage (Pap.) (from [19]).
greatly in longitudinal direction.

Another observation of BONNEVIE establishes a very interesting link between the ridge anlage and the nervous system. The embryonic fingertip is innervated by two main nerves. Both of them project toward the skin surface and converge to each other. BONNEVIE noticed that they finally end underneath the proximal border of the ridge anlage. This observation holds also true in cases where two ridge anlagen are observed. In these cases, the nerve twigs fail to converge and a nerve projects to each ridge anlage. BONNEVIE believed that this phenomenon is the reason for so-called accidental patterns, such as double loops or more complex whorls. This relationship between ridge anlage and nerve innervation has been disputed by some authors but was recently confirmed by DELL and MUNGER [31] and MOORE [66].

A crucial question concerning causality now arises: Does the nerve determine the location of the ridge anlage or, vice versa, is the nerve attracted in some way by the ridge anlage, or are both phenomena created by a third factor? BONNEVIE tried to answer this important question in several of her papers, however, she could not present a convincing answer for this problem. We will come back to this issue later in Section 5.1.

In addition to the position of the ridge anlage and the geometry of the volar pads, BONNEVIE mentioned yet another factor that influences the pattern on the fingers. This factor is related to the properties of the epidermis. BONNEVIE noticed that the epidermis is thickened in some embryos [20, 22]. The layers of the epidermis, including the basal layer, have enlarged cells, due to an increase in cytoplasm and nucleus volume. The borders of the different layers become obscured. BONNEVIE conjectured that this increase in volume is due to increased absorption of water. She called this phenomenon cushioning. Such cushions can spread over the whole volar surface and even cover the dorsal side of the finger where the nail develops. In other cases they spread only over a smaller part of the finger.

Cushions seem to be important for ridge development in several ways. At first,
FIGURE 3.2. Two cross-sections through an embryonic fingertip. The left figure reveals that two ridge anlagen are present. A more proximal cut shows incomplete convergence of the papillary nerves (N.p. I and N.p. II) that point to the centers of the ridge anlagen (Pap. I and Pap. II) (from [19]).

they influence the shape of the finger. Figures showing pads with partly thin, partly cushioned epidermis show that the epidermis is well rounded in regions with thin epidermis but appears stiffer and less rounded in regions with cushioned epidermis. BONNEVIE interpreted this finding as showing that there is an inner pressure of the mesenchymal tissue that shapes the pads. In the case of cushioned epidermis, there is more resistance and the fingers are not as rounded. In some extreme examples of heavy cushioning, the fingers remain in the blocky shape they were given when the fingers originally separated from each other.

Dramatically, this fight between inner mesenchymal pressure and resistance of the epidermis is decided in a few cases by violent rupture of the epidermis. In this case, BONNEVIE conjectured that aberrant ridge patterns may arise like the ones described by ABEL [3], which exhibit irregular, labyrinthian, or even completely dissociated ridge patterns (see Figure 3.3).
There is yet another connection between cushioning and fingerprint patterns. BONNEVIE noticed that embryos with cushioned epidermis on all fingers exhibit arches or tented arches on all fingers. She also observed that ridges form almost simultaneously on the apical phalanges although on every finger a ridge anlage can be identified. BONNEVIE called this fast mode of ridge spread *continuous* and claimed that it does not allow for the formation of triradial. She further observed that the ridges spread continuously in areas of cushioned epidermis if the center of the ridge anlage is itself located in a patch of cushioned epidermis. Otherwise discontinuous spread occurs with the formation of triradial.

We do not think that the formation of triradial (and hence loops and whorls) is solely a matter of continuous ridge spread or cushioning but a topological necessity when ridge patches of different orientation converge toward each other. However, the effect cushioned epidermis apparently has on ridge alignment proves that properties of the epidermis itself influence fingerprint pattern formation. At this point, however, it is not clear in what way cushioned epidermis differs mechanically from uncushioned epidermis.

BONNEVIE explained the greater thickness and opaque appearance of the cushioned epidermis by greater fluid absorption. Caused by this fluid excess the inter-
mediate layer sometimes detaches from the basal layer and the volume in between is filled by a fluid. These blisters open in extreme cases and the outer layers of the epidermis are shed away.

Bonnevie did not offer a conclusive mechanism explaining how cushioning takes place. She speculated that the crucial events start early in hand development before the nail anlagen form because cushioned epidermis can also be found on the dorsal side of the hand.

Schaeuble was able to confirm Bonnevie's observations in the configuration-forming region of the distal palm, around the interdigital pads. He also found regions of increased proliferation at the summit of the pads, which he called ridge anlagen as well. Apparently, there is the same relation between the ridge anlage and innervating nerves as in the ridge anlagen of the fingertips. Another important observation concerns the spread of ridges on the palm away from the interdigital pads. He found that the formation of ridges starts along the major flexion creases. These ridges are parallel to the flexion creases and spread out to cover the whole palm.

The folding hypothesis, as it was described by Bonnevie, was accepted by German researchers of the 1930s such as Abel [4] and Steffens [102] and guided them in their thinking about fingerprint phenomena. However, the idea almost became extinct in Germany after the second World War. It also seems that the folding hypothesis never became popular among English speaking fingerprint researchers.

3.1.3 Other theories based on mechanics and criticism of the folding hypothesis

The idea that mechanical forces determine the ridge patterns was popularized by Harold Cummins, one of the most influential researchers in the field of dermatoglyphics. Based on his observations on malformed fingers he argued in a 1926 paper [26] that there is a connection between hand geometry and ridge patterns. He be-
lieved that this connection is established by growth forces acting on the fetal skin. However, even though he included a chapter named "The mechanism of conditioning ridge direction", he did not specify exactly how the ridge configurations are established. Instead, he wrote:

"Although attempting to show that ridge direction is governed by mechanical epigenetic factors and to demonstrate the medium through which such factors operate, the present paper does not embrace an analysis of the specific histomechanics of ridge alignment."

And he went on:

"It seems unnecessary, therefore, to review an extensive literature dealing with the mechanical factors factors of growth, especially when a full discussion of the subject is presented by Ludwig ... in his work on hair direction".

Although possible, it is not at all obvious that there is a connection between hair direction and ridge alignment. Cummins did not even give a hint where the growth forces that control ridge development are supposed to come from. There is nothing in his vague descriptions that could be used as a basis for a mathematical model.

Later researchers followed a similar path: describing the connection between the pad geometry and the ridge patterns without giving a model how this connection is actually established. It is for this reason that the "mechanical" hypothesis did not become universally accepted and competing theories arose.

Typical for a discussion of the connection between pad geometry and ridge pattern is the well-researched article by Mulvihill and Smith [69]. They hypothesized that

"...the dermal ridge configurations are the immediate result of physical and topographic growth forces ...."

So far we can agree, however they continue to say that these forces act on a
"...skin which is predisposed in a polygenic manner to form parallel dermal ridges."

It is essential that epidermal ridges form parallel curves with a defined wavelength. In the opinion of the authors this is somewhat mysteriously achieved by the hard teamwork of the genes, but even these have to find a way to do the job. The significance of the mechanical forces is solely seen as a kind of bias that arranges the ridges. One can hardly say that the influence of Mulvihill's topographic forces on the ridges is immediate if a hidden background machine is needed to accomplish the task.

Mulvihill and Smith allude to several mathematical principles possibly involved in ridge alignment: that the ridges are laid as economically as possible (in what sense?); that the patterns resemble solutions of differential equations or lines of navigation; that the triradii are reminiscent of soap films under tension. However, it is unlikely that there is a connection to any of these fields.

Mulvihill and Smith deserve credit for summarizing all the evidence that geometry and forces are the right concepts for understanding fingerprint patterns. However, they did little to identify the way in which these factors act. It is telling that the folding hypothesis is not even discussed in their work.

Indeed, there seems to exist a language barrier for the English-speaking fingerprint community when it comes to the works of Kollmann and Bonnevie. Although their works are frequently referenced, their content is often represented wrongly or incompletely.

Explicit criticism of the folding hypothesis is rarely found. An example are Wilder's harsh words in the introduction to [106]:

"His (Kollmann's, the author) study of the ridges in the embryo has the historic value of being the first work devoted to the subject, but his errors in interpretation are such that his conclusions are in the main unreliable and misleading."
FIGURE 3.4. On this section the epidermis is shown under the dermis. The basal layer is clearly seen darkly stained. Formation of primary ridges has just been initiated on the outside parts of the section and is more advanced toward the center. The very large number of basal layer cells in the center regions indicates that cell proliferations not only lead to compressive stress and folding, but also deepen the primary ridge once folding has occurred (from [19]).

Unfortunately WILDER did not specify in what sense KOLLMANN’s interpretations are “misleading”, especially because KOLLMANN’s observations and theories are not much different from the ones of other researchers.

Another typical caveat is mentioned by WHIPPLE [106] who wrote:

“These ridges of the under epidermic surface were called by BLASCHKO “Falten” (folds, the author) because he believed them to arise as folds and not as true proliferations, a standpoint which has not been universally accepted.”

This short passage illustrates a commonly mentioned alternative to the folding hypothesis: that the primary ridges purely arise as proliferations of the stratum basalis. This hypothesis has a lot of problems: for instance, it is not obvious how the wavelength is organized or why the pad geometry influences the ridge alignment. To our knowledge, no mechanism based on pure cell proliferations of the stratum basalis leading to fingerprint patterns has ever been suggested. Increased cell proliferation of the lower layers of the epidermis, however, is an important ingredient of BONNEVIE’s ideas. Further, it is plausible that the topology of the primary ridge pattern is established by folding and then, once the ridges have been created, their depth is increased due to cell proliferations (see Figure 3.4).
Another argument against the folding hypothesis was provided by Hale who wrote [46]:

"Another misinterpretation, evident in the bulk of the literature cited on the subject of developmental mechanics is the idea that the epidermal ridge is a folding produced by mechanical forces alone."

He elaborated as follows:

"If the primary ridge developed as the result of inequalities in growth rate between the stratum basalis and the stratum intermedium a loose bond must exist between the two. The stratum intermedium is derived from the basal layer cell layer and it is hardly possible that such a bond exists."

We do not think that bonds between the basal layer and the intermediate layer are necessary to cause buckling. Quite the contrary seems to be true, the absence of bonds would establish the basal layer as a sheet of cells whose stress response is independent from the surrounding tissues.

To make the folding theory more credible (or any other theory that is based on growth stresses and volar pad geometry) it is necessary to test it. This should be done by identifying sources for the growth stresses that are supported by biological observations and by studying the stress distribution that follows from then. It is not enough to draw pictures of fingertips that exhibit certain geometrical features (such as high pads or asymmetric pads) and superimpose the kind of pattern one expects (such as whorls or arches) as has been done by Wertheim and Maceo [104]. In fact, these kind of illustrations do not replace actual observations of ridge spread and obscure the fact that the relations they assert have not yet found a satisfying explanation.

However, we agree with the words of Newell-Morris and Wienker [75]:

We also suggest that some of the basic questions about the development of the dermatoglyphic system may be addressed initially with computer
modeling rather than with fetal materials themselves. For example, the hypothesis that pad topography and timing of pad elevation determine pattern type dates back to the early 1900s and has guided much of our thinking about the association of pattern type and pad morphology ..., but to date it has not been tested. It is exactly to this type of question that computer simulation leads itself;....

Indeed, it is the aim of this work to implement such simulations.

3.2 The nerve hypothesis

We have already referred to BONNEVIE's observation that nerve twigs of the apical nerves are related to the location of the ridge anlage and therefore the morphology of the ridge pattern as well. In more recent work it was frequently observed that there are connections between the systems of nerves and blood vessels and the positioning of the primary ridges.

HIRSCH and SCHWEICHEL [49] found nerve fibers surrounded by blood vessels (thus forming what they call a vessel-nerve pair) in the dermis projecting to the foot of the primary ridges. However, it is not clear whether the vessel-nerve pairs determine the formation of the primary ridges or if, vice versa, the vessel-nerve pairs are attracted to the site of the primary ridges. The authors speculate that the absence of epidermal ridges and unusual dermatoglyphic patterns could be caused by neurohypotrophies. This was confirmed by SCHAUMANN [94] who observed unusual dermatoglyphic patterns in cases of spina bifida. The severity of the syndrome was correlated with the abnormality of the dermatoglyphic patterns, in some cases no epidermal ridges could be found on the skin surface at all.

These ideas are related to the ones by BLECHSCHMIDT [17] who claimed that the ridges simply follow the network of capillaries underneath them. Unfortunately, he limited his observations to the distal and proximal parts of the fingertip in order to
avoid the "complicated individual development" in the center of the fingertip. The center of the fingertip is really the location of the interesting patterns. Unless there is data confirming a relationship between ridges and capillaries in this area as well, BLECHSCHMIDT's observations can be dismissed as mere coincidence.

More elaborately formulated, and with more evidence, the nerve theory was presented in a 1986 work of DELL and MUNGER [31]. Using light and electron microscopy of rhesus monkey fetuses, these authors identified growth cones of nerve fibers that project to the epidermis. These growth cones are organized hexagonally with a distance of 40µm between each other. This approximately coincides with the wave length of the primary ridges. This innervation takes place in embryos of 55 mm CRL well before any primary ridges become visible. The authors speculated that

"...the afferent nerve fibers provide a two-dimensional grid that could modulate the spacing and arrangement of the papillary or sweat gland ridges."

Therefore abnormal fingerprints could

"...reflect abnormalities of the differentiation of dorsal root ganglion neuroblasts as they integrate the dermatotopic map during dermal and epidermal differentiation."

The authors conceded that earlier researchers such as HALE failed to see innervation in all regions prior to primary ridge formation. However, they noted that the growth cone of an axon can be very hard to identify and does not always exhibit the typical axonal characteristics such as the presence of microtubules, neurofilaments and small mitochondria. Further DELL and MUNGER pointed out that other epidermal organs are also proposed to be neurally induced. Examples are teeth, taste buds and feathers. These findings on monkey embryos have been confirmed in a follow-up study by MOORE and MUNGER [66] on human embryos.
All these studies suggest that the nervous system plays an important role in the development of the ridge system. However, it seems unlikely that the nerves generate the pattern all by themselves. For instance, the ridge direction cannot be determined by the hexagonal planform of the innervating axons. Furthermore, it is not obvious how the larger scale configurations like whorls and loops are formed and what role the pad geometry would play in this process.

Further insight into the relationship between the nervous system and the ridge pattern comes from two papers by MOROHUNFOLA et al. from 1992 [67, 68]. The authors studied the development of the primary ridges at the opossum *Monodelphis domesticus*. This animal is advantageous for experimental study because the development of the hind- and forepaws at birth is comparable with human limbs at 10 weeks of pregnancy. Primary ridge development occurs entirely postnatally where (similarly as in humans) the hindpaw lags in development compared to the forepaw.

In the first paper MOROHUNFOLA et al. observed that, prior to primary ridge development, four events take place: “skin innervation, Merkel cell differentiation, mesenchymal condensation, and epidermal proliferation”. Well-developed volar pads were found that were well innervated and were the regions first covered by primary ridges. These observations seem to further support the hypothesis that there is an influence of innervation on ridge development.

Mesenchymal condensations before primary ridge formation would suggest that the dermis is the pattern forming system and the dermal prepattern is then imposed on the epidermis. If this were true it would support the mathematical model by BENTIL, which we will review in the following section. However, we remain doubtful because the figures do not support these conclusions. On Figure 7a (here reproduced as Figure 3.5) in the second part of the paper, the dermis is shown at the onset of ridge development. Even with good intentions, we are unable to recognize mesenchymal condensations below the shallow primary ridges.

In the second paper by MOROHUNFOLA et al., the problem was investigated ex-
FIGURE 3.5. The dermis and epidermis of the opossum at the time of ridge initiation (from [68]).

experimentally. The left lumbosacral ganglia which project nerves to the left hindpaw were removed. The right hindpaw served as a control. It was found that volar pad development in the affected limb is less pronounced and development of mesenchymal cells and the extracellular matrix was retarded. However, and this is crucial here: primary ridge development still took place in the absence of nerves and the ridge wavelength remained unchanged, although the initiation of ridges was delayed and the depth of the ridges diminished. Another interesting result is the fact that no clear alteration of ridge patterns could be demonstrated. Unfortunately no further details or figures concerning this important question were given.

The authors wrote:

"The absence of neuronal processes in most of the papillary ridges in the left hindpaw ... suggests that peripheral nerves are not an absolute requirement for papillary ridge development. Perhaps the decision to form papillary ridges is inherent in the volar pad epidermis and such a decision is probably made in utero when pads form. The role of sensory nerves may be that of a biological clock which regulates and synchronizes the timing of various developmental parameters including papillary ridge development." (not emphasized in original)
Apparently, innervation is important for a normal development of the pads and it seems that trophic factors play a role for initiating and modulating the ridge forming process. It is also very probable, after BONNEVIE's and SCHAEUBLE's work, that the projection of the papillary nerves toward the epidermis determines the center of patterns, and thus influences the development of the ridge system. However, there is good evidence that the ability of the epidermis to form primary ridges is intrinsic.

3.3 The topological approach and the fibroblast hypothesis

Topological methods for understanding fingerprint patterns were first employed by L. S. PENROSE [62, 84, 87]. The only feature that is considered in this approach is the fact that ridges form nearly parallel and nearly equidistant lines. At every point of this ridge field one can define a director field \( k \) (obtained from a vector field by identifying \( k \) and \(-k\)), which is aligned perpendicular to the ridge crest and whose amplitude is \( 2\pi/\lambda \) where \( \lambda \) is the pattern wavelength. The basic objects of research in this approach are the singularities of this wave director field which are also the canonical singularities of two-dimensional patterns with local translational and rotational invariance. It is well-known that loops (called \textit{convex disclinations} in pattern theory, see Figure 3.6 (a)) and triradii (called \textit{concave disclinations} in pattern theory, see Figure 3.6 (b)) are the fundamental point singularities of such fields [83]. Other singularities can be seen as a composition of these fundamental singularities. For example, a whorl (often called \textit{target} in pattern theory) is the combination of two loops (see Figure 3.7).

If we follow the direction of the field on a circle around a singularity counterclockwise the director field will rotate by a certain angle, which is called \textit{twist}. It is easily seen that the twist around a loop is \( 180^\circ \) and the twist around a triradius is \(-180^\circ \) (see Figure 3.6).

It can now be shown that the rotation \( \rho \) of the director field around an arbitrary
FIGURE 3.6. (a) Surrounding a loop core in counterclockwise direction results in a twist of the wave director of 180°. (b) Similarly, surrounding the center of a triradius gives a −180° twist.

FIGURE 3.7. (a) A whorl has a twist of 360°. This is not surprising considering it is the composition of two loops (b).
FIGURE 3.8. Consider a ridge field on a disk that is always perpendicular to the boundary of the disk. Then the twist of the corresponding director field along the disk boundary is $360^\circ$.

The palm with the fingers can topologically seen as a simple disk. If the ridges pointed everywhere perpendicular to the margin of the disk we would have $\rho = 360^\circ$ (see Figure 3.8). However, this is not the case in humans. At the fingertip and the wrist the ridges become parallel to the margin. It is easily seen that each such occurrence reduces the field rotation by $180^\circ$ (see Figure 3.9). Therefore we have

$$\rho = 360^\circ - 180^\circ \cdot (\# \text{ of digits} + 1)$$

where $D$ is the number of digits. Combining (3.1) and (3.2) yields the following formula, that relates the number of digits $D$, the number of loops $L$ and the number of triradii $T$:

$$T + 1 = L + D$$

It is called Penrose's formula. It was very useful for developing a new nomenclature.
of classifying fingerprints and can serve to check whether all singularities have been found [85].

R. Penrose [87] concluded from the occurrence of structures like triradii and loops that the ridge direction is not determined

"...by a field of force, or as the lines along which some quantity takes a constant value, say."

Instead,

"...we must expect something of a tensor character, such as stress or strain, or perhaps the curvature of a surface". (emphasized in original)

The reason for this argument is the fact that director fields can be described by a pair of functions $f$ and $g$ up to a sign as $\pm(f, g)$. The sign dependence can be avoided by taking the dyadic product with itself giving

$$\begin{pmatrix} f^2 & fg \\ fg & g^2 \end{pmatrix},$$

which is a tensorial quantity.
Around certain singularities (such as targets) the twist is a multiple of 360°. In that case we can define a vector field in the neighborhood of the singularity, and there is another topological invariant, called circulation. It can be found using the formula
\[
\text{circulation} = \oint_C k \cdot dx.
\]
For example, the circulation of a target is 0.

However, apart from these observations we do not think that the topological approach has more to say about fingerprint development. The topological approach uses too little of the observations that can be made about fingerprints. It does not take into account that a uniform wavelength is established. It does not relate pad geometry and ridge direction. It does not address small-scale defects. It does not even discriminate between a whorl and two loops.

Nevertheless, other pattern forming systems have been related to fingerprints solely based on the fact that they exhibit similar topological features such as triradii and loops.

An example is a 1978 paper by Green and Thomas [40] that deals with patterns arising in cultures of disaggregated human epidermal cells. The authors observed that keratinocytes in petri dishes form, after 30 to 40 days, directional patterns that are reminiscent of whorls or triradii. The conclusion that this has a relationship to fingerprint patterns, however, is daring. We have already mentioned that triradii and loops are generic features of director fields. The system is far from establishing a wavelength and the considerable amount of time (several weeks) needed for establishing the pattern also does not support the author's conclusions.

Another example are patterns that were obtained by Elsdale and Wasoff [32] in fibroblast cultures. Fibroblasts are undifferentiated elongated cells in the dermis that can potentially differentiate into muscle, connective tissue or fat cells. In culture, fibroblasts like to align themselves and form director fields. In their 1976 paper "Fibroblasts Cultures and Dermatoglyphics: The Topology of Two Planar Patterns"
Elsdale and Wasoff analyzed the topological index of certain defects and their relation to the boundary. Although they did not make a causal connection between fibroblast patterns and dermatoglyphics themselves, this connection has been suggested later on, for instance in Bard's interesting book "Morphogenesis" [12] or in Murray's "Mathematical Biology" [70].

Indeed, apart from similarities in the topology there does not seem to be a lot of evidence that fibroblast patterns determine dermatoglyphics. The flow of the pattern looks quite different than the one of fingerprints and fibroblast patterns do not select a wavelength. However, the fibroblast hypothesis has received considerable interest from the mathematical side and Bentil developed a model [15, 16] on its basis that was supposed to model fingerprint formation. This model is the two-dimensional version of a model first developed by Murray and Oster [71, 81] to describe the interaction between fibroblasts and extracellular matrix and the pattern-forming ability of these processes. A system of PDEs is based on the observation that fibroblasts can generate strong tensile forces that act on the surrounding extracellular matrix [48]. It also takes into account that fibroblast cells have been observed to move toward areas of greater stickiness. This phenomenon is called haptotaxis. The model equations include balance equations of the density of fibroblast cells, of the extracellular matrix and a force equilibrium equation.

The model was used to describe mesenchymal condensations that are believed to play a role in cartilage and bone formation. Certain parameter regimes lead to hexagonal or ridge patterns. In the chick such patterns in the density of dermal cells are believed to control feather development. The model has also been applied to wound healing. The authors made a good case that the model equations describe these processes adequately. Merits of the model are the fact that it generates patterns with a defined finite wavelength and is able to produce stripe and hexagon patterns, depending on the choice of parameters. Drawbacks are the many parameters and singularities in the dispersion relation.
On the other hand there are a lot of reasons why this model is not useful for understanding fingerprint development. We have mentioned that the topological properties are the only experimental connection between fibroblast behavior and dermatoglyphics, and this is not a strong one. BENTIL, himself, was not able to make his model more credible either. He noted that the "epidermal pattern" is a reflection of the "dermal pattern". Because it is the interface of dermis and epidermis that codes the pattern this is a trivial observation and does not account for much. He further misquoted the article of GREEN and THOMAS that dealt with cultures of epidermal cells. BENTIL, however, wrote:

"In vitro experiments by Green and Thomas ...indicate that patterns develop by a process of cell movement in the dermis, which first produces ridges and then curves these ridges into figures of increasing complexity ...." (not emphasized in original)

Certain other problems of BENTIL's model are evident as well. He used a model that describes how differences in dermal cell density arise. However, there is very little evidence that such condensations occur in the dermis prior to primary ridge development. The only paper mentioning these condensations is the one by MOROHUNFOLA et al. but the arguments they present are not conclusive. Further BENTIL did not tell us how these hypothetical condensations affect the dermis–epidermis interface. Although he mentioned the classical work of CUMMINS dealing with the connection between pad geometry and ridge direction, he did not use it for his analysis.

From all this we have to conclude that the biological foundations of BENTIL's model are very weak. This may still be tolerable if his model were able to explain a variety of fingerprint phenomena. However, this is not the case either. The only fingerprint feature that is recovered is the fact that they form roll patterns with a defined wavelength, and even this feature did not come easily. The model is only able to obtain stationary waves (as opposed to traveling waves) after the introduction of
a fairly obscure parameter that is supposed to model the long range interaction of fibroblast cells and extracellular matrix.

BENTIL presented some numerical results and claimed that there is considerable similarity between these experimentally obtained patterns and fingerprint patterns. We cannot follow these conclusions. Of course, two director fields will always resemble each other to a certain degree, but BENTIL’s computer prints exhibit features that are very unusual for real fingerprints. The most striking ones are distinct saddle structures that are never seen in dermatoglyphics. Finally, it seems that BENTIL’s work has been completely ignored by the fingerprint community and the quest for a mathematical description of fingerprints goes on.

Nevertheless, the fibroblast hypothesis has survived in the work of FLEURY [34]. In an as yet unpublished article, he dealt with the morphogenesis of fingers and branched organs. Apparently he took the connection between fibroblasts and fingerprints for granted, at least he gave no references and even used the term “fingerprint crystal” for the direction of fibroblasts. We are not competent to discuss the details of his arguments (that do not deal with fingerprint formation in the first place) but emphasize again that a relationship between fibroblasts and dermatoglyphics is unproven and probably nonexistent.

3.4 Miscellaneous theories

3.4.1 Evolution as the creator of fingerprints

Interesting material regarding the purpose of fingerprints and the evolutionary pressure to form certain patterns can be found in the works of WHIPPLE [106] and SCHLAGINHAUFEN [96].

WHIPPLE comments that ridges have the

“...purpose of increasing friction between the skin and the surface with
which it comes in contact, either in walking or in prehension. Incidentally
the ridges acquire an important tactile function."

Because slipping forces would act from the summit of a pad the concentric arrange­
ment of ridges on the pads prevents slipping best. WHIPPLE supported these ideas
with the fact that most primates only exhibit ridges on volar skin that is in contact
with another surface during walking or grasping.

SCHLAGINHAUFEN emphasized the role of epidermal ridges for the tactile senses.
It can easily be noted that the volar side of the hand is much more sensitive than
the dorsal side. Experimentally, SCHLAGINHAUFEN demonstrated that tactile spatial
resolution is higher in directions perpendicular to the epidermal ridges than in direc­
tions parallel to them. This could be a reason why the ridges point in many different
directions at the fingertips, where tactile resolution is most important.

The idea that spatial tactile discrimination is paramount for ridge alignment was
used for a mathematical theory by KOLOSSOFF and PAUKUL in 1906 [57]. They
argued that tactile discrimination is best in directions of zero strain. We do not want
to indulge in a closer discussion of their ideas because of some apparent problems.
Directions of zero strain need not generally exist (when the principal strains have the
same sign). These directions certainly depend on the kind of forces acting on the
skin surface. In general, there are not one but two directions of zero strain (if the
principal strains have different signs). And finally: how should the embryo anticipate
during its genesis what the future lines of zero strain are?

WHIPPLE and SCHLAGINHAUFEN have collected useful material to explain why the
ridge patterns are useful for the animals whose palms and soles they cover. However,
we should not be tempted to think that just “evolution” created these patterns.
Even the most useful feature in the construction plan of an organism has to be
implemented efficiently by a physical process. It is this physical process that we
attempt to understand in this work.
3.4.2 The Penrose hypothesis

L. S. Penrose formulated an elegant hypothesis of how the geometry of the volar pads could translate into the alignment of ridge direction and the formation of patterns [62, 86]. He thought that the ridges run along the lines of largest curvature of the skin. At first glance, there are a number of facts that speak in favor of this hypothesis. For instance, in the absence of pads, the ridges tend to run perpendicular to the axis of fingers in the direction of largest curvature.

The hypothesis has received some attention by mathematics and it was shown by Smith [100] and Mardia et al. [63] that all occurring fingerprint patterns can be reproduced as lines of curvature of certain surfaces. Unfortunately, the authors did not construct the surfaces explicitly and did not attempt to predict the ridge pattern from the volar pad geometry. Instead, they considered a two-dimensional quadratic form whose coefficients change continuously as a function of two variables. They determined the direction of the largest half-axis of the ellipse associated with the quadratic form and showed that the curves tangent to these directions at any point form triradii, loops or whorls. The curves do not necessarily run parallel to each other and do not quite resemble real fingerprints, but they exhibit the right topological structure. It is further worth pointing out that the authors did not use the fact that the quadratic form originated from the curvature tensor. Indeed, any two-dimensional tensor could give rise in the fashion outlined above to the patterns obtained. In fact Mardia et al. concede that "the dermatoglyphic ridges might follow what were originally lines of stress rather than lines of curvature."

A closer look at the examples of dermatoglyphic patterns in humans and primates reveals soon that there seem to be a large number of exceptions to Penrose’s rule. The hypothenar pad of the palm in Figure 2.15 shows ridges running along the summit of the pad, therefore clearly not in the direction of largest curvature. We have mentioned examples of primates before where the ridges do not run perpendicular to
the axis of the finger. Further, the hypothesis fails to reproduce the fact that high rounded pads produce whorls (i.e. concentric circles surrounding the summit). To understand this argument let us consider the surface of revolution that arises if the curve
\[ y = e^{-x^2} \]
is rotated around the \( y \)-axis. The resulting surface can be seen as an idealization of a pad. The meridians and parallel circles are the principal lines of curvature. Denote \( \kappa_m \) the curvature in radial direction and \( \kappa_\theta \) the curvature in circumferential direction. Obviously both quantities only depend on the distance from the origin \( r \).
A calculation reveals that
\[ \kappa_m(r) = \frac{4r^2e^{-r^2} - 2e^{-r^2}}{4r^2e^{-2r^2} + 1} \]
\[ \kappa_\theta(r) = -2e^{-r^2} . \]
It turns out that
\[ |\kappa_\theta| > |\kappa_m| \text{ if and only if } 0 < r < r_l \]
where \( r_l \) is the only solution of the equation
\[ 2r^2e^{-2r^2} + 1 - r^2 = 0 . \]
The solution can only be found numerically and we have \( r_l \approx 1.101882 \).

This means that the lines of greatest curvature are indeed concentric circles if we are close to the summit of the elevation. However, they turn to rays projecting from the center as we leave the vicinity of the summit. This kind of behavior would be predicted by PENROSE's hypothesis, but it is never seen in nature. As a by-product this calculation also rules out the possibility that the ridges follow the lines of smallest curvature.

Even more disturbing than the lack of empirical evidence is the fact that PENROSE was not able to provide a mechanism that could explain why the curvature actually plays a determining role in fingerprint development. All he wrote [86] is:
FIGURE 3.10. This is figure 2.22 from [62]. According to LOESCH pressure is built up in the lower layers of the epidermis and tension is formed in the upper layers. Note that the ridges follow the lines of greatest curvature.

“One possibility is that the cells in the lower part of the epidermis are, at a very early period, sensitive to curvature and that the slight pressure produced by the concavity in the primitive basal layer induces the cells to multiply and form folds, with rows of papillae, along the lines of pressure rather than in other directions.”

Unfortunately, L.S. PENROSE's untimely death did not allow him to publish a more elaborate discussion of his ideas. Such a discussion, however, can be found, to some extent, in LOESCH's book [62]. Her Figure 2.22 is reproduced here as Figure 3.10. It shows a curved surface indicating tension in the upper layers and pressure in the lower layers. However, it is wrong to believe that such a stress distribution will arise just because the surface is curved. It is true that a surface will exhibit such stresses if it was bent to acquire this shape. No such stresses will be present if a material is cast or if biological effects alter the stress distribution. So it is not true that the curvature necessarily puts a compressive stress on the basal layer cells. And even if there were such stresses, it is unclear how they, acting uniformly on all cells, are able to produce a pattern. Interestingly the ridges in Figure 3.10 run parallel to the compressive stress which is quite in contradiction to an opinion published earlier by L.S. PENROSE that the ridges are formed perpendicular to compressive forces.

Another problem of models using only curvature arguments is the fact that the
curvatures can change drastically even if the surface is only slightly perturbed. Embryo photos suggest that such perturbations actually occur, but dramatic changes in ridge direction are relatively rare and mostly confined to the center of the volar pads.

In view of these arguments, we have to reject Penrose's hypothesis and with it any other hypothesis that attempts to explain ridge alignment solely using curvature arguments.

3.4.3 Reaction–diffusion models

There has been one paper explaining fingerprint formation using a reaction–diffusion model. It was written by Nagorcka and Mooney and published in 1992 [72]. The authors were interested in reaction–diffusion models that are able to exhibit both stripes and spots. The model they used consists of two substances (activator and inhibitor) that diffuse in the epidermis and migrate at a certain rate into the dermis. Depending on this rate, stripes, imperfect stripes, hexagons and irregular spot patterns are produced. If the prepattern of the morphogens is established, primary ridges or hair follicles form because of increased proliferation in areas of high (or low) concentration of morphogens.

A very desirable feature of this approach is the fact that a single parameter is responsible for the transition between a spot pattern (such as hairs) and a stripe pattern (such as epidermal ridges). If a stripe pattern is produced, the model predicts that a secondary pattern consisting of maxima and minima will form within the stripes. This could explain how the sweat gland ducts form within the primary ridges. Maybe the ratio of the primary ridge wavelength to the the distance between the sweat gland ducts could give insight into the magnitude of the model parameters and possibly help to validate the model.

The authors of the model did not address the question as to what factors condition the formation of fingerprint patterns nor do they cite much literature about fingerprint
formation. In spite of the mentioned advantages there is not overwhelming support for the specific reaction–diffusion model NAGORCKA and MOONEY present. However, considering the connection between nerve development and ridge formation it is very likely that chemical signals play some role in the picture. They might just be responsible as an activator of cell activity, but it is not impossible that a reaction–diffusion system in the classic sense produces a prepattern that translates into increased mitotic activity and thereby into the formation of primary ridges. Such a model would have to take into account of what is known about the connection between volar pads and ridge formation.
CHAPTER 4

A BUCKLING MODEL FOR FINGERPRINT FORMATION

After reviewing the extensive literature dealing with fingerprint formation we think that the folding hypothesis is the most promising for understanding fingerprint formation. Our basic hypothesis is as follows:

*The epidermal ridge pattern is established as the result of a buckling instability acting on the basal layer of the epidermis and resulting in the primary ridges.*

The hypothesis is supported by the following observations and analyses; some of these arguments will be developed in later parts of this work:

- The basal layer can be considered as an elastic sheet that is subject to compressive stress (see Section 4.1).
- At the risk of stating the obvious, buckling of an elastic sheet directly results in folds. It is known that these folds appear nearly parallel and have a well-defined wavelength (see Section 4.4).
- If the stress is anisotropic ridges are the preferred pattern type. The ridges will form perpendicular to the greatest stress (see Section 4.4). Therefore, the ridge system is governed by a tensor quantity describing a director field (as opposed to a vector field).
- In certain cases, dot (hexagon) patterns, as found on some marsupials, are a possible buckling pattern (see Section 4.5).
- Mechanical concepts like forces are related to geometrical concepts like curvature. Therefore the folding hypothesis allows us to tackle the question why
fingerprints are related to the pad geometry. Indeed, in Chapter 5 we will find an answer to this question.

- Computer simulations of the model in Chapter 6 exhibit many correct fingerprint features.

After our analysis we will present a more refined version of the basic hypothesis.

We want to point out that in this work we are mainly concerned with the question of how ridge alignment is conditioned and how the patterns are laid out. Therefore we are mainly interested in the instability and the behavior of the system shortly after the instability. This restriction allows us to use linear and weakly nonlinear analysis of the well-known von Karman equations. Although we cannot understand all the features of the fingerprint system this way, we will gain information about wavelength, wave direction and pattern type which is essential information for any further more refined models. This procedure is promising because there is considerable empirical knowledge on the pattern wavelength and the relation between ridge direction and geometry.

In a system as complex as a developing embryo, it is very likely that not only mechanical effects like elastic buckling influence ridge development. Plasticity and viscoelasticity might play an important role. We have already mentioned that cell proliferations may deepen the primary ridges, once they have been created. Other biological phenomena might also be important. Such effects, if they can be isolated, could be incorporated in our present model. In this sense, the buckling model should be seen as the simplest model that captures the most important fingerprint features and serves as a starting point to which new elements can be added.

In this chapter, we want to give biological arguments for the validity of our hypothesis. Then we derive the shallow shell equations and use them to understand the influence of curvature and stresses on the ridge direction. Next we will use weakly-nonlinear analysis to investigate for what parameters ridges and hexagons are stable
FIGURE 4.1. The cytoskeleton of the basal layer cells are attached to each other by desmosomes and to the basal lamina by hemidesmosomes.

solutions. Finally, we will use the analysis results to estimate the involved model parameters.

4.1 Biological arguments

There are certain requirements that are necessary for a surface to buckle. These are elastic material properties and compressive stress. Further (in the absence of curvature) there has to be resistance in normal direction to obtain a finite wavelength. From our review of the literature we will argue that these requirements are met in the process of ridge formation.

1. The basal layer of the epidermis is a sheet of cells having elastic properties.

The cells of the basal layer are connected to each other by desmosomes. These are structures that link adjacent cells by tonofilaments that deeply penetrate the cell and attach to the cell’s cytoskeleton [86]. The basal layer cells are also attached by hemidesmosomes to a protein sheet called the basal lamina which acts as a seal between the dermis and the epidermis. Considering that these desmosomes are connected to each other by the cell’s cytoskeleton we can consider the basal layer as a sheet that is able to support compressive forces and resist bending.

Further, it seems that basal layer cells are more intimately connected to each other than to cells of the intermediate layer. This is suggested by BONNEVIE’s observations.
of blisters in the epidermis of certain embryos whose intermediate layer detaches from the basal layer. In these cases the structural integrity of the basal layer remains unaffected.

It is known that plastic and viscoelastic behavior also plays a role in the mechanics of epithelial tissue and cells can rearrange themselves if the shearing stress is large enough [65, 103]. We will not consider such effects, although they possibly play a role, especially after buckling has occurred and should be considered in a more sophisticated model.

2. Compressive stress acts on the basal layer.

The necessity of this requirement for buckling is obvious. It is not difficult to identify the source of the required stress. Many researchers explicitly mentioned the increased proliferation of basal layer cells at the time of primary ridge formation and connect this proliferation to the formation of compressive stress. BONNEVIE and KOLLMANN referred to histological sections of the ridge anlage which show many mitoses and basal layer cells that can hardly find place in the space offered. Also, the cylindrical shape of the basal layer cells with the cylinder axis pointing perpendicular to the skin surface is evidence that the basal layer is put under compressive stress by differential
growth.

These observations were confirmed and extended by Gould. She observed that, in relatively immature embryos, the basal layer consists of flattened cells. They are still flattened on the dorsal surface of hands and feet, as well as on areas with late ridge development, when primary ridges first appear on the distal phalanges of the digits. However, the basal layer cells become densely packed and highly columnar prior to primary ridge formation. Primary ridges first become visible as undulations of the dermis–epidermis interface. Later on, basal layer cells on the top of the ridges appear more elongated than between the ridges.

Further, we want to mention an interesting observation made by Okajima and Newell–Morris who examined the embryology of epidermal ridges in the pigtailed macaque (*Macaca nemestrina*) [80]. Using toluidine staining techniques, they found that, prior to ridge formation, the dermal–epidermal interface exhibits a fine crêpelike appearance of faint stripes that predicts the direction of primary ridges to come. To us, these faint stripes appear as stress lines of the yet unbuckled dermal surface. Admittedly, the authors of the study arrived at a different conclusions. A similar observation was made by Gould who observed “faint striations” of the dermis–epidermis interface before primary ridge formation occurs. Unfortunately, she did not provide images of these striations.

Bonnevie suggested another possibility of how compressive stress in the basal layer could arise. She noticed that mitoses are rare in areas away from the ridge anlage after primary ridge formation has been initiated at the ridge anlage. She conjectured that the stress could be put on the basal layer by contraction of the periderm in the volar pad areas. This might also account for the fact that the volar pads become less prominent during the time of ridge proliferation. However, even if a contracting periderm has some influence on the basal layer near the pads, it certainly cannot influence the basal layer in areas which are not covered by them (like the midpalm). Further compressive stress can be formed not only by proliferation of cells, but also
FIGURE 4.3. The differential growth produces compressive stress in the basal layer. For the subsequent buckling process, the dermis and the intermediate layer act as beds of resisting springs.

by a change in cell volume. This could indeed be the case, as Gould has observed such phenomena during the time of ridge differentiation. Unfortunately, many of these questions have not been experimentally addressed at this point.

In this work, we think of differential growth as cells in the basal layer just proliferating faster than the cells in the layers above. However, this approach is probably too simplistic. For one, the cells in the intermediate layer are mostly recruited from the basal layer. Also, it is possible that basal layer cells control the stress they are subjected to by some feedback loop. The processes of developing epithelial tissues and the chemical cues for cell growth or cell death are still an important and little understood part of current biological research [5, 12, 98, 105].

3. The stresses in the basal layer are resisted by other structures of the developing skin.

This sentence has to be understood in two ways. First the compressive stress in the basal layer has to be resisted. And second, there has to be normal resistance (in the absence of curvature) against normal displacement, so that a wavelength can be established.

According to Bondevie the compressive stress in the basal layer is resisted by the
tension in the periderm. This tension in the periderm is indicated by long elongated
cells. Together with the inner pressure of the mesenchymal tissue it is responsible for
the harmoniously rounded shape of the apical pads. BONNEVIE has shown [20, 22]
how disturbances in this balance (like ruptures of the epidermis or failure of sufficient
inner pressure) can lead to irregularities in finger and ridge pattern development. The
periderm cannot resist basal layer expansion in flat regions of the volar surface. Here
the resistance is provided by anatomical landmarks like creases and furrows and the
margin of palm and fingers.

If the stress in the basal layer reaches a critical point buckling will occur. Both
the intermediate layer of the epidermis and the dermis will resist this buckling. It is
quite evident, however, that the dermis with its less rigid structure offers much less
resistance than the stratified intermediate layer. Therefore buckling occurs towards
the dermis. The resistance of the dermis is represented by densified tissue underneath
the primary ridges as reported both by BONNEVIE and SCHAEUBLE.

There are a few other examples where buckling has been established to be respon-
sible for the formation of certain morphological features. Examples are the ciliary
body of the eye and, possibly, the furrows of the brain [12]. Recently buckling theory
has been used to understand growth patterns in plants [41, 42, 101, 99].
4.2 An overview of shell theory and derivation of the shallow shell equations

To understand the effects of local stress and local curvature on buckling we will use shell theory, developed by engineers to understand the statics and dynamics of thin curved bodies. A shell is a three-dimensional body that extends a small distance from a two-dimensional surface, which is usually called the middle surface. The thickness of the shell is denoted by \( h \). We will assume that it is constant.

There is actually not just one shell theory, but rather several ones. They all have in common that the three-dimensional elasticity equations are reduced to two-dimensional equations that are mathematically much more feasible. As the basic idea, the three-dimensional forces, moments and stresses are integrated over the thickness of the shell. Thus, new equations relating these quantities are obtained.

The different shell theories differ in the assumptions that are used. For example, the classic theories use the Kirchhoff’s assumption, that normals to the middle surface still remain normal after deformation. This hypothesis yields equations for special cases that are analytically solvable, but introduces inconsistencies in the theory. Kirchhoff’s hypothesis requires that the transverse shear stress is constant and nonzero throughout a section of the shell, although it is clearly zero at the upper and lower surface if no shear stress is applied there. These inconsistencies are tolerable when the thickness of the shell is small.

More advanced theories like the one by Reissner–Mindlin assume that normals remain lines (not necessarily normal anymore) after deformation. The resulting equations are significantly more complicated but this approach gives better results for thick plates. It is also easier to implement numerically using finite element methods. More information on shell theory is provided in [37, 39, 44, 76].

A specialized shell theory that is especially well-suited for analyzing buckling is connected with the names Föppl, Donnell, Mushtari, Vlasov who developed
the theory independently from each other. The resulting equations are analogous to the von Kármán equations of plate theory. The theory neglects certain linear terms and retains the most dominant nonlinear terms. These nonlinear terms are important for the analysis of the postbuckling behavior. The following assumptions are made in the theory:

1. The thickness of the shell is much smaller than the radius of curvature.
2. The thickness of the shell is much smaller than the plane dimension of the shell.
3. The thickness of the shell is much smaller than the buckling wavelength.
4. The plane dimension of the shell is much smaller than the radius of curvature. Shells satisfying this assumptions are called shallow shells.
5. The principal radii of curvature greatly exceed the buckling wavelength.
6. The derivatives in the normal displacement may be much larger than the in-plane strains but still have to be small compared to unity.

In a strict sense it has to be verified a posteriori if the assumptions were actually satisfied. In our application, the ratio of shell thickness to the radius of curvature is about 1:100; the ratio of shell thickness and buckling wavelength is about 1:5; and the ratio of wavelength to radius of curvature is about 1:20. Therefore conditions 1, 2 and 5 are clearly fulfilled. Condition 3 can still be accepted. Condition 4 is satisfied if we limit ourselves to a small patch of skin. Condition 6 is satisfied if we only consider the immediate postbuckling state when normal displacement derivatives are still small.

In the following, we derive the shallow shell equations. We assume that $x$ and $y$ are coordinates in the in-plane directions. To simplify the equations we assume that the $x$ and $y$ coordinate lines are lines of smallest and largest curvature, respectively. Further, we use the planar metric $ds^2 = dx^2 + dy^2$ (see Appendix A for a justification). The coordinate perpendicular to the shell will be called $z$. The displacements of the
The strains \( \epsilon_x \) and \( \epsilon_y \) denote the (small) relative increase of a small piece of the shell in the \( x \)-direction and \( y \)-direction, respectively. The strain \( \epsilon_{xy} \) is a measure for the geometric distortion.

In-plane stresses acting on a small area of the shell. As usual we have \( \sigma_{xy} = \sigma_{yx} \).

The shell in \( x \), \( y \) and \( z \) directions will be called \( u \), \( v \) and \( w \), respectively. The shell is assumed to have a constant thickness, \( h \). The surface \( z = 0 \) corresponds to the midsurface of the shell. The surfaces \( z = h/2 \) and \( z = -h/2 \) describe the upper and the lower surfaces of the shell.

We will first find an expression for the elastic energy functional \( E \). By taking functional derivatives with respect to the spatial coordinates we find the body forces \( B_x \), \( B_y \) and \( B_z \). That is we have [58]

\[
B_x = -\frac{\delta E}{\delta u}, \quad B_y = -\frac{\delta E}{\delta v} \quad \text{and} \quad B_z = -\frac{\delta E}{\delta w}.
\]
Setting these functional derivatives to zero gives us equations for static (i.e. time-independent) solutions. Including inertia and damping terms gives the dynamic equations which we are mostly interested in. We will ignore the normal stresses and only consider the in-plane stresses and strains. Therefore the elastic energy can be written as \[ \mathcal{E}_{el} = \frac{1}{2} \int_V \sigma_x \varepsilon_x + \sigma_y \varepsilon_y + \sigma_{xy} \varepsilon_{xy} \, dV \]

\[ = \frac{1}{2} \int_A \int_{-h/2}^{h/2} \sigma_x \varepsilon_x + \sigma_y \varepsilon_y + \sigma_{xy} \varepsilon_{xy} \, dz \, dx \, dy . \]  

(4.1)

Here \( \varepsilon_x, \varepsilon_y \) and \( \varepsilon_{xy} \) are the in-plane strains and \( \sigma_x, \sigma_y, \sigma_{xy} \) are the in-plane stresses (see Figure 4.6). The integration is performed over the volume of the shell. Next we use the stress-strain relations, which are simply Hooke’s law,

\[ \sigma_x = \frac{E}{1 - \mu^2} (\varepsilon_x + \mu \varepsilon_y) \]
\[ \sigma_y = \frac{E}{1 - \mu^2} (\varepsilon_y + \mu \varepsilon_x) \]
\[ \sigma_{xy} = \frac{E}{2(1 + \mu)} \varepsilon_{xy} \]  

(4.2)

or, solved for the strains,

\[ \varepsilon_x = \frac{1}{E} (\sigma_x - \mu \sigma_y) \]
\[ \varepsilon_y = \frac{1}{E} (\sigma_y - \mu \sigma_x) \]
\[ \varepsilon_{xy} = \frac{2(1 + \mu)}{E} \sigma_{xy} . \]  

(4.3)

As common, \( E \) denotes Young’s modulus and \( \mu \) is Poisson’s ratio. For constant \( x \) and \( y \), the strains still depend on the \( z \)-coordinate because of possible bending effects. To express this dependence we write \( \varepsilon_x = \varepsilon_x(z) \) and obtain

\[ \varepsilon_x(z) = \varepsilon_x(0) - zw_{xx} \]
\[ \varepsilon_y(z) = \varepsilon_y(0) - zw_{yy} \]
\[ \varepsilon_{xy}(z) = \varepsilon_{xy}(0) - 2zw_{xy} . \]  

(4.4)
Refer to Appendix A for a derivation of these formulas. Here and in the remainder of this section derivatives are denoted by commas. Substituting the expressions (4.2), (4.3) and (4.4) into the elastic energy density yields three terms: a term independent of \( z \), one that is proportional to \( z \) and one that is proportional to \( z^2 \).

The term independent of \( z \) is

\[
\frac{1}{E} \left( \sigma_x (\sigma_x - \mu \sigma_y) + \sigma_y (\sigma_y - \mu \sigma_x) + 2(1 + \mu) \sigma_{xy}^2 \right) = \frac{1}{E} \left( \sigma_x^2 + \sigma_y^2 - 2\mu \sigma_x \sigma_y + 2(1 + \mu) \sigma_{xy}^2 \right). \tag{4.5}
\]

Integrating over the thickness of the shell yields

\[
\frac{1}{Eh} \left( N_x^2 + N_y^2 - 2\mu N_x N_y + 2(1 + \mu) N_{xy}^2 \right)
\]

where \( N_x = h \sigma_x, N_y = h \sigma_y, N_{xy} = h \sigma_{xy} \) are the stresses integrated over the thickness as well. The term proportional to \( z \) vanishes after integration over the thickness of the shell. The remaining term, we write in terms of \( w \). It has the form

\[
\frac{z^2 E}{1 - \mu^2} \left( (w_{,xx} + \mu w_{,yy}) w_{,xx} + (w_{,yy} + \mu w_{,xx}) w_{,yy} + 2(1 - \mu) w_{,xy}^2 \right) = \frac{z^2 E}{1 - \mu^2} \left( w_{,xx}^2 + w_{,yy}^2 + 2\mu w_{,xx} w_{,yy} + 2(1 - \mu) w_{,xy}^2 \right). \tag{4.6}
\]

Integrating over the thickness of the shell gives us

\[
D \left( w_{,xx}^2 + w_{,yy}^2 + 2\mu w_{,xx} w_{,yy} + 2(1 - \mu) w_{,xy}^2 \right)
\]

where

\[
D = \frac{Eh^3}{12(1 - \mu^2)}.
\]

This gives the following expression for the elastic energy

\[
\mathcal{E}_{el} = \frac{1}{2} \int_A \frac{1}{Eh} \left( N_x^2 + N_y^2 - 2\mu N_x N_y + 2(1 + \mu) N_{xy}^2 \right) + D \left( w_{,xx}^2 + w_{,yy}^2 + 2\mu w_{,xx} w_{,yy} + 2(1 - \mu) w_{,xy}^2 \right) \, dx \, dy. \tag{4.7}
\]
The first term measures the energy due to traction or compression of the midsurface of the shell, the second term measures the bending energy.

Let us now assume that the shell is subjected to a perpendicular load that depends on \( w \). Such a load could be pressure on the shell or springs that resist normal displacement. Assume that this load has a potential \( V(w) \). For instance, the potential \( V(w) = pw \) corresponds to a normal pressure on the system; and \( V(w) = \frac{2}{3}w^3 + \frac{2}{3}w^3 \) corresponds to a nonlinear elastic foundation, where \( \gamma \) is the linear spring constant and \( \alpha \) the next order correction. Then the energy of the system is

\[
\mathcal{E}^* = \int_A \frac{1}{2Eh} \left( N_x^2 + N_y^2 - 2\mu N_x N_y + 2(1 + \mu)N_{xy}^2 \right) + \frac{D}{2} \left( w_{xx}^2 + w_{yy}^2 + 2\mu w_{xx} w_{yy} + 2(1 - \mu)w_{xy}^2 \right) + V(w) \, dx \, dy \quad (4.8)
\]

We wish to take functional derivatives. In doing this we have to take account of the fact that \( N_x, N_y, N_z \) are not independent of \( w \). These dependencies become apparent in the strain–displacement relations, that have the following form in our situation:

\[
\begin{align*}
\frac{1}{Eh} (N_x - \mu N_y) &= \epsilon_x = u_x + \frac{w}{R_x} + \frac{1}{2} w_x^2 \\
\frac{1}{Eh} (N_y - \mu N_x) &= \epsilon_y = v_y + \frac{w}{R_y} + \frac{1}{2} w_y^2 \\
\frac{2(1 + \mu)}{Eh} N_{xy} &= \epsilon_{xy} = u_y + v_x + w_{x,y} \quad (4.9)
\end{align*}
\]

According to these formulas, shell displacements cause strain due to three mechanisms.

- A gradient in the in–plane displacements as seen in Figure 4.7 (a).
- A normal displacement in the presence of curvature as seen in 4.7 (c).
- A gradient in the normal displacement with a slope of \( \tan \alpha \) illustrated in 4.7 (b).

After the deformation a small piece of length \( L \) now has length

\[
L_{\cos \alpha} = L \sqrt{1 + \tan^2 \alpha} \approx L \left( 1 + \frac{\tan^2 \alpha}{2} \right) = L \left( 1 + \frac{w_x^2}{2} \right)
\]

Note that this is a nonlinear contribution to the strain.
FIGURE 4.7. The three mechanisms responsible for the formation of strain, illustrated at \( \varepsilon_z \). The solid curves represent the undeformed state and the dashed lines the deformed state. See text for further explanation.
Here \( R_x \) and \( R_y \) denote the radii of curvature in \( x \) and \( y \) direction, respectively. In the last term we anticipate that the slopes \( w_{x} \) and \( w_{y} \) are not necessarily small after buckling has taken place. This way we introduce a geometric nonlinearity to our equations. Refer to Appendix A for a more rigorous derivation of the strain–displacement relations.

There are two equivalent possibilities of dealing with the strain–stress relations:

1. When taking the variation with respect to \( u, v \) and \( w \) represent \( N_x, N_y \) and \( N_{xy} \) in terms of the displacements and use the chain rule.

2. Formulate the problem as a restricted extremization problem. Use Lagrange parameters to incorporate the restrictions into the elastic energy expression and take variations as usual.

We chose the second approach because the algebra seemed to be somewhat easier. Accordingly, we introduce Lagrange multipliers, \( \lambda_1, \lambda_2, \lambda_3 \) (which are functions of \( x \) and \( y \) here), and formulate a modified energy:

\[
\mathcal{E} = \int_A \frac{D}{2} \mathcal{L}_1 w + \frac{1}{2Eh} \left( N_x^2 + N_y^2 - 2\mu N_x N_y + 2(1+\mu)N_{xy}^2 \right) \\
+ \lambda_1 \left( u_x + \frac{w}{R_x} + \frac{1}{2} w_{xx} - \frac{1}{Eh} (N_x - \mu N_y) \right) \\
+ \lambda_2 \left( v_y + \frac{w}{R_y} + \frac{1}{2} w_{yy} - \frac{1}{Eh} (N_y - \mu N_x) \right) \\
+ \lambda_3 \left( u_y + u_x + w_{,x} w_{,y} - \frac{2(1+\mu)}{Eh} N_{xy} \right) \\
+ V(w) \ dx \ dy . \quad (4.10)
\]

Here we have introduced the differential operator \( \mathcal{L}_1 \) which is defined as follows

\[
\mathcal{L}_1 w = w_{xx}^2 + w_{yy}^2 + 2\mu w_{,xx} w_{,yy} + 2(1-\mu)w_{,xy}^2.
\]

The functional \( \mathcal{E} \) depends on \( N_x, N_y, N_z, \lambda_1, \lambda_2, \lambda_3, u, v \) and \( w \). We first consider
the derivatives with respect to the stresses

\[ \frac{\delta \mathcal{E}}{\delta N_x} = \frac{1}{Eh} \left( N_x - \mu N_y - \lambda_1 + \mu \lambda_2 \right) \]

\[ \frac{\delta \mathcal{E}}{\delta N_y} = \frac{1}{Eh} \left( N_y - \mu N_x - \lambda_2 + \mu \lambda_1 \right) \]

\[ \frac{\delta \mathcal{E}}{\delta N_{xy}} = \frac{1}{Eh} \left( 2(1 + \mu) N_{xy} - 2(1 + \mu) \lambda_3 \right). \] (4.11)

Setting these expressions to zero gives

\[ \lambda_1 = N_x, \quad \lambda_2 = N_y, \quad \lambda_3 = N_{xy}. \]

This result, put back into the expression for \( \mathcal{E} \) (4.10) gives

\[ \mathcal{E} = \int_A \frac{D}{2} \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} - 2\mu \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} + (1 + \mu) N_x^2 \right) \]

\[ + N_x \left( \frac{\partial u}{\partial x} + w \frac{\partial^2 w}{\partial x^2} \right) + N_y \left( \frac{\partial v}{\partial y} + \frac{w}{R_y} + \frac{1}{2} \frac{\partial^2 w}{\partial y^2} \right) \]

\[ + N_{xy} (u + v_x + w_x w_y) + V(w) \, dx \, dy. \] (4.12)

Now we take the derivatives with respect to \( u \) and \( v \). These give us the body force in \( u \) and \( v \) direction, respectively. We get

\[ B_x = -\frac{\delta \mathcal{E}}{\delta u} = N_{x,x} + N_{x,y,y} \quad \text{and} \quad B_y = -\frac{\delta \mathcal{E}}{\delta v} = N_{y,y} + N_{y,x,x}. \]

From this we obtain two dynamic equations for the tangential variables as follows

\[ su_{tt} + \kappa' u_t = N_{x,x} + N_{x,y,y} \]

\[ sv_{tt} + \kappa' v_t = N_{y,y} + N_{y,x,x}. \] (4.13)

The expressions \( su_{tt} \) and \( sv_{tt} \) denote inertia terms whereas \( \kappa' u_t \) and \( \kappa' v_t \) model damping effects due to friction. Because movement in the developing embryo occurs slowly we assume that the friction terms are far more important than the inertia terms. Therefore we will neglect the inertia terms.

The final form of the equations simplifies greatly if we additionally assume that \( \kappa' u_t \) and \( \kappa' v_t \) vanish as well. This is justified considering that we are interested in
the instabilities of the \( w \) variable. In this case we treat the shell as if it is always in force balance tangentially; and we introduce a new function, \( F \), with the following properties:

\[
N_x = F_{yy} \quad N_y = F_{xx} \quad N_{xy} = -F_{xy}.
\]

(4.14)

Note that (4.13) is always satisfied in that case. The function \( F \) is commonly called Airy stress function. In terms of \( F \), we obtain the following form for \( \mathcal{E} \)

\[
\mathcal{E} = \int_A \frac{D}{2} \mathcal{L}_1 w - \frac{1}{2Eh} \mathcal{L}_2 F + F_{yy} \left( u_x + \frac{w}{R_x} + \frac{1}{2} w_x^2 \right) \\
+ F_{xx} \left( v_y + \frac{w}{R_y} + \frac{1}{2} w_y^2 \right) - F_{xy} (u_y + v_x + w_x w_y) + V(w) \, dx dy
\]

(4.15)

where

\[
\mathcal{L}_2 F = F_{xx}^2 + F_{yy}^2 - 2\mu F_{xx} F_{yy} + 2(1 + \mu) F_{xy}^2.
\]

Let us now find the derivative with respect to \( F \). Using the identities

\[
\frac{\delta \mathcal{L}_2 F}{\delta F} = 2 \nabla^4 F
\]

\[
\frac{1}{2} (w_x^2)_{yy} + \frac{1}{2} (w_y^2)_{xx} - (w_x w_y)_{xy} = w_{xx}^2 - w_{xx} w_{yy}
\]

\[
(u_x)_{yy} + (v_y)_{xx} - (u_y)_{xy} - (v_x)_{xy} = 0
\]

we find

\[
\frac{\delta \mathcal{E}}{\delta F} = -\frac{1}{Eh} \nabla^4 F + \frac{w_{yy}}{R_x} + \frac{w_{xx}}{R_y} + w_{xy}^2 - w_{xx} w_{yy}.
\]

Similar calculations are needed for the derivative with respect to \( w \). We use the identities

\[
\frac{\delta \mathcal{L}_1 w}{\delta w} = 2 \nabla^4 w
\]

\[
- (F_{yy} w_x)_x - (F_{xx} w_y)_y + (F_{xy} w_y)_x + (F_{yy} w_x)_y
\]

\[
= 2 F_{xy} w_{xy} - F_{yy} w_{xx} - F_{xx} w_{yy}
\]
to obtain

\[-B_x = \frac{\delta E}{\delta w} = D \nabla^4 w + \frac{F_{yy}}{R_x} + \frac{F_{xx}}{R_y} + 2F_{xy}w_{,xx} - F_{yy}w_{,xy} - F_{xx}w_{,yy} + V'(w).\]

Setting up the force balance equation yields (again neglecting the inertia term)

\[\kappa w_t + D \nabla^4 w + \frac{F_{yy}}{R_x} + \frac{F_{xx}}{R_y} - [F, w] + V'(w) = 0 \tag{4.16}\]

\[
\frac{1}{Eh} \nabla^4 F - \frac{w_{yy}}{R_x} - \frac{w_{xx}}{R_y} + \frac{1}{2}[w, w] = 0 \tag{4.17}
\]

where

\[[F, w] = F_{xx}w_{,yy} + F_{yy}w_{,xx} - 2F_{xy}w_{,xy}.\]

Note that the resulting equations are no longer dependent on \(u, v\) and the only \(\mu\) dependence is through \(D\). The only remaining dependent variables are \(F\) and \(w\). We will refer to these equations as the von Karman equations. They are relatively simple, however it is difficult to specify physically relevant boundary conditions. The term \([w, w]\) is often referred to as Gaussian curvature because it is the Gaussian curvature of a plate with deflection \(w\).

It is sometimes useful to use the displacement equations of shallow shell theory. These equations are easily obtained combining (4.9), (4.13), (4.14) and (4.16).

\[\kappa' u_t = N_{x,x} + N_{x,y,y}\]
\[\kappa' v_t = N_{y,y} + N_{x,x}\]
\[\kappa' w_t = -D \nabla^4 w - \frac{N_x}{R_x} - \frac{N_y}{R_y} + N_{x,x}w_{,xx} + N_{y,y}w_{,yy} + 2N_{xy}w_{,xy} - V'(w) \tag{4.18}\]

where

\[N_x = \frac{Eh}{1 - \mu^2} \left( u_{,x} + \frac{w}{R_x} + \frac{w_x^2}{2} + \mu \left( v_{,y} + \frac{w}{R_y} + \frac{w_y^2}{2} \right) \right)\]
\[N_y = \frac{Eh}{1 - \mu^2} \left( v_{,y} + \frac{w}{R_y} + \frac{w_y^2}{2} + \mu \left( u_{,x} + \frac{w}{R_x} + \frac{w_x^2}{2} \right) \right) \tag{4.19}\]
\[N_{xy} = \frac{Eh}{2(1 + \mu)} \left( u_{,x} + v_{,x} + w_{,x}w_{,y} \right).\]
We will assume in the following that \( V(w) \) is given by a power series around \( w = 0 \) where we retain the terms up to the fourth power as follows

\[
V(w) = pw + \frac{\gamma}{2} w^2 + \frac{\alpha}{3} w^3 + \frac{\beta}{4} w^4 .
\]

This should only be seen as a first attempt to model the mechanical response of the dermis and the intermediate layer. Parameter \( \gamma \) is the linear spring parameter that is, as we will shortly see, important for the buckling wavelength. Parameter \( \alpha \) measures the degree of asymmetry in the elastic response. We associate \( w > 0 \) to the upper epidermis layers and \( w < 0 \) to the dermis. Therefore, if \( \alpha > 0 \) there is more resistance from the intermediate layer than from the dermis, as it is, in fact, observed. Parameter \( \beta \) is a nonlinear spring constant, that is important for the saturation of the instability and the buckling amplitude.

### 4.3 Solutions and perturbed equations

To analyze the effect of stress and curvature on the buckling behavior, we follow directions outlined by REISSNER [91] and AXELRAD [7]. We omit the commas for denoting derivatives on \( u, v \) and \( w \) and \( F \), but retain them for \( N_x, N_y \) and \( N_{xy} \) where a confusion would be possible.

We are interested in solutions in a small patch of the shell where the stress field \((N_x = N_{x0}, N_y = N_{y0}, N_{xy} = N_{xy0}, F = F_0)\), the radii of curvature \((R_x, R_y)\) and the deflection \((w = w_0)\) are almost constant. In this case we have

\[
F_0 = \frac{1}{2} N_{x0} y^2 + \frac{1}{2} N_{y0} x^2 - N_{xy0} xy .
\]

Then we have from the third equation of (4.18)

\[
\frac{N_{x0}}{R_x} + \frac{N_{y0}}{R_y} + p + \gamma w + \alpha w^2 + \beta w^3 = 0 .
\]

This cubic obviously always has at least one solution \( w = w_0 \). There cannot be more than one solution because \( V'(w) \) is always increasing (the opposite case would be unphysical).
Further, (4.19) gives us

\[ u_{0y} = \frac{1}{Eh} (N_{x0} - \mu N_{y0}) \]
\[ v_{0y} = \frac{1}{Eh} (N_{y0} - \mu N_{x0}) \]
\[ u_{0y} + v_{0x} = \frac{2(1 + \mu)}{Eh} N_{xy0} . \]  

(4.20)

The concrete solutions of these equations depend on the boundary conditions. We now perturb the variables according to

\[ u = u_0 + \tilde{u} \quad N_x = N_{x0} + \tilde{N}_x \quad F = F_0 + f \]
\[ v = v_0 + \tilde{v} \quad N_y = N_{y0} + \tilde{N}_y \]
\[ w = w_0 + \tilde{w} \quad N_{xy} = N_{xy0} + \tilde{N}_{xy} . \]

For the potential \( V(w) \) we obtain that way

\[ V(w_0 + \tilde{w}) = V(w_0) + \tilde{\gamma} \tilde{w} + \alpha \tilde{w}^2 + \beta \tilde{w}^3 + \text{higher nonlinear terms} \]

We will neglect these nonlinear terms in the following. Performing the perturbation for the von Karman equations gives us

\[ \kappa \ddot{u} + D \nabla^4 \tilde{w} + \frac{1}{R_y} f_{xx} + \frac{1}{R_x} f_{yy} - [f, \tilde{w}] - N_{x0} \tilde{w}_{xx} - N_{y0} \tilde{w}_{yy} - 2N_{xy0} \tilde{w}_{xy} + \gamma \tilde{w} + \alpha \tilde{w}^2 + \beta \tilde{w}^3 = 0 \]
\[ \frac{1}{Eh} \nabla^4 f - \frac{1}{R_y} \tilde{w}_{xx} - \frac{1}{R_x} \tilde{w}_{yy} + \frac{1}{2} [\tilde{w}, \tilde{w}] = 0 . \]  

(4.21)

For the displacement equations we obtain

\[ \kappa' \ddot{u}_t = \tilde{N}_{x,x} + \tilde{N}_{xy,y} \]
\[ \kappa' \ddot{v}_t = \tilde{N}_{y,y} + \tilde{N}_{xy,x} \]
\[ \kappa \ddot{w}_t = -D \nabla^4 \tilde{w} - \frac{\tilde{N}_x}{R_x} - \frac{\tilde{N}_y}{R_y} \]
\[ + (N_{x0} + \tilde{N}_x) \tilde{w}_{xx} + (N_{y0} + \tilde{N}_y) \tilde{w}_{yy} + 2(N_{xy0} + \tilde{N}_{xy}) \tilde{w}_{xy} - V'(w) . \]  

(4.22)
In Section 4.5.1 we will show that the boundary conditions in \( u \) and \( v \) have a significant impact on the solution behavior, however we will not attempt to settle all the questions associated with this problem in this thesis. As for \( w \) we will not take boundary conditions into account because here we are interested in the solution behavior away from the boundaries in a small patch. It is known from other pattern forming systems that the boundary conditions usually only have an effect to the regions close to the boundary.

For simplicity we drop the \( \varrho \) indices and the bars in \( u, v, \bar{w}, \bar{\alpha}, \bar{\beta} \) and \( \bar{\gamma} \). We will now go ahead and analyze (4.21) and the corresponding displacement equations (4.22).

### 4.4 Linear stability analysis

The displacement equations for \( u \) and \( v \) are elliptic in the spatial coordinates and parabolic in time. Therefore, there are no instabilities in \( u \) and \( v \). For stability we will look at the perturbed von Karman equations (4.21), an analysis of the displacement equations leads to the same result.

We neglect all nonlinear terms in (4.21) and combine the equations as follows

\[
\nabla^4(\kappa \omega_t + D \nabla^4 w - N_x w_{yy} - N_y w_{xx} - 2N_{xy} w_{xy} + \gamma w) = -E\hbar (\nabla_C^2)^2 w \quad (4.23)
\]

where

\[
\nabla_C^2 = \frac{1}{R_y} \frac{\partial^2}{\partial x^2} + \frac{1}{R_x} \frac{\partial^2}{\partial y^2}.
\]
Take

\[ w = e^{\omega t + ik \cdot x} \]

where \( k = (l, m) \). We obtain the dispersion relation

\[
\kappa \omega = -Dk^4 - N_x l^2 - N_y m^2 - 2N_{xy} ml - \gamma \frac{Eh}{k^4} \left( \frac{l^2}{R_y} + \frac{m^2}{R_x} \right)^2.
\]

with \( k = |k| \). It is advantageous to write the dispersion relation in polar coordinates, viz. \( l = k \cos \theta \) and \( m = k \sin \theta \).

\[
\kappa \omega = -Dk^4 - k^2 (\hat{N}_x \cos^2(\theta - \psi) + \hat{N}_y \sin^2(\theta - \psi)) - \gamma \frac{Eh}{R_y} \left( \frac{\cos^2 \theta}{R_y} + \frac{\sin^2 \theta}{R_x} \right)^2. \tag{4.24}
\]

Here \( \hat{N}_x \) and \( \hat{N}_y \) denote the principal stresses and \( \psi \) the angle between the direction of stress \( \hat{N}_x \) and the direction of the curvature \( 1/R_x \). Note that the linear spring constant \( \gamma \) and the curvature terms play equivalent roles in the dispersion relation. Anticipating that buckling patterns only occur when at least one of the stresses is negative (or only very slightly positive), we introduce a control parameter \( S = -\hat{N}_x \).

For simplicity, we assume that \( \hat{N}_x \leq \hat{N}_y \) and \( |1/R_y| \geq |1/R_x| \). To make the following calculations more transparent, we set \( \xi = \hat{N}_y/\hat{N}_x \) and \( \zeta = R_y/R_x \). In this notation, we have the dispersion relation

\[
\kappa \omega = -Dk^4 + Sk^2 (\cos^2(\theta - \psi) + \xi \sin^2(\theta - \psi)) - \gamma \frac{Eh}{R_y^2} \left( \cos^2 \theta + \zeta \sin^2 \theta \right)^2. \tag{4.25}
\]

where \( \xi \leq 1 \) and \( |\zeta| \leq 1 \). We choose \( S \) as the bifurcation parameter on the system and look for the smallest value of \( S \) for which some mode exhibits growth. Further, we are interested in the wavelength \( \lambda = 2\pi/k \) and the direction \( \theta \) of the amplified mode. The general solution of the problem is difficult and leads to trigonometric expressions that are a nightmare. Instead, let us first investigate two special cases.

1. Consider the case \( \hat{N}_x = \hat{N}_y \) or \( \xi = 1 \). Then we have

\[
\kappa \omega = -Dk^4 + Sk^2 - \gamma \frac{Eh}{R_y^2} \left( \cos^2 \theta + \zeta \sin^2 \theta \right)^2.
\]
FIGURE 4.8. The $y$-axis is the direction of greatest curvature, the $x$-axis is the direction of smallest curvature. Further, the $\tilde{N}_z$-axis is the one of largest stress and the $\tilde{N}_y$-axis the one of smallest stress.

Fixing all variables except $\theta$, this expression has a maximum for $\sin \theta = 0$. Therefore the modes first positively amplified are the ones in the direction of largest curvature. Accordingly, the ridges follow the lines of smallest curvature.

2. Similarly we can look what happens if $1/R_x = 1/R_y$ or $\zeta = 1$. Then

$$\kappa \omega = -Dk^4 + S k^2 (\cos^2 (\theta - \psi) + \xi \sin^2 (\theta - \psi)) - \gamma - \frac{Eh}{R^2_y}.$$ 

Again, fixing all variables except $\theta$ reveals that the maximum is achieved for $\cos(\theta - \psi) = 0$. In other words, the buckling mode is the direction of largest compressive stress. Accordingly, the ridges follow the lines of smallest compressive stress.

In view of these observations, we can consider curvature and stress as competing factors for the buckling direction. Even without solving the problem rigorously, and in spite of the lack of important mechanical parameters (such as $E$, $\mu$), we are able to
estimate the coefficients in front of the stress and curvature terms in the dispersion relation. We will show that the coefficient of the stress term is much larger than the one in front of the curvature terms.

For this purpose we will determine lower bounds on the critical stress and the wavelength. To obtain these bounds we use measurements of wavelength, basal layer thickness and radius of curvature provided by histological photographs. We use numbers provided by BONNEVIE's pictures of Embryo No. 11, digit II. ([19], Figures 6a and 16a). We find

\[ R = 780 \mu m, \quad \lambda = 36 \mu m \quad \text{and} \quad h = 8.0 \mu m. \]

Now we find an expression for the critical wave number by setting \( \omega = 0 \) in the dispersion relation (4.25) and solving for the stress parameter \( S \)

\[ S(\cos^2(\theta - \psi) + \xi \sin^2(\theta - \psi)) = Dk^2 + \frac{\gamma}{k^2R_y^2} \left( \cos^2 \theta + \zeta \sin^2 \theta \right)^2. \]

Minimizing for \( k \) yields

\[ k_c^4 = \left( \frac{2\pi}{\lambda} \right)^4 = \frac{Eh}{DR_y^2} (\cos^2 \theta + \zeta \sin^2 \theta)^2 + \frac{\gamma}{D}, \]

where we utilized \( D = Eh^3/(12(1 - \mu^2)) \). The first term is associated with the effects of curvature on the buckling wavelength, the second one measures the contribution of the spring foundation. Even though we do not know \( \theta \) we can overestimate the curvature contribution by neglecting the trigonometric terms. We assume \( \mu = 0.45 \) (see Section 4.6 for a discussion of this choice). Now we can estimate the curvature contribution to the wavelength and then determine the foundation contribution by solving for \( \gamma/D \). Using BONNEVIE's numbers we estimate the wave number \( k_c \) that would be obtained if there were no elastic foundation and compare it to the wave
number $k_+$ that is actually observed.

$$k_*^4 = \frac{12(1 - \mu^2)}{h^2 R_y^2} = 2.5 \cdot 10^{-7} \text{ (\mu m)}^{-4}$$

$$k_+^4 = \left(\frac{2\pi}{\lambda}\right)^4 = 9.3 \cdot 10^{-4} \text{ (\mu m)}^{-4}.$$  

We see that the curvature terms are not nearly large enough to enforce the correct wavelength. If there were no elastic foundation and the curvature terms would solely determine the ridge wavelength we would have

$$\lambda_+ = 280 \mu m$$

which is almost 8 times larger than the actually observed wavelength. We conclude that the wavelength is enforced by the elastic foundation and have

$$k_+^4 = \left(\frac{2\pi}{\lambda}\right)^4 = \frac{\gamma}{D} = 9.3 \cdot 10^{-4} \text{ (\mu m)}^{-4}.$$  

Now we are able to estimate the magnitude of the curvature and stress terms in the dispersion relation (4.25). We have

$$S(\cos^2(\theta - \psi) + \xi \sin^2(\theta - \psi)) k^2 = Dh^4 + \gamma + \frac{E_h}{R_y} \left(\cos^2 \theta + \zeta \sin^2 \theta\right)^2$$

$$> Dh^4$$

$$= D \cdot 9.3 \cdot 10^{-4} \text{ (\mu m)}^{-4}.$$  

On the other hand,

$$\frac{E_h}{R_y^2} \left(\cos^2 \theta + \zeta \sin^2 \theta\right)^2 < \frac{E_h}{R_y^2}$$

$$= D \cdot 2.5 \cdot 10^{-7} \text{ (\mu m)}^{-4}.$$  

These estimations show that the curvature terms in the dispersion relation are much smaller (by a factor of about 3,500) than the stress terms. So, we conclude that even small differences in the principal stresses are sufficient to determine the wave direction.
4.5 Nonlinear analysis

4.5.1 Analysis of the displacement equations

We are interested in the behavior of the shell once buckling has occurred. A technique suitable for the analysis of the postbuckling state is weakly nonlinear analysis that describes the state of the system close to the onset of buckling. Weakly nonlinear analysis is performed by deriving equations for the amplitude of the unstable variable (if the superposition of cosine waves forms the pattern) that give insight into the pattern type and the pattern amplitude.

In this section we focus on the idea of weakly nonlinear analysis (see also [59, 60]) and discuss certain boundary effects in the displacement equations that are difficult to capture using the von Karman equations. We will not yet consider the most general case, for instance, we neglect space dependence of the amplitude here. In the following section 4.5.2 we consider the amplitude equations for a wave triad using the less complex von Karman equations. In 4.5.3 we present a more general calculation that takes a slowly varying amplitude into account.

We neglect the curvature terms and, to simplify the discussion, set \( N_y = 0 \) and \( N_{xy} = 0 \). Hence, we start with the following equations:

\[
\begin{align*}
\kappa' u_t &= \overline{N}_{x,x} + \overline{N}_{xy,y} \\
\kappa' v_t &= \overline{N}_{y,y} + \overline{N}_{xy,x} \\
\kappa w_t &= -DN^2w + (N_x + \overline{N}_x)w_{xx} + \overline{N}_yw_{yy} + 2\overline{N}_{xy}w_{xy} \\
&= -\gamma w - \alpha w^2 - \beta w^3
\end{align*}
\]

where

\[
\begin{align*}
\overline{N}_x &= \frac{Eh}{1-\mu^2} \left( u_x + \frac{w_x^2}{2} + \mu \left( v_y + \frac{w_y^2}{2} \right) \right) \\
\overline{N}_y &= \frac{Eh}{1-\mu^2} \left( v_y + \frac{w_y^2}{2} + \mu \left( u_x + \frac{w_x^2}{2} \right) \right) \\
\overline{N}_{xy} &= \frac{Eh}{2(1+\mu)} \left( u_y + v_x + w_xw_y \right).
\end{align*}
\]
Assume that we are still close to the onset of the instability, that is we have

\[ N_x = -N_c(1 + \epsilon^2 \chi') \quad \text{where} \quad N_c = 2\sqrt{\gamma D} \]

(note again that the stress has to be negative for an instability). Here \( \epsilon \) is a small parameter and \( \chi' \) a free parameter. Let us assume the postbuckling solution is approximately a cosine wave in the following form

\[ w = \epsilon w_0 + \epsilon^2 w_1 + \epsilon^3 w_2^2 + \ldots \]

where

\[ w_0 = A e^{ik_c x} + \text{c.c.} \quad . \]

Here \( k_c = \sqrt{\gamma D} \) is the critical wave number. We introduce a slow time scale \( T = \epsilon^2 t \) and assume that \( \vec{N}_x \) is of order \( O(\epsilon^2) \).

Now we insert these expressions into the force balance equation for \( w \). At order \( O(\epsilon) \) we obtain

\[ D \nabla^4 w_0 + N_c w_{0xx} + \gamma w_0 = 0 \quad . \]

This equation is satisfied due to the choice of \( w_0 \). At the next order \( O(\epsilon^2) \) we get

\[ D \nabla^4 w_1 + N_c w_{1xx} + \gamma w_1 = 0 \quad . \]

We choose \( w_1 = 0 \) as a solution of this equation. Taking into account that

\[ (A e^{ik_c x} + \text{c.c.})^3 = 3A^2 A^* e^{ik_c x} + A^3 e^{3ik_c x} + \text{c.c.} \]

we get at order \( O(\epsilon^3) \)

\[ D \nabla^4 w_2 + N_c w_{2xx} + \gamma w_2 = (-\kappa A_T + 2N_c k_c^2 A - 3\beta A^2 A^* - k_c^2 \vec{N}_x A)e^{ik_c x} \]

+ terms independent from \( e^{ik_c x} \).

Because \( e^{ik_c x} \) solves the homogeneous equation this equation only has bounded solutions if all terms involving \( e^{ik_c x} \) (so-called secular terms) vanish. This requirement gives rise to the amplitude equation

\[ \kappa A_T = 2N_c k_c^2 \chi' A - 3\beta A^2 A^* - k_c^2 \vec{N}_x A \]
unbuckled shell


buckled shell

Case 1.) $\bar{N}_x = 0$ and $u_x < 0$

Case 2.) $\bar{N}_x < 0$ and $u_x = 0$

Figure 4.9. Two different buckling scenarios. In the first case the stress on the system is maintained, the shell is squashed together. If the shell cannot move at the ends, the stress on the system decreases and the amplitude is reduced compared to the first case.

where

$$\bar{N}_x = \frac{Eh}{1-\mu^2} (u_x + AA^* f_x^2 + \mu x) .$$

The stationary solutions of the amplitude equations depend on the boundary conditions. For simplicity assume that the equations are given on a box with corners $(0,0)$, $(0,L)$, $(M,0)$ and $(M,L)$. Further assume that $v(x,0) = 0$ and $v(x,L) = 0$ implying $v = 0$. Now there are two important cases

1.) $\bar{N}_x(0,y) = 0$ and $\bar{N}_x(M,y) = 0$

This boundary condition means that there is no change in boundary stress after buckling occurs. It implies $\bar{N}_z = 0$ everywhere. Then the amplitude equation has the following form

$$\kappa A_T = 2Nc_k^2 \chi' A - 3\beta A^2 A^* .$$
This equation has the solutions $A = 0$ (which is unstable for $\chi' > 0$) and

$$|A|^2 = \frac{2N_ck_c^2\chi'}{3\beta}.$$  

Notice that in this case the instability is only saturated by the cubic term in the elastic foundation. Further we can determine $u_\pi$ to be

$$u_\pi = -AA^*k_c^2.$$  

Therefore compression of the shell takes place here.

2.) $u(0, y) = 0$ and $u(M, y) = 0$

This boundary condition means that the shell cannot move during buckling at the boundary. It implies $u = 0$ everywhere. In this case we have

$$\bar{N}_x = -\frac{Ehk_c^2AA^*}{1 - \mu^2}.$$  

It is not surprising that the stress on the system in the direction of the instability decreases because buckling with fixed ends allows material expansion. The amplitude equation becomes

$$\kappa A_T = 2N_ck_c^2\chi' A - \left(3\beta + \frac{Ehk_c^2}{1 - \mu^2}\right) A^2 A^*.$$  

Compared to the amplitude equation the nonlinear resistance term is increased. For the amplitude we obtain

$$|A|^2 = \frac{2N_ck_c^2\chi'}{3\beta + \frac{Ehk_c^2}{1 - \mu^2}}.$$  

Even if $\beta = 0$ the instability saturates. See also Figure 4.9 for a visualization.

Unfortunately it is difficult to decide what case is actually important in the buckling process we are interested in. If the buckling process takes place everywhere almost simultaneously case 2.) is more realistic. It is indeed more reasonable that
the boundary does not invade the epidermis and the stress acting on the system becomes smaller during the buckling process.

However, if buckling only takes place in a small patch it seems that 1.) is more appropriate. The unbuckled epidermis is still able to "push in" toward the buckled epidermis and the stress decreases only slightly. To keep matters as simple as possible we will choose case 1.) in the discussion of the next two sections. Therefore whenever we have to choose a change in the stress field during buckling we will define it in a way that it vanishes for straight parallel rolls. Note that case 2.) is still included in our calculations because it can be obtained by a change in the nonlinear spring constant $\beta$.

Although the displacement equations are more intuitive than the von Karman equations and although it is easier to include relevant boundary conditions we will use the von Karman equations in the following analysis because they are less complex and only involve two independent variables.

4.5.2 Analysis of the von Karman equations: weakly anisotropic case

In humans, ridge patterns are almost always observed. Among mammals and marsupials ridge patterns dominate as well and examples of dot patterns (so-called hexagons) are rare. On the other hand, it is well-known theoretically [59, 60] and empirically [23] that hexagons are a possible buckling pattern for elastic shells. In this section we investigate what circumstances cause ridges or hexagons and why hexagons almost never form on human skin.

For simplicity, we neglect the curvature terms and assume a constant stress field in a small region. Now we use a coordinate system that has the principal stresses as coordinate axes. We are interested in the bifurcations of the system and the parameters that govern the transition from hexagons to stripes. The equations of
interest have the following form.

\[ \kappa w_t + D \nabla^4 w + S w_{xx} + U w_{yy} + \gamma w + \alpha w^2 + \beta w^3 - [f, w] = 0 \]
\[ \frac{1}{Eh} \nabla^4 f + \frac{1}{2} [w, w] = 0 \]

where \( S = -\bar{N}_x \) and \( U = -\bar{N}_y \). From the results of the previous section, we know the critical stress and the critical wave number,

\[ N_c = 2\sqrt{\gamma D} \quad \text{and} \quad k_c^4 = \frac{\gamma}{D}. \tag{4.26} \]

Note that, in view of the results of the last section, \( f \) tells us about the change in the stress after buckling has occurred. Even if \([w, w] = 0\) it does not necessarily follow that \( f = 0 \) because of the boundary conditions. We will, however, draw exactly this conclusion in the following with the understanding that boundary forces causing the instability are still maintained after buckling. Note also, that \([w, w] \neq 0\) implies \( f \neq 0 \) regardless of the boundary conditions imposed.

To investigate the behavior near the onset of instability we now introduce a small parameter \( \epsilon \) such that

\[ S = N_c(1 + \epsilon^2 \chi') \]
\[ U = N_c(1 + \epsilon^2 \tau') \]

This means we assume that both principal stresses are close to the critical stress. We will take \( S \geq U \) implying \( \chi' \geq \tau' \). We introduce a slowly varying time scale, \( T = \epsilon^2 t \).

We also assume that the quadratic nonlinearity in the spring response is small and set \( \alpha = \epsilon \alpha' \).

It is well-known that for certain parameter values above critical stress both ridges (represented by a single cosine wave) and hexagons (represented by a sum of three cosine waves) are stable solutions of the equations. Therefore, we are interested in solutions that are approximately given by the sum of three cosine waves. We obtain hexagons if their amplitudes are equal, and we we obtain ridges if two of the three amplitudes vanish.
For this goal let $k_j$, $j = 1, 2, 3$ be three wave vectors with $|k_j| = |k_c|$ and

$$\sum_{j=1}^{3} k_j = 0 \quad (4.27)$$

which is possible if the three vectors are separated by angles of $120^\circ$ (see Figure 4.10). Now assume that $w$ and $f$ can be represented by an asymptotic series in $\epsilon$ as follows

$$w(x, T) = \epsilon w_0 + \epsilon^2 w_1 + \epsilon^3 w_2 + \cdots$$
$$f(x, T) = \epsilon f_0 + \epsilon^2 f_1 + \epsilon^3 f_2 + \cdots$$

where

$$w_0 = \sum_{j=1}^{3} A_j e^{ik_j \cdot x} + \text{c.c.} .$$

We assume that the $A_j$ depend on the slow time scale $T$ as follows $A_j = A_j(T = \epsilon^2 t)$. It is easily seen that $f_0 = 0$.

Because $S \geq U$ the wave vector most amplified will be $(1, 0)$. Therefore we include it as one of the three wave vectors. This defines the other two wave vectors via (4.27) and the requirement that the vectors are separated by $120^\circ$. Therefore, we have

$$k_1 = (1, 0), \quad k_2 = (-1/2, \sqrt{3}/2) \quad \text{and} \quad k_3 = (-1/2, -\sqrt{3}/2) .$$

These expressions are substituted back into the main equations and we find $w_j$ and $f_j$ by matching terms of the same $\epsilon$ power. At order $O(\epsilon)$ we obtain

$$D \nabla^4 w_0 + N_c w_{0xx} + N_c w_{0yy} + \gamma w_0 = 0 .$$

This equation is satisfied due to the choice of $w_0$. At the next order $O(\epsilon^2)$ we get

$$D \nabla^4 w_1 + N_c w_{1xx} + N_c w_{1yy} + \gamma w_1 = 0 .$$

We choose $w_1 = 0$ as a solution of this equation. At order $O(\epsilon^3)$ we get

$$D \nabla^4 w_2 + N_c w_{2xx} + N_c w_{2yy} + \gamma w_2 = -\kappa w_{0t} - N_c (\chi' w_{0xx} + \tau' w_{0yy}) - \alpha' w_0^2$$
$$- \beta w_0^3 + [w_0, f_1] . \quad (4.28)$$
FIGURE 4.10. A triad of wave vectors. The vectors all have the magnitude of the critical wave number and are spaced 120° away from each other. The sum of three cosine waves with these wave vectors results in hexagons. Note that we have $k_1 + k_2 + k_3 = 0$. This property is crucial for quadratic interactions that stabilize hexagon patterns.
The three waves, $e^{ik_1 \cdot x}$, are solutions of the homogeneous equation. Therefore, in order to obtain a solution for (4.28) these secular terms must cancel. Requiring all these terms to vanish gives the amplitude equations.

Below are the contributions of each right hand side term to the secular terms and, consequently, to the amplitude equations:

\[
\kappa w_{0t} \rightarrow \kappa \sum_{j=1}^{3} \frac{\partial}{\partial T} A_j e^{i k_j \cdot x} + \text{c.c.}
\]

\[
N_c \chi' w_{0xx} \rightarrow -N_c |k_c|^2 \chi' \left( A_1 e^{i k_1 \cdot x} + \frac{1}{4} A_2 e^{i k_2 \cdot x} + \frac{1}{4} A_3 e^{i k_3 \cdot x} \right) + \text{c.c.}
\]

\[
N_c \tau' w_{0yy} \rightarrow -N_c |k_c|^2 \tau' \left( \frac{3}{4} A_2 e^{i k_2 \cdot x} + \frac{3}{4} A_3 e^{i k_3 \cdot x} \right) + \text{c.c.}
\]

\[
\alpha' w_0^2 \rightarrow 2\alpha' \left( A_2^* A_3 e^{i k_1 \cdot x} + A_1^* A_3^* e^{i k_2 \cdot x} + A_1^* A_2^* e^{i k_3 \cdot x} \right) + \text{c.c.}
\]

\[
\beta w_0^3 \rightarrow 3\beta \sum_{j=1}^{3} A_j^2 A_j e^{i k_j \cdot x} + 6\beta \sum_{j=1}^{3} A_i A_j^* A_j e^{i k_j \cdot x} + \text{c.c.}
\]

\[
[f_1, w_0] \rightarrow -\frac{9}{8} E h k_c^4 \sum_{j=1}^{3} \sum_{i \neq j} A_i A_i^* A_j e^{i k_j \cdot x} + \text{c.c.}
\]

Note that we used (4.27) for finding the contribution of $\alpha' w_0^2$. For the expression $[f_1, w_0]$ we used

\[
f_1 = - \frac{E h}{k_c^2} \cdot \frac{1}{2} [w_0, w_0]
\]

\[
= - \frac{E h}{k_c^2} \cdot \left( \frac{\sqrt{3}}{2} k_c^2 \right)^2 \left( A_2^* A_3 e^{i k_1 \cdot x} + A_1^* A_3^* e^{i k_2 \cdot x} + A_1^* A_2^* e^{i k_3 \cdot x}
+ A_1 A_3 e^{i (k_1 - k_2) \cdot x} + A_1 A_3 e^{i (k_1 - k_3) \cdot x} + A_2 A_3 e^{i (k_2 - k_3) \cdot x} + \text{c.c.} \right)
\]

Using (4.26), we obtain the following amplitude equations:

\[
\frac{dA_1}{dT} = \hat{\eta} A_1 - 2\alpha' A_2^* A_3^* - A_1 \left( \hat{\delta} A_1 A_1^* + \hat{\phi} \sum_{i \neq 1} A_i A_i^* \right)
\]

\[
\frac{dA_2}{dT} = \hat{\sigma} A_2 - 2\alpha' A_1^* A_3^* - A_2 \left( \hat{\delta} A_2 A_2^* + \hat{\phi} \sum_{i \neq 2} A_i A_i^* \right)
\]

\[
\frac{dA_3}{dT} = \hat{\sigma} A_3 - 2\alpha' A_1^* A_2^* - A_3 \left( \hat{\delta} A_3 A_3^* + \hat{\phi} \sum_{i \neq 3} A_i A_i^* \right)
\]
where

\[ \hat{\eta} = 2\gamma\chi', \quad \hat{\sigma} = \frac{\gamma}{2} (\chi' + 3\tau'), \quad \hat{\delta} = 3\beta, \quad \hat{\phi} = 6\beta + \frac{27\gamma(1 - \mu^2)}{2h^2}. \]

Now we scale the equations as follows

\[ A_i = -\frac{2\alpha'\hat{\delta}}{\hat{\delta}} A_i, \quad T = \frac{(2\alpha')^2}{\hat{\delta}} T \]

and obtain

\[
\begin{align*}
\frac{dA_1}{dT} &= \eta A_1 + A_2 A_3 - A_1 \left( A_1 A_1 + \phi \sum_{i \neq 1} A_i A_i \right) \\
\frac{dA_2}{dT} &= \sigma A_2 + A_1^* A_3 - A_2 \left( A_2 A_2 + \phi \sum_{i \neq 2} A_i A_i \right) \\
\frac{dA_3}{dT} &= \sigma A_3 + A_1^* A_2^* - A_3 \left( A_3 A_3 + \phi \sum_{i \neq 3} A_i A_i \right)
\end{align*}
\]

(4.29)

where

\[ \sigma = \frac{\hat{\sigma} \hat{\delta}}{(2\alpha')^2}, \quad \eta = \frac{\hat{\sigma} \hat{\delta}}{(2\alpha')^2}, \quad \phi = \frac{\hat{\phi}}{\hat{\delta}}, \]

again we have dropped the bars. Note that \( \sigma \leq \eta \) and, in the absence of the term \([f, w]\), we would have \( \phi = 2 \).

Let us now investigate the stability of certain solutions of these equations:

**Uniform Solution**

\[ A_1 = A_2 = A_3 = 0 \]

The solution is stable for \( \max(\sigma, \eta) < 0 \), otherwise it is unstable.

**Roll Solution Type I**

\[ A_1 = r \neq 0, \quad A_2 = A_3 = 0 \]
where $r = \pm \sqrt{\eta}$. We perturb the stationary solutions according to

$$A_1 = r + u, \quad A_2 = v \quad \text{and} \quad A_3 = w$$

where $u$, $v$ and $w$ are small numbers (a confusion with the displacements that are also denoted $u$, $v$ and $w$ is unlikely in this section). The linearized equations are

$$u_T = -2\eta u$$
$$v_T = \sigma v + rw^* - \phi r^2 v$$
$$w_T = \sigma w + rv^* - \phi r^2 w.$$  

The first equation decouples from the other two and one easily concludes that the solution $u = 0$ is always stable. We get the real equations by adding the complex conjugate. This gives us

$$(v + v^*)_T = \sigma(v + v^*) + r(w + w^*) - \phi r^2(v + v^*)$$
$$(w + w^*)_T = \sigma(w + w^*) + r(v + v^*) - \phi r^2(w + w^*).$$

The system has decaying solutions if the matrix

$$A = \begin{bmatrix} \sigma - \phi r^2 & r \\ r & \sigma - \phi r^2 \end{bmatrix}$$

has negative eigenvalues. The eigenvalues of the matrix are

$$\lambda_{1,2} = \sigma - \phi r^2 \pm r.$$  

They will be negative if and only if

$$\sigma < \phi \eta - \sqrt{\eta}.$$  

Considering the imaginary equations for $v - v^*$ and $w - w^*$ gives the same criterion.

**Roll Solution Type II**

$$A_2 = r \neq 0, \quad A_1 = A_3 = 0$$
where \( r = \pm \sqrt{\sigma} \). It is evident that the case \( A_3 \neq 0, A_1 = A_2 = 0 \) is absolutely analogous. We perturb the stationary solutions according to

\[
A_1 = u, \quad A_2 = r + v, \quad A_3 = w
\]

and linearize the equations. This gives

\[
\begin{align*}
t &= u + ru^* - \phi r^2 u \\
\phi &= -20v \phi \\
v^* &= \sigma w + ru^* - \phi r^2 w.
\end{align*}
\]

Again we consider equations for \((u + u^*)\) and \((w + w^*)\),

\[
\begin{align*}
(u + u^*)_t &= \eta(u + u^*) + r(w + w^*) - \phi r^2(u + u^*) \\
(w + w^*)_t &= \sigma(w + w^*) + r(v + v^*) - \phi r^2(w + w^*)
\end{align*}
\]

and find the eigenvalues of

\[
A = \begin{bmatrix}
\eta - \phi r^2 & \sigma \\
r & \sigma - \phi r^2
\end{bmatrix}
\]

which are

\[
\lambda_{1,2} = \frac{\sigma + \eta}{2} - \phi \sigma \pm \sqrt{\sigma + \left( \frac{\sigma - \eta}{2} \right)^2}.
\]

A necessary condition for the eigenvalues to be negative is

\[
\frac{\sigma + \eta}{2} - \phi \sigma < 0,
\]

or

\[
\sigma > \frac{\eta}{2\phi - 1}. \tag{4.30}
\]

If this condition is fulfilled we can set the larger eigenvalue to zero and solve for \( \sigma \).

We obtain

\[
\sigma > \frac{1 + \eta(\phi - 1)}{\phi^2 - \phi}.
\]

It is easy to see that this condition is more restrictive than the one given in (4.30).

As above, the equations for \((u - u^*)\) and \((w - w^*)\) give the same stability criterion.
Hexagons

\[ A_1 = e^{i\theta_1} \neq 0, A_2 = e^{i\theta_2} \neq 0, A_3 = e^{i\theta_3} \neq 0 \]

The amplitude equations require

\[ \theta_1 + \theta_2 + \theta_3 = 0 \quad \text{or} \quad \pi. \]

However, we will only look at the case when the sum is zero, because the other case is already included if we set \( s \to -s \) and \( r \to -r \). Putting the above expressions for the three amplitudes into the amplitude equations (4.29) gives us

\[ 0 = \eta s + r^2 - s^3 - 2\phi sr^2 \]
\[ 0 = \sigma r + rs - (1 + \phi)r^3 - \phi rs^2. \]  

(4.31)

From the second equation we obtain

\[ r^2 = \frac{\sigma + s - \phi s^2}{1 + \phi}. \]  

(4.32)

If we substitute this into the first equation we find a cubic equation in \( s \) as follows

\[ (2\phi^2 - \phi - 1)s^3 - 3\phi s^2 + (1 + (1 + \phi)\eta - 2\phi \sigma)s + \sigma = 0. \]  

(4.33)

Given \( \sigma, \eta, \phi \) we can find \( s \) using (4.33) and then \( r \) using (4.32). As a complication we can have one or three solutions for \( s \) from equation (4.33). We will see, however, that hexagons are always unstable if the cubic has only one solution. If it has three solutions the middle solution is the only one that may be stable.

As usual, we perturb the solution according to

\[ A_1 = (s + u)e^{i\theta_1}, \quad A_2 = (r + v)e^{i\theta_2}, \quad A_3 = (r + w)e^{i\theta_3}. \]

Linearizing and using (4.31) yields

\[ u_T = \frac{-r^2}{s}u + ru^* + rw^* - s^2(u + u^*) - \phi sr(v + v^*) - \phi sr(w + w^*) \]
\[ v_T = -sv + sw^* + ru^* - r^2(v + v^*) - \phi r^2(w + w^*) + \phi rs(u + u^*) \]
\[ w_T = -sw + sv^* + ru^* - r^2(w + w^*) - \phi r^2(v + v^*) + \phi rs(u + u^*). \]
Subtracting the complex conjugate yields
\[
(u - u^*)_T = -\frac{r^2}{s}(u - u^*) - r(v - v^*) - r(w - w^*)
\]
\[
(v - v^*)_T = -s(v - v^*) - s(w - w^*) - r(u - u^*)
\]
\[
(w - w^*)_T = -s(w - w^*) - s(v - v^*) - r(u - u^*)
\]
The eigenvalues of the matrix
\[
A = \begin{bmatrix}
-\frac{r^2}{s} & -r & -r \\
-r & -s & -s \\
-r & -s & -s
\end{bmatrix}
\]
are
\[
\lambda_{1,2} = 0 \quad \text{and} \quad \lambda_3 = -\frac{2s^2 + r^2}{s}
\]
This way we obtain $s > 0$ as a necessary criterion for stability. Let us now investigate
the equations for $(u + u^*)$, $(v + v^*)$, $(w + w^*)$
\[
(u + u^*)_T = (-\frac{r^2}{s} - 2s^2)(u + u^*) + (r - 2\phi sr)(v + v^*) + (r - 2\phi sr)(w + w^*)
\]
\[
(v + v^*)_T = (-s - 2r^2)(v + v^*) + (s - 2\phi r^2)(w + w^*) + (r - 2\phi sr)(u + u^*)
\]
\[
(w + w^*)_T = (-s - 2r^2)(w + w^*) + (s - 2\phi r^2)(v + v^*) + (r - 2\phi rs)(u + u^*)
\]
We consider the matrix
\[
B = \begin{bmatrix}
-\frac{r^2}{s} - 2s^2 & r - 2\phi sr & r - 2\phi sr \\
-r - 2\phi sr & -s - 2r^2 & s - 2\phi r^2 \\
-r - 2\phi sr & s - 2\phi r^2 & -s - 2r^2
\end{bmatrix}
\]
Its characteristic polynomial is
\[
c(B, \lambda) = -\frac{1}{s}(\lambda + 2s - 2(\phi - 1)r^2)(\lambda^2 + p\lambda + q),
\]
where
\[
p = 2s^3 + 2(\phi + 1)r^2s + r^2 \quad \text{and} \quad q = r^2(4(1 + \phi - 2\phi^2)s^3 + 8\phi s^2 - 2s + 2(1 + \phi)r^2).
\]
Because $p > 0$ (at least if $s > 0$), the eigenvalues defined by the quadratic will both
be negative or have negative real part if and only if $q > 0$. Dividing by $r^2$, we obtain
the following condition for stability,
\[
2(1 + \phi)r^2 > 4(2\phi^2 - \phi - 1)s^3 - 8\phi s^2 + 2s.
\] (4.34)
Requiring the remaining eigenvalue to be negative gives another condition,

\[(\phi - 1)r^2 < s.\]  \hspace{1cm} (4.35)

These inequalities immediately provide the area in the \(rs\)-plane for which hexagons are stable. However, it is much more informative to visualize the results in the \(\sigma\eta\)-plane. This is, however, a difficult task because \(s\) depends on \(\sigma\) and \(\eta\) via the solution of a cubic equation. Therefore, the problem has to be solved numerically. It is advantageous to rewrite the stability conditions (4.34), (4.35) in terms of \(\sigma\) and \(s\) as follows

\[\sigma > 2(2\phi^2 - \phi - 1)s^3 - 3\phi s^2\]
\[\sigma < \frac{2}{\phi - 1}s + \phi s^2.\]  \hspace{1cm} (4.36)

Here we have used (4.32). The regions of stability for the solution types are given in Figures 4.11, 4.12 and 4.13 for \(\phi = 2\). Even if \(\eta\) is only slightly larger than \(\sigma\) hexagons quickly become unstable and, hence, are not observed. If we increase \(\phi\) the stability domain for hexagons shrinks further.

Equations (4.36) are also useful to better understand the bifurcation type at the onset of instability. Let us introduce the variable \(\iota = \sigma - \eta\). Now solve equation (4.33) for \(\sigma\) and we get

\[\sigma = f_1(s) = \frac{-(2\phi^2 - \phi - 1)s^3 + 3\phi s^2 - (1 + (1 + \phi)\iota)s}{s(1 - \phi) + 1} = (2\phi + 1)s^2 - s + \frac{(\phi + 1)\iota s}{s(\phi - 1) - 1}.

Let us denote

\[g(s) = (2\phi + 1)s^2 - s\]

and further introduce the functions

\[f_2(s) = 2(2\phi^2 - \phi - 1)s^3 - 3\phi s^2\]
\[f_3(s) = \frac{2}{\phi - 1}s + \phi s^2.\]
For given $\phi$ and $\iota$ the variables $\sigma$ and $s$ have to satisfy the relationship $\sigma = f_1(s)$. The system is stable if $\sigma > f_2(s)$ and $\sigma < f_3(s)$. It is easy to see that $f_1$ is undefined for $s = \frac{1}{\phi-1}$. At the singularity we observe that

$$f_2 \left( \frac{1}{\phi-1} \right) = f_3 \left( \frac{1}{\phi-1} \right) = g \left( \frac{1}{\phi-1} \right) = \frac{\phi+2}{(\phi-1)^2}.$$  

It is easy to show that for $s > \frac{1}{\phi-1}$ we have

$$g'(s) < f_2'(s) \quad \text{and} \quad g'(s) > f_1'(s).$$

This implies that in this case

$$g(s) < f_2(s) \quad \text{and} \quad g(s) > f_3(s).$$

It follows that

$$f_1(s) < f_2(s) \quad \text{if} \quad \iota \leq 0 \quad \text{and} \quad f_1(s) > f_3(s) \quad \text{if} \quad \iota \geq 0.$$  

This shows us that the system cannot be stable for $s > \frac{1}{\phi-1}$. Because the case $s < 0$ also gives instabilities, we can restrict ourselves to the case $0 < s < \frac{1}{\phi-1}$. This shows that the stability domain for hexagons shrinks if we increase $\phi$ and becomes arbitrarily small as $\phi \to \infty$.

A tedious but elementary calculation confirms the identity

$$\frac{df_1}{ds} = \frac{f_2(s) - f_1(s)}{s(s(\phi-1) - 1)}.$$  

This equation can be interpreted as follows

The stability condition $f_2(s) < f_1(s) \equiv \sigma$ is fulfilled if and only if $f_1$ is increasing.

Because $f_1(s) \sim (2\phi + 1)s^2$ for large $|s|$ we know that $f_1$ must be decreasing for small $s$. Therefore, if the equation $\sigma = f_1(s)$ has for given $\sigma$ a solution $s$ with positive $f_1'(s)$ there has to be at least another solution. In fact, there have to be two such
Figure 4.11. The stability domain in the $\sigma\eta$-plane for hexagons for $\phi = 2$
FIGURE 4.12. The stability domain in the $\sigma\eta$-plane for rolls type I for $\phi = 2$
FIGURE 4.13. The stability domain in the $\sigma\eta$-plane for rolls type II for $\phi = 2$
solutions because the equation \( \sigma = f_3(s) \) can be written as the cubic equation (4.33). This finally proves our previous claim that (4.33) has to have three solutions for a stable regime and the middle solution is the only one that may be stable. It also tells us that for increasing \( \sigma \) and constant \( \iota \) the occurrence of hexagons is a saddle-node bifurcation.

4.5.3 Analysis of the von Karman equations: strongly anisotropic case

In the previous section we discussed the competition of different pattern planforms (rolls and hexagons) when the principal stresses are almost equal and the system is almost isotropic. Now we investigate the strongly anisotropic case. We learned that rolls are the only stable solution in this case. The linear analysis suggests that the wave vector of the rolls are roughly parallel to the largest principal stress. We wonder whether rolls with a slightly different wave vector are also stable solutions and whether there are differences in stability for different parameters. To ease the notations we will nondimensionalize the governing equations. We start with the following equations

\[
\begin{align*}
\kappa w_t + D \nabla^4 w + N_c(1 + \chi)w_{xx} + N_c(1 + \tau)w_{yy} + \gamma w + \alpha w^2 + \beta w^3 - [f, w] &= 0 \\
\frac{1}{Eh} \nabla^4 f + \frac{1}{2}[w, w] &= 0.
\end{align*}
\]

Here \( N_c \) is the critical stress. As usual, we assume that the direction of greatest compressive stress is along the \( x \)-axis, so we have \( \chi > \tau \).

Motivated by the results of the previous section we scale as follows

\[
\begin{align*}
t &\rightarrow \frac{\kappa t}{\gamma} \\
x &\rightarrow k_c^{-1} x = \sqrt[4]{\frac{D}{\gamma}} x \\
y &\rightarrow k_c^{-1} y = \sqrt[4]{\frac{D}{\gamma}} y \\
w &\rightarrow \sqrt{\frac{\chi}{\beta}} w \\
f &\rightarrow Df.
\end{align*}
\]
After scaling the following equations arise

\[ w_t + \nabla^4 w + 2(1 + \chi)w_{xx} + 2(1 + \tau)w_{yy} + w + \alpha \sqrt{\frac{\chi \gamma}{\beta}} w^2 + \chi w^3 - [f, w] = 0 \]

\[ \nabla^4 f + \frac{6(1 - \mu^2)\chi \gamma}{k^2 \beta} [w, w] = 0. \]  

(4.37)

It turns out that there are only four parameters left. Parameters \( \chi \) and \( \tau \) tell us how much the principal stresses are above (or below) critical stress. Further we have a parameter

\[ \hat{\alpha} = \alpha \sqrt{\frac{\chi \gamma}{\beta}} \]

that measures the degree of up–down asymmetry. Finally there is the parameter

\[ \Lambda = \frac{12(1 - \mu^2)\chi \gamma}{k^2 \beta} \]

that determines the amount of feedback of the second equation into the first equation.

Referring to the results of the previous section notice that the amplitude of a roll is given by

\[ A = \sqrt{\frac{2\gamma \chi}{3\beta}}. \]

Then we have

\[ \Lambda = 2(1 - \mu^2) \left( \frac{3A}{k} \right)^2. \]  

(4.38)

Therefore \( \Lambda \) depends on the ratio of wave amplitude and thickness. We note that this ratio has to be small because otherwise the assumptions for the von Karman equations are not satisfied. We also want to point out the dependence of \( \phi \), that was introduced in the last section and used as a bifurcation parameter, on \( \chi \) and \( \Lambda \):

\[ \phi = 2 + \frac{3\Lambda}{8\chi}. \]  

(4.39)

It is easy to see that rolls given by \( w = e^{i\sqrt{1 + \chi}}x \approx e^{ix} \) and \( f = 0 \) are a solution of the equations above (following our convention of using \( f = 0 \) for straight parallel
rolls). We are now interested in solutions of almost straight rolls. Following the ideas in [73] we assume that $w$ has the form

$$w = A(X,Y,T) e^{iz}$$

where $X = \epsilon x$ and $Y = \epsilon y$ are slowly varying spatial coordinates and $T = \epsilon^2 t$ is a slowly varying time scale. This time we did not include an $\epsilon$ in $w$ because that was already scaled in during the nondimensionalization. Further we assume that $\chi = \epsilon^2 \chi' = O(\epsilon^2)$ meaning that the maximal stress is still close to the critical stress.

As in the above section we collect terms of same order and derive solution conditions that give rise to amplitude equations. Because we look at the anisotropic setting we will not include the quadratic term because there is only one favored wave direction and quadratic interaction of wave triads does not take place. Let us now expand the various terms in powers of $\epsilon$:

$$\nabla w = \left( \left( \epsilon \frac{\partial}{\partial X} + i \sqrt{1 + \chi} \right) i + \epsilon \frac{\partial}{\partial Y} j \right) A e^{i \sqrt{1 + \chi} z}$$

$$\nabla^4 w = \left( \left( \epsilon \frac{\partial}{\partial X} + i \sqrt{1 + \chi} \right)^2 + \left( \epsilon \frac{\partial}{\partial Y} \right)^2 \right)^2 A e^{i \sqrt{1 + \chi} z}$$

$$w_{xx} = \left( -1 + 2i \epsilon \frac{\partial}{\partial X} + \epsilon^2 \frac{\partial^2}{\partial X^2} \right) A e^{i \sqrt{1 + \chi} z}$$

$$w_{yy} = \epsilon^2 \frac{\partial^2}{\partial Y^2} A e^{i \sqrt{1 + \chi} z}$$

$$w^3 = 3A^2 A^* e^{i \sqrt{1 + \chi} z} + \text{higher harmonics}$$

$$-\frac{1}{2} [w, w] = \epsilon^4 \frac{\partial^2}{\partial Y^2} (AA^*) + O(\epsilon^5).$$

From the last equation we obtain for the second von Karman equation at leading order and ignoring higher harmonics:

$$\nabla^4 f = \Lambda \frac{\partial^2}{\partial Y^2} (AA^*).$$

(4.40)
Here $\nabla$ is the nabla operator in the slow coordinates $X$ and $Y$. This implies that the second order derivatives of $f$ are of order $O(\epsilon^2)$. Thus we have

$$-[f, w] = \epsilon^3 f_{yy} A e^{ix}.$$  

Regarding the first von Karman equation it is easy to see that all terms of order less then $\epsilon^2$ vanish. At order $\epsilon^3$ we obtain:

$$A_T = 2\chi' A + 4A_{xx} + \Gamma A_Y - f_{yy} A - 3\chi' A^2 A^*$$  \hspace{1cm} (4.41)

where

$$\Gamma = 2(\chi - \tau).$$

Equations (4.40) and (4.41) are the Newell–Whitehead–Segel equations for the von Karman equations in the anisotropic case.

Now we want to investigate the stability of solutions of the form

$$A = \alpha e^{ik\cdot x}.$$  

Such solutions correspond to rolls whose wave vector slightly deviates from the most amplified one. Again, we use $f = 0$. From (4.41) we obtain a relation between $\alpha$, $k$ and $\chi'$ as follows:

$$2\chi' = 3\chi' \alpha^2 + 4k_x^2 + \Gamma k_y^2.$$  \hspace{1cm} (4.42)

Hence, solutions exist if the wave vector components lie inside the ellipse

$$4k_x^2 + \Gamma k_y^2 \leq 2\chi'.$$

We investigate stability by perturbations as follows

$$A = \alpha e^{ik\cdot x} \left(1 + a^+ e^{iQ\cdot x} + a^- e^{-iQ\cdot x}\right)$$

where $Q = (Q_x, Q_y)$. Further, $a^+$ and $a^-$ depend on $T$. From this ansatz we obtain

$$AA^* = \alpha^2 \left(1 + a^+ e^{iQ\cdot x} + a^- e^{-iQ\cdot x} + a^+ e^{-iQ\cdot x} + a^- e^{iQ\cdot x}\right)$$

$$+ \text{higher order terms in } a^+ \text{ and } a^- \text{ that we neglect}$$
and it follows
\[ \nabla^4 f = \nabla^2 (AA^*) = -\Delta \alpha^2 Q_y^2 \left( a^+ e^{iQ_x} + a^- e^{-iQ_x} + a^+ e^{-iQ_x} + a^- e^{iQ_x} \right). \]

This yields
\[ f_{yy} = \frac{\Lambda \alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2} \left( a^+ e^{iQ_x} + a^- e^{-iQ_x} + a^+ e^{-iQ_x} + a^- e^{iQ_x} \right). \]

Using these results we can write down the linear terms in \( a^+ \) and \( a^- \) of equation (4.41):
\[
A_T \to A \left( a^+ e^{iQ_x} + a^- e^{-iQ_x} \right) \\
4A_{XX} \to 4A \left( -a^+ (k_x + Q_x)^2 e^{iQ_x} - a^- (k_x - Q_x)^2 e^{-iQ_x} \right) \\
2\Gamma A_{YY} \to 2\Gamma A \left( -a^+ (k_y + Q_y)^2 e^{iQ_x} - a^- (k_y - Q_y)^2 e^{-iQ_x} \right) \\
2\chi' A \to 2\chi' A \left( a^+ e^{iQ_x} + a^- e^{-iQ_x} \right) \\
3\chi'A^2 A^* \to 3\chi' \alpha^2 A \left( 2a^+ e^{iQ_x} + a^- e^{-iQ_x} + 2a^- e^{iQ_x} + a^+ e^{-iQ_x} \right).
\]

Substituting these equations into (4.41) and collecting all the \( a^+ \) and \( a^- \) terms gives rise to two linear equations as follows
\[
\begin{pmatrix}
  a^+ \\
  a^-
\end{pmatrix}_T =
\begin{pmatrix}
  c_1 & c_2 \\
  c_2 & c_3
\end{pmatrix}
\begin{pmatrix}
  a^+ \\
  a^-
\end{pmatrix}
\]

where we have using (4.42)
\[
c_1 = -4(k_x + Q_x)^2 - \Gamma(k_y + Q_y)^2 + 2\chi' - 6\chi' \alpha^2 - \frac{\alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2} \\
= -8k_x Q_x - 2\Gamma k_y Q_y - 4Q_x^2 - \Gamma Q_y^2 - 3\chi' \alpha^2 - \frac{\alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2} \\
c_2 = -3\chi' \alpha^2 - \frac{\alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2} \\
c_3 = -4(k_x - Q_x)^2 - \Gamma(k_y - Q_y)^2 + 2\chi' - 6\chi' \alpha^2 - \frac{\alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2} \\
= +8k_x Q_x + 2\Gamma k_y Q_y - 4Q_x^2 - \Gamma Q_y^2 - 3\chi' \alpha^2 - \frac{\alpha^2 Q_y^4}{(Q_x^2 + Q_y^2)^2}.
\]
The characteristic polynomial of the matrix in (4.43) is
\[ c(\lambda) = \lambda^2 + p\lambda + q \]
where \( p = -(c_1 + c_3) \) and \( q = c_1c_3 - c_2^2 \). It is easy to see that we always have \( p > 0 \). Therefore the characteristic polynomial has roots with negative real value (implying stability) if and only if \( q > 0 \). This gives us the following stability criterion. \( A = e^{i\lambda - \omega} \) is stable if and only if we have for all \( Q_x, Q_y \)
\[ q = c_1c_3 - c_2^2 = (4Q_x^2 + \Gamma Q_y^2)^2 + \left( 6\chi\alpha^2 + \frac{2\Gamma\alpha^2 Q_y^2}{Q_x^2 + Q_y^2} \right) (4Q_x^2 + \Gamma Q_y^2) - (8k_x Q_x + 2\Gamma k_y Q_y)^2 > 0 \]
(4.44)
We introduce the vectors
\[ \hat{Q} = (2Q_x, \sqrt{\Gamma} Q_y) \quad \text{and} \quad \hat{k} = (2k_x, \sqrt{\Gamma} k_y) . \]
Denote \( \delta \) the angle between the \( x \)-axis and \( \hat{Q} \), further let \( \omega \) be the angle between \( \hat{k} \) and the \( x \)-axis. The only term that we cannot immediately represent in terms of these vectors is \( Q_y^2/(Q_x^2 + Q_y^2)^2 \). But we have the following identity
\[
\frac{Q_y^4}{(Q_x^2 + Q_y^2)^2} = \left( \frac{4Q_x^2 + Q_y^2}{Q_x^2 + Q_y^2} \right)^2 = \left( \frac{\cot^2 \delta + 1}{\frac{2}{Q_x} \cot^2 \delta + 1} \right)^2 \text{ where } \cot^2 \delta = \frac{4Q_x^2}{\Gamma Q_y^2} .
\]
This implies
\[
\frac{Q_y^4}{(Q_x^2 + Q_y^2)^2} = \left( \frac{\cot^2 \delta + 1}{\frac{2}{Q_x} \cot^2 \delta + 1} \right)^2 \sin^4 \delta .
\]
We now rewrite the stability criterion (4.44) in terms of \( \hat{Q} \) and \( \hat{k} \) and obtain
\[
\|\hat{Q}\|^4 + \alpha^2 \left( 6\chi + 2\Lambda \left( \frac{\cot^2 \delta + 1}{\frac{2}{Q_x} \cot^2 \delta + 1} \right)^2 \sin^4 \delta \right) \|\hat{Q}\|^2 - 4\|\hat{k}\|^2\|\hat{Q}\|^2 \cos(\delta - \omega) > 0 .
\]
Dividing by $\|\hat{Q}\|^2$ we obtain

$$\|\hat{Q}\|^2 + \alpha^2 \left( 6\chi' + 2\Lambda \left( \frac{\cot^2 \delta + 1}{\frac{1}{4} \cot^2 \delta + 1} \right)^2 \sin^4 \delta \right) - 4\|\hat{k}\|^2 \cos^2 (\delta - \omega) > 0.$$ 

The most unstable case clearly occurs when $\|\hat{Q}\| \to 0$. This shows that in the case of an instability small wave numbers will be amplified most. Setting $\|\hat{Q}\|$ to zero and using

$$\alpha^2 = \frac{2}{3} - \frac{\|\hat{k}\|^2}{3\chi'},$$

gives us the following inequality for $\|\hat{k}\|$:

$$\|\hat{k}\|^2 < \frac{4\chi' + \frac{5}{3}\Lambda \left( \frac{\cot^2 \delta + 1}{\frac{1}{4} \cot^2 \delta + 1} \right)^2 \sin^4 \delta}{2 + 4 \cos^2 (\delta - \omega) + \frac{2}{3\chi'} \Lambda \left( \frac{\cot^2 \delta + 1}{\frac{1}{4} \cot^2 \delta + 1} \right)^2 \sin^4 \delta}. \quad (4.45)$$

For stability $\|\hat{k}\|$ has to be bounded by this expression for all $\delta$. Unfortunately the corresponding optimization problem is not trivial in general and has to be solved numerically. However there is an easier special case:

$\Lambda = 0$

In this case we totally ignore the contribution of the second von Karman equation. The remaining equation is a nonlinear version of the Swift–Hohenberg equation. It is easy to see that the stability condition for $\|\hat{k}\|$ reads as

$$\|\hat{k}\| < \frac{2\chi'}{3}.$$ 

Contrast that with the domain of all wave vectors for which the amplitude equations have straight roll solutions

$$\|\hat{k}\| < 2\chi'.$$

The most unstable mode is the one with $\delta = \omega$, therefore $Q$ points in the same direction as $k$. The stability domain is an ellipse around the origin (see Figures
Figure 4.14. The domain of the modes $k$ for which $A e^{i\omega}$ with $A = e^{i\lambda k}$ is stable. The stability domain (black) lies inside the domain of amplified wave vectors (grey). Parameters $\chi = 0.2$ and $\Gamma = 0.9$ were used for all examples. Further we have (a) $\Lambda = 0$, (b) $\Lambda = 1$, (c) $\Lambda = 5$, (d) $\Lambda = 30$.
Figure 4.15. The domain of the modes $k$ for which $A e^{i k \cdot \omega}$ with $A = e^{i k \cdot \omega}$ is stable. The stability domain (black) lies inside the domain of amplified wave vectors (grey). Parameters $\chi = 0.2$ and $\Gamma = 0.2$ were used for all examples. Further we have (a) $\Lambda = 0$, (b) $\Lambda = 1$, (c) $\Lambda = 5$, (d) $\Lambda = 30$.
If $A > 0$ the stability domains enlarges and contains more of the modes with large $|k_y|$ (see Figures 4.14 and 4.15). This effect has also been observed in the numerical simulations (see Section 6.2) If $A$ is very large the most unstable mode is the one with $\delta = 0$ (except for $\omega \approx \pi/2, 3\pi/2$); in that case $Q$ points in the direction of the largest stress. Decreasing $\Gamma$ elongates the stability domain and the situation approaches the isotropic case.

### 4.6 Determination of parameters

Ideally, biological models can be tested by checking parameter relations against real measurements. Often, as in the case of our model, these parameters are not available. The magnitudes of many geometric quantities are known, but we do not have mechanical parameters of the basal layer, such as Poisson's ratio $\mu$ and Young's modulus $E$ and the resistance parameters of the foundation, $\gamma$, $\alpha$ and $\beta$.

Even though we cannot confirm the model by testing parameters it is still important to get an idea of their magnitudes which we can use in the computer experiments. It is possible, of course, that they can be confirmed (or disproved) later on. Fortunately Poisson's ratio does not play a very significant role and a lot of other quantities will prove to be just proportional to $E$.

Let us get started with the geometrical parameters. In a previous section we estimated the buckling wavelength $\lambda$, the buckling wave number $k = 2\pi/\lambda$ and the shell thickness $h$ as follows.

$$
\lambda = 36 \mu m, \quad k = 0.17 \frac{1}{\mu m}, \quad h = 8.0 \mu m.
$$

Without further reference to Young's modulus, $E$, we can obtain an approximation of the strain, $\epsilon$, in the basal layer when buckling takes place. We use Hooke's Law,
assuming that the principal stresses have roughly the same value, and obtain

$$\frac{N_c}{h} = E\epsilon_c(1 + \mu).$$

Using (4.26), and solving for \(\epsilon_c\), yields

$$\epsilon_c = \frac{2\pi^2}{3(1 + \mu)} \left(\frac{h}{\lambda}\right)^2.$$

Using the numbers above, we obtain a strain of about 22%, which seems somewhat excessive, especially if one considers the displacements that are necessary to give this strain. Therefore, we use a smaller thickness \(h\) in the simulations. If we decrease \(h\) to 4 or 5\(\mu m\), the strain is less then 10%, which seems to be a more realistic value. The original estimate for the thickness was obtained by measuring the thickness of the basal cell layer, however, this is not necessarily the mechanically correct value. Possibly, cell connections are concentrated closer to the middle surface and the mechanically effective thickness is reduced this way. This modified value for \(h\) weakens our previous arguments somewhat that the forces, and not the curvatures control buckling, but does not invalidate them. For the remainder of this section we will use \(h = 5\mu m\), which was also used in the simulations.

Another somewhat problematic parameter is the ridge depth \(d\). From measurements at the fully-developed primary ridges we get an estimation of \(d = 12\mu m\). However, we have mentioned before that buckling is unlikely to establish the full ridge depth. Even if it did establish the full ridge depth we could not use the von Karman equation to model this process because the gradients \(w_x\) and \(w_y\) would be far too large. For the validity of the equations these gradients should be much smaller than unity, but as seen on Figure 2.5 these gradients often even exceed unity. We will rather model the situation depicted on Figure 2.4 where the primary ridges have already been created as shallow undulations but have not reached their final depth yet. Here we have a ridge depth of about \(d = 2\mu m\); we will use this estimate in the following.
Looking at the scaled von Karman equations (4.37) we notice that there are two
important parameters: \( \Lambda \), measuring the feedback from the second von Karman equa-
tions; and \( \chi \), the ratio of current stress to critical buckling stress. From equation
(4.38) we see that \( \Lambda \) essentially depends on geometric quantities as follows
\[
\Lambda = 2(1 - \mu^2) \left( \frac{3A}{h} \right)^2 \approx 0.57.
\]
Here we used \( \mu = 0.45 \) (see below for explanation) and note that \( d = 2A \). It is difficult
to deduce an estimate for \( \chi \) from the geometric data. However, we can use the fact
that the number of dislocations in the patterns increases as \( \chi \) increases. On the basis
of the computer simulations (see Chapter 6) we observed that a values of \( \chi = 0.2 \)
gives good results.

To continue, we need some estimate of Young's modulus \( E \). As we have pointed
out, no real measurements are available, but we can compare the basal layer to tissues
where Young's modulus has been measured. Skin, however, is not a very suitable
comparison object. Its properties are mostly characterized by keratin proteins that
are not expressed at primary ridge formation. Furthermore, it seems likely that the
mechanical properties of skin are more determined by the keratin rich outer layers
than by the basal layer.

For a more reasonable estimate we should look at other epithelial tissues like liver,
brain, bladder which all have a Young's modulus in the range from 1 to 100kPa [82].
In the following, we will therefore use an estimate of
\[
E = 10 \text{kPa} = 10^{-8} \frac{N}{(\mu m)^2}.
\]
This estimate could be significantly inaccurate, but it is the best we can currently
obtain and deviations from it do not significantly affect our conclusions.

Poisson's ratio in physical materials ranges from 0 (fully compressible) to 0.5
(incompressible). In our equations, \( \mu \) only appears in the bending modulus \( D \) via the
expression \( 1 - \mu^2 \). For the possible ranges of \( \mu \) this expression can only vary from
0.75 to 1, which is not very much. We estimate Poisson’s ratio as

\[ \mu = 0.45, \]

again this value could be inaccurate, but, as we have argued, this hardly influences our results.

From these elasticity constants we determine the bending modulus,

\[ D = \frac{Eh^3}{12(1 - \mu^2)} = 1.3 \cdot 10^{-7} \frac{N}{(\mu m)^3}. \]

Now we can use equations (4.26) to find the critical stress, \( N_c \), and the linear foundation constant, \( \gamma \),

\[ N_c = 2k^2D = 7.9 \cdot 10^{-9} \frac{N}{\mu m}, \quad \gamma = k^4D = 1.2 \cdot 10^{-10} \frac{N}{(\mu m)^3}. \]

Next we can find an estimate for the cubic foundation term, \( \beta \), using the ridge depth \( d \); we find, from the amplitude equations, the relation

\[ d = \frac{2\gamma X}{3\beta}. \]

Solving for \( \beta \) yields

\[ \beta = \frac{8\gamma X}{3d^2} \approx 1.6 \cdot 10^{-11} \frac{N}{(\mu m)^5}. \]

Further we can find the parameter \( \phi \) from equation (4.39),

\[ \phi = 2 + \frac{3\Lambda}{8\chi} \approx 3.07. \]

It remains the quadratic foundation constant, \( \alpha \). At this point we can only give an upper bound. We use the fact that ridges are virtually never seen on fingertips. We assume that ridges always appear if the ratio of the largest stress to the smallest stress exceeds a certain value \( 1/(1 + \rho) \). For our estimation we will use \( \rho = -0.01 \). We now find the largest \( \alpha \) that still gives us ridges for this assumption. From the bifurcation diagram 4.11 we determine for \( \eta = 0 \) the smallest value of \( \sigma \) that still gives unstable
hexagon solutions. This value is about $\sigma = -0.1$. According to our assumptions this point in the bifurcation diagram corresponds to $\varepsilon^2 \tau = \rho$ and $\varepsilon^2 \chi = 0$. Now we have

$$0.1 < -\sigma$$

$$= -\frac{\bar{\delta}}{(2\alpha')^2}$$

$$= -\frac{\frac{3}{4}(\chi + 3\tau)3\beta}{(2\alpha')^2}$$

$$= -\frac{9\gamma\beta e^2 \tau}{2(2\alpha)^2}$$

$$= -\frac{9\gamma\beta \rho}{8\alpha^2}.$$ 

Solving for $\alpha$ gives

$$|\alpha| < \sqrt{-\frac{9\gamma\beta \rho}{8 \cdot 0.1}}.$$ 

Putting in all numbers we finally obtain

$$|\alpha| < 1.5 \cdot 10^{-11} \frac{N}{(\mu m)^4}.$$ 

Because $\alpha$ measures the degree of up-down asymmetry in the spring response, it is clear that a considerable amount of asymmetry is still compatible with roll solutions.

### 4.7 Summary

The estimations made in the linear analysis helped to prove a very important hypothesis.

*The buckling process underlying fingerprint development is controlled by the stresses formed in the basal layer, not by the curvatures of the skin surface.*

At first sight this result seems somewhat contrary to the mainstream of research on epidermal ridges. It sounds paradoxical after all we have said about the connections between volar pad geometry and ridge development. However, it confirms the popular
idea that epidermal ridges form perpendicular to growth stresses in the epidermis. Let us summarize the facts that support the hypothesis above.

- Connections between skin geometry and ridge alignment have only been shown to exist on the volar pads. Muscular eminences, like the thenar or the calcar, fail to influence ridges.

- The geometry does not always determine ridge direction. For example, the lower phalanges of the fingers can exhibit both transverse lines (as on humans) and wedge-like patterns (as in certain monkey species). Other examples are the pads on monkeys and the pads on vulpine phalangers. Monkey pads are covered by ridges running along the elevations, the vulpine phalanger shows ridges running across the elevations, in the direction of greatest curvature. Finally, the human interdigital areas are rarely covered by whorls, in spite of often well-developed pads.

- We have seen that models relying on curvature alone, in the spirit of Penrose's hypothesis, cannot correctly predict whorl patterns on symmetric elevations.

- Curvature effects fail to reproduce the correct wavelength. Their magnitude in the dispersion relation is considerably smaller than the stress effects.

These arguments do not invalidate our discussion on the connection between pad geometry and ridge patterns, but they show that local curvature effects cannot account for the direction of the ridges. So far it is still possible that the curvature contributes to the formation of stress and influences ridge configurations this way. We shall see soon in the next chapter that this is indeed the case.

The weakly nonlinear analysis suggests that ridges are the preferred pattern to arise. This is due to two reasons:

- Even small anisotropies in the principal stresses drastically decrease the region of stability for hexagons.
In our model we have $\phi \approx 3$, which further slightly reduces the stability region of hexagons.

There is, however, a caveat. There are regions where the principal stresses are in fact very close in magnitude. Examples are the larger scale singularities, such as the center of whorls, the concave disclination in loop patterns and the triradii. The direction of largest stress changes rapidly in these regions, which is only possible if the principal stresses are equal at the singularity center and almost equal close to the defect center. It can still be argued that ridges show up at these locations because of the bias introduced by ridges in the surrounding area. This argument is especially strong for the triradii which are the areas last covered by ridges.

Finally we want to briefly discuss the hexagon patterns of the vulpine phalanger and the koala. As large areas of the volar surface of these species exhibit hexagons the model implies that these areas are characterized by an almost isotropic stress field, large up-down asymmetry (as measured in $\alpha$) and small $\phi$. We cannot directly confirm these predictions as measurements at the fetal stage are not available. However, there is one argument that suggests that anisotropies in the stress, indeed, determine the pattern type. The hexagon pattern dominates in flat, featureless areas where almost isotropic stress is a reasonable assumption. However, we encounter ridges around the nail furrow. As will be shown in the next chapter, the area around the nail furrow is expected to show significant stress anisotropy. Therefore, it is not surprising that ridges appear in these areas. The other areas of the volar surface of the vulpine phalanger where ridges are formed are elevations of the palm. Why ridges show up here and why the ridge direction seems to be along the lines of greatest curvature is still a mystery.

In contrast to ridges, hexagons do not have up-down symmetry. Therefore it makes a difference whether the dot pattern “points up” toward the skin surface or “points down” toward the dermis. The dot pattern of the koala and the vulpine
phalanger points up which means that the amplitudes of the three cosine waves that form the pattern are positive. This implies that the quadratic spring parameter, $\alpha$, is negative. This means that there is more resistance from the dermis than from the upper epidermis layers. If true, this situation would be different from the situation in humans, where the opposite is believed to be true.
CHAPTER 5

THE FORMATION OF STRESS IN THE EPIDERMIS

5.1 Biological considerations

We have argued in the previous chapter that the surface curvatures of the embryo do not directly determine the ridge pattern. Instead, the ridge direction is locally established by the direction of smallest stress in the basal layer. In this chapter, we will discuss how stresses arise in the epidermis and how these mechanisms are able to produce the pattern configurations we see. At the same time, we will establish how the mechanisms connect finger geometry and ridge direction; thereby providing a link to the empirical observations. These ideas are then tested by computer experiments.

To start, let us consider the case of a flat rectangular plate that is unable to expand at its boundaries. Assume that stresses arise in the basal layer due to differential growth, either by increased cell proliferation, or by an increase in cell volume. It is easy to see that a flat plate that experiences uniform growth but is not able to expand due to boundary restrictions will be subjected to uniform biaxial stress, so the stresses in all directions will be equal to each other. There will be no dominant stress direction to determine the ridge direction in the fashion outlined in the last chapter. Ridge patches with random direction and labyrinthic pattern will form, clearly not what we observe in actual fingerprints.

In the remainder of this section we will present mechanisms that avoid this problem. First, we will outline the two mechanisms we believe are most important for understanding fingerprint configurations. Then, we will discuss two other mechanisms, that at first sight seem plausible, and argue why they are not important.
1. Boundary effects

These effects are important to understand why the ridge direction in certain areas is almost always the same in different humans. To understand the basic idea refer to Figure 5.1. Here the major palm flexion creases, the wrist crease, the phalangeal creases and the nail furrow are shown as thick lines. We notice that the ridges run parallel to these lines. In other words,

*the ridges tend to align themselves parallel to the creases and furrows.*

Of course, this is only true for the creases that arise prior to ridge formation and not for the ones that form later. At this point, we do not have embryological evidence that the wrist crease is formed before ridges do. This, however, is clearly true for the other creases mentioned and the nail furrow.

Further, we notice that there is a relation between ridge direction and the margin of the palm. We see that

*the ridges usually arrive at a steep angle at the periphery of the volar surface.*

This angle is often very close to a right angle and almost never less than 45°.

There are a few minor exceptions of these two observations but they do not change this picture significantly. An example are the ridges close to the metacarpophalangeal crease (MCPC) of the thumb, which often cross the MCPC at angles up to 20°. However, these ridges usually reach the first interdigital space almost perfectly perpendicular to the palm periphery. Also, ridges tend to cross the flexion creases in areas where the creases are not very distinct and fade out. Exceptions like these usually occur if the ridges are subject to conflicting requirements and some kind of compromise has to be found.

The conclusions of the previous chapter provide us with a framework to understand these observations. Remember, that we consider an expanding cell sheet in which
FIGURE 5.1. The relation between folds and furrows (thick lines), the periphery of the palm and the epidermal ridges are shown. The ridges tend to run parallel to the thick lines and perpendicular to the margin of the palm.
FIGURE 5.2. (a) A cross-section of a surface with curvature $1/R$ (solid curve) is displaced by $w$ toward the center of the curvature circle (dashed curve). The cross-section becomes strained. Stress proportional to $w/R$ is induced. (b) No stress will be induced if the cross-section is free to expand in tangential direction.

compressive stress is generated due to resistance of the surrounding structures. It is likely that the nail furrow and the flexion creases prevent tangential expansion, whereas the margin of the palm may not (or only to a lesser degree). As the basal layer cannot expand toward the creases it will be subjected to compressive forces acting perpendicular to the creases. Since the ridges align themselves along the lines of smallest stress, they will form along the creases, as it is actually observed. The situation is exactly opposite at the margin of the palm. Because basal layer expansion is not resisted here, there are no forces perpendicular to the margin of the palm. This is clearly the direction of smallest stress. Hence, the ridges will align perpendicular to the palm periphery.

We should emphasize again that we are always interested in relative growth. All parts of the embryo's hand are growing, but the basal layer expands at a faster rate. The increased proliferation stress in the basal layer will push cells to the dorsal side of the hand, wherever this is possible. We postulate that such movement takes place across the palm margin, but is prevented across flexion creases and the nail furrow.

As the anatomical position of the flexion creases and the nail furrow in humans is usually very similar, the forces induced by boundary effects are also similar among
humans. This explains why the ridge direction is similar in certain regions for almost all people. However, boundary effects do not account for the more individual configurations that cover the blank areas of Figure 5.1. The next mechanism aims to explain how the stress field is determined in these regions.

2. *Normal displacements due to the regression of the volar pads induce tangential stress. This effect is most pronounced close to the ridge anlage.*

It is well-known that normal displacements of curved surfaces induce in-plane stress if tangential displacements are prevented (see Figure 5.2). The induced stress is compressive if the displacement occurs toward the center of the curvature circle. As we increase the curvature, the compressive stress increases as well. Therefore, normal displacements could possibly explain the observed connection between skin geometry and the stress field that determines ridge direction. But do we actually have normal displacements in the fetal skin when the primary ridges develop?

Primary ridge formation starts at a time when the volar pads digress and become less prominent. Therefore, the assumption that normal displacements create tangential stress is very reasonable. However, we have to take into account that the finger as a whole is actually growing at the time of pad digression. Even in this case, our argument is still valid. If all parts of the embryo finger grew at exactly the same rate, the finger would increase its size but exactly preserve its shape. No growth forces are induced in that case. In contrast, changes in the shape of the growing finger will induce forces. These forces can be deduced by relating the finger shape after pad regression to the original finger shape before pad regression (see Figure 5.3).

Looking at several of BONNEVIE's pictures of cross-sections through fetal fingertip pads (like in Figure 3.2) we notice that the usually nicely rounded outline of the pad often becomes flat or even slightly concave at the ridge anlage. Strangely, almost nobody has given this phenomenon much attention. BONNEVIE attributed it to the
FIGURE 5.3. (a) A cross-section through the apical volar pad at the beginning of volar pad digression. (b) If the fingertip grew exactly uniformly the fingertip would preserve its shape and only increase in size. (c) The fingertip changes its shape due to volar pad regression. The growth forces induced by the regression are obtained by morphing figure (b) into figure (c). To accomplish this task normal displacements toward the center of the pad are necessary.

fact that buckling has taken place and the stress is relieved. However, this argument does not explain the degree of concavity found in some specimen. BONNEVIE's observations have been confirmed by SCHAEUBLE [93] at the ridge anlagen of the interdigital pads and more recently by MOORE and MUNGER [66].

This change in concavity at the ridge anlage indicates that the normal displacements are especially large in that area. Indeed, it seems as if the epidermis is pulled in by some force here. We may even speculate about the probable origin of this force. The main clue is provided by Figures 3.1 and 3.2 which show the papillary nerve projecting to the ridge anlage. It is plausible that the presence of the nerve induces forces that pull in the epidermis. For instance, the nerve could absorb fluids and create an underpressure in the volar pad below the ridge anlage. In fact, this fluid absorption could possibly be the very mechanism for pad digression. If this interpretation is true it would mean that the position of the ridge anlage is indeed caused by the incoming papillary nerves and confirm the importance of the nervous system on the development of epidermal ridges.

In the next section we will discuss in more depth how the normal displacements create a stress field that predicts the most common fingerprint configurations. We will also explain how the geometry of the volar pad influences this process. These ideas
will then be tested by computer experiments. We will see that normal displacements, combined with certain boundary conditions, produce the common configurations, predict the right timing sequence of pattern formation and confirm the relationship between geometry and pattern type.

Nevertheless, there remain questions. For example, it is not clear how the stress field is established on the palms of certain monkey species, where the pads persist until adulthood. However, we may speculate that even here some digression of the pads takes place. The almost complete lack of specific information concerning the embryological history of monkey pads does not yet allow an answer to this question.

In our investigations, we considered the following mechanism as a possibility to explain the stress field responsible for creating fingerprint configurations.

3. The growth rate of the basal layer depends on the position on the volar surface.

There is some empirical evidence for this assumption. BONNEVIE’s and SCHAEUBLE’s investigations of the ridge anlage established it as an area of increased cell proliferations. It is intuitively clear (and can be confirmed by the computer experiments) that such a center of increased proliferation could give rise to stresses whose lines of greatest stress project radially from the center. As ridges align perpendicular to the greatest stress they form concentric circles and a whorl emerges. This effect has been reproduced in computer experiments.

Another reason why different growth rates may exist in different areas of the palm is the fact that some volar surfaces (apical and interdigital pads) are much earlier covered by ridges than others (proximal and middle phalanges). If the basal layer expands faster in a certain area more stress will form here and ridge formation will be initiated early. It seems that the timing of ridge initiation in the different parts of the volar surface is correlated to the amount of mesenchymal tissue underneath. Ridge formation takes place first on the volar pads. The first finger with visible primary
ridges is the thumb, which presents the largest amount of dermal tissue; the last one is the small finger, which has the least amount of dermal tissue. Finally, ridge formation at the fleshier proximal finger phalanx precedes development at the skinnier middle phalanx.

However, these facts could also be explained by normal displacements. Areas with pronounced pad digression will experience large stress on the basal layer. Therefore, ridges are first formed in such areas.

In spite of these findings we have to reject this mechanism as crucial for ridge pattern formation because the computer experiments using it were quite disappointing. Whorls and arches are still fairly easy to reproduce, but realistic loop patterns (the most frequently occurring pattern in humans) are almost impossible to get. The loops that were achieved also do not predict the right sequence of ridge formation (starting in the center and at the nailfurrow). Furthermore, it is unclear how asymmetry of the pad is related to asymmetry of the loop. While performing the computer experiments and using reasonable assumptions concerning basal layer expansion quite frequently bizarre patterns were encountered that did not resemble anything seen on real fingers. The combination of the two previously mentioned mechanisms does not suffer from these problems.

This following mechanism was pointed out in fingerprint literature and may be important to understand the stress field in some instances.

4. *Stresses arise due to geometry changes of the palm.*

Besides the regression of the volar pad, there are other changes in the shapes of palm and fingers. For certain monkey species, the suggestion has been made that these changes induce stress that determines ridge direction. The idea is as follows: if an area of the palm predominantly grows in a certain direction at the time of ridge formation, growth stress in the basal layer will build up perpendicular to this direction.
and, hence, determine ridge direction. This sounds reasonable, but there is little evidence that such an effect is important for the pattern in human palms where the ridge direction is not uniform and seems to be more influenced by the crease lines and interdigital pads. The mechanism possibly plays a role for the ridges on the sole, proximal to the interdigital areas. Here, in the absence of pads and crease lines, the ridges run transverse to axis of the sole and change little in their orientation. The effect has also been related to the ridge direction in certain monkey species where the ridge direction is almost uniform over the palm.

Thus, we can formulate our hypothesis regarding the formation of stresses as follows:

*The stresses that determine ridge direction are themselves determined by boundary forces acting at creases and the nail furrow and normal displacements, which are most pronounced close to the ridge anlage.*

### 5.2 The role of hand geometry – a heuristic explanation

In this section we will explain how the combination of boundary effects and normal displacements leads to the most common fingerprint configurations and what role the volar pad geometry plays in this process. On the one hand, it would not be wise to completely rely on the computer experiments as the assumptions that are used for their realization may influence the results. On the other hand, an analytical solution to the problem is not yet available. Therefore, without a formal mathematical discussion, we will attempt to give an idea why the computer program produces the results it does.

In the previous section we argued that normal displacements toward the center of the curvature circle induce compressive stress if expansion in tangential direction is resisted. As the curvature and the resistance to tangential expansion increases, the induced stress increases as well. Resistance to in-plane expansion is mainly provided
by the following two effects.

- Tangential expansion is not possible across flexion creases and the nail furrow, as we have argued above.

- Expansion is further resisted due to cell connections of the basal layer with the dermis and the intermediate layer. This effect can be thought of as tangential springs that allow small tangential displacements, but prevent large ones.

The lack of resistance can significantly change the stress field that we would expect solely looking at the curvatures. As an example let us consider the human interdigital pads that are located just proximal to the outline of the palm between the fingers (the location is seen as the dot in Figure 5.4 (b)). At the time of ridge initiation the interdigital pads subside, but are often still well-developed. As the pad regresses, there are normal displacements at the center of the pad which is significantly curved in both G and H directions. One would expect that the normal displacements induce stress along these directions, and that the ridges will form transverse to G and H creating a whorl configuration. However, whorls are very rarely formed in the interdigital areas. Why?

The answer can be found taking the boundary conditions into account. According to our theories the distal margin of the palm does not (or only to a small extent) resist tangential expansion. Therefore expansion is not resisted in direction H and, hence, little stress is induced. In direction G, however, there is no close free boundary and expansion is resisted by tangential springs. Therefore, large stress will be formed in this direction. This implies that ridges will form transverse to G and parallel to H, as it is usually observed in humans. This peculiar interplay between curvature and boundary confirms the necessity to look beyond the curvatures in certain cases.

Let us now consider the situation at the fingertips. To introduce some useful definitions let us refer to the illustration of a finger given in Figure 5.4 (a). Here we give the position of the ridge anlage and define directions A–F that radiate from it.
The configuration type (whorl, loop or arch) on a finger can be found if we determine whether the ridges follow the given directions or whether they cross them transversely. We note that ridges in almost all patterns run transverse to A, B and C. If they also run transverse to, D, E and F we have a whorl. If the ridges run transverse one of the directions D and F but not to the other a loop is present; and if they run parallel to both directions D and F an arch or tented arch.

At first, consider that we have a highly rounded, almost spherical, pad with the ridge anlage close to the summit. At the ridge anlage, curvature is present in all directions and there is no close free boundary. Therefore in–plane expansion is resisted by tangential springs. Large stress radiating from the ridge anlage is induced and a whorl forms. The behavior at the nail furrow and the digital interphalangeal crease (DIC) is largely determined by boundary forces that act perpendicular to the flexion crease and the nail furrow. This situation is illustrated in Figure 5.5 (a).

Let us now imagine that the pad is somewhat less pronounced. In this case, the curvatures along A–C exceed the ones along B–E, because the finger is significantly
FIGURE 5.5. The forces shaping a whorl. (a) The volar pad is high, rounded and normal pressure in the area produces radial stress away from the ridge anlage. (b) The volar pad is less pronounced and stress along A–C dominates in the center.

curved along A–C even in the absence of apical pads. Therefore, normal displacements induce larger stress along A–C, that is well–resisted by the boundaries. This stress dominates at the ridge anlage and the center of the whorl becomes elongated. See Figure 5.5 (b) for an illustration.

As the pad becomes flatter the stress induced by normal displacements decreases because the amount of curvature decreases and not much further digression takes place that could result in normal displacements. Therefore the radial stress along D–F decreases. At some point the radial stress along these directions is overpowered by the boundary stress. The principal stress is now transverse to D–F, which makes the ridges align parallel to D–F. A tented arch has formed. See Figure 5.6 (a) for a visualization. If the amount of normal displacements is decreased further, hardly any stress is generated away from the boundaries. The stress field is then dominated by the boundary stress, resulting in an arch. This situation is seen in Figure 5.6 (b).

In order to understand how the loops enter the picture let us consider the asymmetric case. Let us start with the tented arch and consider two possible asymmetries:
FIGURE 5.6. The forces shaping tented arches and arches. (a) For a flatter pad little stress is generated in direction D–F. Boundary effects dominate and a tented arch arises. (b) If the pad is even flatter and little normal pressure is applied the boundary forces dominate the stress pattern and arches form.

asymmetries in the volar pad and asymmetries in the location of the ridge anlage.

By asymmetry on the volar pad, we mean that it is slanted toward one side. For clarity, let us say it is the radial side, the case for the ulnar side is equivalent. The location of greatest curvature for each cross-section can now be found on a curve radial to the ridge anlage, as demonstrated in figure 5.7 (a). The stress produced by the digression of the pad at the ridge anlage is predominantly along A–C and the ridges form along B–E. An ulnar loop forms in this situation if the boundary forces from the ulnar side have a larger effect on the stress field at the center of the pad than the boundary forces from the radial side. In this case, the ridges in the center are slightly bend toward the ulnar side. Even a small change of this kind destroys the tented arch (a rare configuration after all) and creates a loop (the most common configuration). But why is the effect from the ulnar side larger than the one from the radial side? The boundary forces are not only resisted by tangential springs, but also by normal springs if large curvature is present. The asymmetry of the pad implies more curvature on the radial side than on the ulnar side. Therefore, the boundary
FIGURE 5.7. The forces shaping loops. (a) If we have an asymmetric pad (the line of the location of greatest radius is indicated by the dashed line) the pattern becomes asymmetric as well. For moderate pad elevation stresses in direction A–C dominate. The direction of the ridges in the center of the pad is influenced by the boundary forces. (b) If the ridge anlage is asymmetric the pattern becomes asymmetric as well due to interaction of the stress arising at the ridge anlage and the stress arising at the boundary.

stress from the radial side is more resisted and has a smaller effect on the stress field in the center of the pad. The effect from the ulnar side dominates, bends the ridges toward the ulnar side and creates an ulnar loop. In fact, the ridges of the loop core are often remarkably parallel to the ulnar nail furrow.

Now let us discuss the case when the shape of the volar pad is still symmetric, but the location of the ridge anlage is displaced from the center of the pad; say it is displaced to the ulnar side. The argument outlined above can be made in a similar fashion. Again, the influence of the boundary forces decides the direction toward which the ridges at the center of the pad are bent. Stress induced by the ridge anlage is best resisted toward the ulnar side of the nail furrow because this is the closest resisting boundary. Therefore stress at the ridge anlage tends to be perpendicular to the ulnar nail furrow. Therefore, the ridges at the center form parallel to the ulnar nail furrow. An ulnar loop arises.
Finally we want to point out that certain pattern configurations do not appear on fingerprints although they are not unplausible at first sight. As an example we refer to Figure 5.8 which shows a whorl pattern with inner ridges running perpendicular to the finger axis. If the ridges in the center were influenced by the boundary stress we would expect that an arch forms. If, however, the ridges were determined by stresses arising due to curvature, the curvature along A–C would have to be larger than the one along B–E. This is not the case in embryonic fingers. Consequently this kind of whorl is not observed.

This theory is consistent with the empirical observations on the connection between pad geometry and ridge configurations. Indeed, bulgy pads give rise to whorls and flat pads give rise to arches. Asymmetries in the pad or the position of the ridge anlage lead to asymmetries in the ridge pattern. An interesting direction of future research would be an extension of the theory to include more complex fingerprint configurations to understand how they form.
CHAPTER 6

COMPUTER EXPERIMENTS

To test the ideas discussed in this work, a computer program was written that simulates the conditions that we assume are present when fingerprint formation takes place. We show that our results are consistent with the empirical observations and produce the three common configuration types.

In the simulations we will proceed in two steps. At first, we determine how different forces, growth rates and geometries produce a certain stress field. It will be seen that this stress field anticipates in many respects the later buckling pattern. This first step will be accomplished by a finite element algorithm. Second, we will use the stress field that was obtained in the first step as input to the von Karman equations. The equations will be solved using a spectral method. In this chapter we will mainly discuss the simulation results. We refer to Appendix B for the details of the numerical algorithms we used.

6.1 Finding the stress field

In order to simulate the situation in the embryo finger during ridge formation two different approaches are possible.

1. Specify certain expansion rates which are resisted by boundary, tangential and normal springs.

2. Specify forces on the shell that are motivated by the model.

Approach 1 has the advantage that it is directly obtained from the model presented above and very intuitive. Approach 2 is not as direct because the forces have to be generated somehow. However, we chose approach 2 for our simulations because
it is much simpler to implement and much easier to obtain meaningful results. In approach 2 it is not obvious to implement resistance to growth at the boundary. Further, one has to specify in what coordinate system (expanding or not-expanding) the tangential springs act. Finally, unrealistic stress patterns are generated for plates with very different side lengths.

Approach 1 was used for a now obsolete model of a flat finger with nonconstant expansion rates and springs acting from the boundary. This model could reproduce all major patterns but the parameters had to be carefully tuned (especially for loops). It also produced a lot of other bizarre patterns and exhibited large regions of almost equal principal stresses. In view of these effects this flat fingerprint model was finally abandoned.

In approach 2 the stress field was obtained using a shell with a fingertip geometry. Different geometries (high pad/low pad, symmetric pad/asymmetric pad) were
considered. On this shell the following forces were applied:

- Boundary forces perpendicular to the nail furrow and the digital interphalangeal crease (DIC) compressing the shell.
- Normal load concentrated at the ridge anlage leading to normal displacements
- Normal and tangential springs.

The boundary forces and the normal and tangential spring constants were the same for all the examples presented here. The normal load and the geometry, however, were varied. The direction of the boundary forces is given by figure 6.2. The magnitude of the boundary force is maximal in the spherical part of the fingertip (corresponding to the nail furrow) and decreases along the cylindrical part going down. The forces acting on the bottom (corresponding to the DIC) are smaller than those applied to the spherical part.
FIGURE 6.3. (a) A pad slanted to the left side is subjected to a normal pressure centered at the ridge anlage which is shifted to the right side. (b) A loop forms.

FIGURE 6.4. (a) The pad is flat and little normal load is applied. (b) An arch arises.
In the following we give examples for the most frequently occurring fingerprint patterns. For each example we provide the magnitude of the normal pressure as given in loadn.tri (large load – black, small load – white) and the direction of smallest compressive stress as given in maxstress.tri. Note that the direction of smallest compressive stress predicts the buckling direction. The file maxstress.tri also provides the magnitude of the largest principal stress (small stress – black, large stress – white) and therefore predicts in what region fingerprint formation takes place first.

*Whorl*

The pad is highly rounded and a lot of normal load is applied at the summit of the pad. The stress pattern predicts a whorl, the largest stress is found around the center of the whorl and at the periphery. See Figure 6.2.

*Loop*

The pad is strongly slanted to the left hand side and the center of largest normal load is shifted to the right side. The normal load is not as concentrated as in the case of a whorl. A loop pattern arises, the loop opens to the right hand side. The largest stress occurs at the periphery and close to the loop core. See Figure 6.3.

*Arch*

The pad is flat and symmetric. Small normal pressure acts on the pad. The normal load is uniform over the fingertip and very small. The stress pattern predicts an arch. Largest stress is found at the periphery. See Figure 6.4.

Both the connection between pad geometry and configuration type and the predicted timing on ridge spread is in perfect agreement with the descriptions of BONNEVIE, GOULD and others.
6.2 Numerical Observations at the von Karman equations

In this section we discuss numerical simulations of the von Karman equation for certain simple stress fields to get some intuition for how the model parameters influence the buckling pattern. We use the following version without the curvature terms:

\[
\begin{align*}
&\kappa \psi_t + D \nabla^4 \psi - N_x \psi_{xx} - N_y \psi_{yy} - 2N_{xy} \psi_{xy} - [f, \psi] + \gamma \psi + \alpha \psi^2 + \beta \psi^3 = 0 \\
&\frac{1}{Eh} \nabla^4 f + \frac{1}{2}[\psi, \psi] = 0.
\end{align*}
\]

(6.1)

We focus on the effects of \( \chi \) (the relative amount over critical stress) and \( \Lambda \) (the amount of feedback from the second von Karman equation) because these are the most important parameters in the nondimensionalized equation (4.37).

In our scheme we used periodic boundary conditions. Further the Airy stress function \( f \) is chosen to vanish for straight parallel rolls. Remember that \( f \) models the decrease in the stress field after buckling occurs. Regardless what boundary conditions are actually present, they are certainly not periodic (unless the finger looks like a torus). However, in our later simulations the decrease in stress will not be strong (as signified by a small \( \Lambda \)) because the wave amplitude is still quite small, therefore, the boundary conditions are not very important and the scheme is acceptable. In this section we also explore the effects of larger amplitudes (as signified by a large \( \Lambda \)).

For the definition of the stress fields we use polar coordinates. Here \( r \) is the radial coordinate and \( \theta \) the circumferential coordinate. SI to SII are defined to simulate target patterns that can be seen as idealized whorls:

\[
\begin{align*}
\text{SI} : & \quad N_r = (1 + \chi)N_c \quad N_\theta = (1 + \tau)N_c \\
\text{SII} : & \quad N_r = (1 + \chi)N_c(1 - dr) \quad N_\theta = (1 + \tau)N_c(1 - dr)
\end{align*}
\]

where \( \chi > \tau \). The largest stress in this stress field points toward the origin and a target (ridges forming concentric circles around the origin) is expected to arise. Different from the whorls we observe in fingerprints the ratio of the principal stresses does not tend to 1 as we move toward the whorl center.

As the patterns in Figures 6.5, 6.6, 6.7 show, such target patterns are indeed found in our simulations, however, the pattern texture varies depending on the chosen
FIGURE 6.5. Buckling patterns for the stress field SI using $\chi = 1.05$ and $\tau = 0.95$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$
FIGURE 6.6. Buckling patterns for the stress field SI using $\chi = 1.2$ and $\tau = 1.1$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$
FIGURE 6.7. Buckling patterns for the stress field SI using $\chi = 1.6$ and $\tau = 1.4$ for different values of $\Lambda$. (a) $\Lambda = 0$ (b) $\Lambda = 0.57$ (c) $\Lambda = 3.1$ (d) $\Lambda = 18.4$
**Figure 6.8.** A target according to stress field SII. Here we used $\chi = 0.2$ and $\Lambda = 0.57$.

**Figure 6.9.** A loop consisting of a convex disclination and many dislocations according to stress field SIII. Here we used $\chi = 0.2$ and $\Lambda = 0.57$. 
FIGURE 6.10. A loop consisting of a convex disclination and many dislocations according to stress field SIV. Here we used $\chi = 0.2$ and $\Lambda = 0.57$.

parameter. For $\chi = 0.05$ we find very few dislocations that become more numerous as we raise this value. Hence, the number of dislocations was the criterion for choosing $\chi$. A value of $\chi = 0.2$ seems to roughly correspond to the number actually seen in fingerprints, therefore it was used for the simulations in the following section.

Raising $\Lambda$ leads to two interesting phenomena. Whereas the pattern is very smooth for $\Lambda = 0$, it has more wriggles as the parameter increases. This makes sense in view of our analysis in Section 4.5.3 that predicts such a behavior (modes not aligned with the critical stress become more stable). The second effect is a very pronounced decrease in amplitude at the target center if $\Lambda$ is large. Even for $\Lambda = 0.57$ (that we will use in the next section for the simulations) this effect is present, but not very strong. The reason for the amplitude decrease is the decrease in the stress field. In a target pattern the rolls are not straight anymore, the $[w,w]$ term does not vanish and actually has an effect on $f$.

Because the scheme we use for the solution of the von Karman equation enforces
periodic boundary conditions (see Section B.2) one could expect numerical artefacts at the boundary. These artefacts occur, sometimes ridges extend to the other side of the domain; but overall these effects are very mild.

A version of a target with a radially decreasing stress (SII) is given in Figure 6.8. It can be understood as an idealized developing whorl configuration.

Further we define two more fields that represent a convex disclination and represent idealized loops:

\[
\text{as in SI for } y > 0 \\
N_x = (1 + \chi)N_c \\
N_y = (1 + \tau)N_c \\
\text{otherwise}
\]

\[
\text{as in SII for } y > 0 \\
N_x = (1 + \chi)N_c(1 - dx) \\
N_y = (1 + \tau)N_c(1 - dx) \\
\text{otherwise}
\]

Simulation results are given in Figures 6.9 and 6.10. Changing the parameters (not shown here) produces similar results as in the target case.

6.3 Simulating the ridge pattern

We are going to approach the problem by taking the stress field obtained by the finite element program and use it to parameterize the von Karman equations where we dropped the curvature terms. Inevitably, this process has the disadvantage that the stress distribution that was obtained on the curved model has to be projected into the plane. Regardless what method is used for this process, distortions are impossible to avoid. These distortions are especially pronounced in the spherical geometry close to the fingertip and in whorls which deviate significantly from the cylindrical fingershape.

However, there are a number of reasons why these distortions are still acceptable. First, the spherical area close to the fingertip usually just exhibits straight ridges following the nail furrow and is not that interesting for the understanding of larger-scale pattern types like whorls or loops. Second, highly curved pads are destined to digress during fetal development and flatten out. And finally, the very process of
actual fingerprinting involves a projection from the ridge pattern on the finger onto a sheet of paper.

Generally we expect that the ridge lines form perpendicular to the lines of greatest stress. Therefore we anticipate that the direction of the ridge pattern will be given by the picture obtained in \textit{maxstress.tri}. Deviations from this picture are likely to occur especially in areas like centers of whorls, loops or triradii where the ratio of the principal stresses is close to 1.

Even after finding the model parameters and the stress field, there are different opportunities to use these in order to simulate the ridge pattern. In the following we will illustrate the following two procedures.

1. At every point the stress field is scaled such that the greatest principal stress is exactly $1.2N_c$. The instability then occurs everywhere simultaneously.

2. The complete original stress field is scaled such that it is subcritical everywhere. Then the stress is raised slowly. The instability starts in areas of greatest stress and then slowly spreads over the whole surface. The stress is not raised anymore at a certain point if the critical stress reaches $1.2N_c$.

There is no question that approach 2 is the more realistic, because the ridges do spread slowly over the volar surface of the embryo. However, the results obtained by approach 1 are much better.

The patterns obtained by approach 1 (Figures 6.11, 6.12 and 6.13) simulate the general "flow" of fingerprint patterns nicely. Also, the "texture" of dislocations (minutiae) is close to the one we observe in real fingerprints. It is observed that the wave vector is not always exactly parallel to the direction of greatest stress. This especially happens in areas where the ratio of the principal stresses is close to 1. Therefore it sometimes happens that the wave vector changes direction quickly when we follow the ridges from one parallel patch to another. Such a situation can be seen in Figure 6.11 above the triradii. Although such situations occur in real fingerprints they are
FIGURE 6.11. A whorl pattern obtained using approach 1. Ridge formation takes place simultaneously over the surface. The characteristic features of the whorl including the dislocations are well-captured. However, note the sharp change in the wave vector above the triradii.

FIGURE 6.12. A loop pattern using approach 1. The pattern exhibits the characteristic flow of loops. The triradii and the loop core are well-represented.
FIGURE 6.13. An arch pattern using approach 1 exhibiting realistic flow of the pattern.

quite rare. Note also that this phenomenon does not occur above the triradius in Figure 6.12 where there is a much slower change in the wave vector.

The dislocations (branches and endings) in our simulations mostly occur in two circumstances. They show up when the ridges diverge from each other and new ridges are inserted. Further, they arise in regions where almost parallel ridge patches meet each other (such as the triradii) and the cores of whorls and loops. These are also areas where the ratio of the principal stresses is close to 1. Our observation is consistent with statistics on fingerprint data in forensics [61] showing that dislocations in such areas are more frequent than at the periphery. In our simulations we obtained very few dislocations at the periphery, where they are in fact not that rarely observed in real fingerprints. Dislocations at the periphery could arise due to growth of the finger that makes insertion of new ridges necessary. At this point we have not attempted to model such effects.

Changing the initial conditions or parameters slightly results in a distinct change
in the placement of the dislocations. In this sense, the dislocations are quite sensitive to their "environment". This is a reason for the well-observed fact that fingerprints are indeed unique and can be used for identification purposes. This observation is especially important considering the recent challenges of fingerprint permissibility in court and could be a starting point for making the foundations of forensic fingerprint science more credible.

There are a few problems in our buckling simulations.

- In some places, especially where the ridge direction changes rapidly, the simulated pattern looks stiffer than real fingerprint patterns. The reason for this is the rigorous enforcement of the wavelength in the simulations that leads to dislocation cascades when the pattern diverges. In actual fingerprints, dislocations are sometimes avoided by a local change in wavelength that tends to smooth out sharp corners. A possible reason for this difference could be nonelastic behavior of the basal layer that makes the material more pliable.

- Although many minutiae arise their appearance does not completely satisfy. It is often somewhat obscure whether we have a bifurcation or a ridge ending. Frequently adjacent ridges exhibit "kinks" that are not seen in real fingerprints that appear very smooth.

Although approach 2 does not produce as convincing pictures (see Figures 6.14 and 6.15) they are still illuminating and point the way to more realistic models. The sequence of ridge formation — starting at the boundary, the core of loops and whorls, the spreading over the surface and finally filling in the triradius — is correct. Further the resulting patterns can clearly be recognized as loops and whorls. However, there are some problematic features. The pattern appears very "stiff", they seem to consist of patches of parallel ridges with discontinuous changes in the wave vector between the patches. Further there are much fewer dislocations, and the ones that arise are distributed in a stereotypical fashion, usually along the patch boundaries.
FIGURE 6.14. A whorl pattern forms. Ridge formation takes place at the periphery and the whorl core and spreads over the remainder of the skin surfaces. Finally the triradii are filled in.
FIGURE 6.15. A loop forms. The ridges start to form at the periphery and the loop core, finally the triradius is filled in.
It is not difficult to locate the reason for these problems. In the course of ridge formation in these simulated patterns the ridges tend to arise adjacent to existing ones. Even if the underlying stress field slightly changes, the wave vector of the existing ridges creates a large enough bias for the new ridges to follow in the same direction (especially if the ratio of the principal stresses is close to 1 and no direction is clearly favored). Patches of parallel ridges arise this way that are not compatible when they meet and the wave vector becomes discontinuous where these patches join each other.

However, approach 1 makes clear that it is possible to use the information coded in the stress pattern to obtain buckling patterns that exhibit the flow of the main fingerprint patterns. Admittedly, we do not have a fully convincing model that also correctly describes the right sequence of ridge formation over the fingertip.

These problems do not mean at all that the equations are useless. Quite the contrary, they were enormously helpful to understand the onset of buckling. The results we obtained from the linear and weakly-nonlinear analysis are consistent with the biological observations and helped us to gain a better understanding how the pattern is laid out. However, in our investigations we limited ourselves to observations that are determined at the instability or shortly after the instability like wavelength, pattern type and ridge direction. We have used very little information about the minutiae and the spread of ridges, in part because little quantitative data is available.

The von Karman equations should be seen as a first model that already captures many important fingerprint features and leaves room for significant improvements. Although buckling is the likely mechanism for the instability that creates fingerprint patterns it cannot explain everything one would like to know about fingerprint development. A more sophisticated model translating the stress pattern into the ridge pattern should use incorporate effects such as:

- The ridge direction should be chosen locally by the stress field.
• The ridges should not form simultaneously but rather first form in patches and then spread out over the palm.

• The model could use biological information concerning the time of ridge spread. Such information is provided by BLECHSCHMIDT [17] and MISUMI [64] (changes in the cell properties during the folding process).

• The model should take account of the fact that the digression of the volar pads changes the geometry of the fingertip.

• The model should, by including nonelastic material properties, improve the pattern flow in areas where the ridge direction changes rapidly.
CHAPTER 7

CONCLUSIONS

7.1 Summary

We are finally able to formulate our hypothesis on the development of epidermal ridges.

The epidermal ridge pattern is established as the result of a buckling instability acting on the basal layer of the epidermis and resulting in the primary ridges.

The buckling process underlying fingerprint development is controlled by the stresses formed in the basal layer, not by the curvatures of the skin surface.

The stresses that determine ridge direction are themselves determined by boundary forces acting at creases and the nail furrow and normal displacements, which are most pronounced close to the ridge anlage.

Referring to this picture we can suggest the following scenario of ridge development

- The fingers separate from each other at the 6th week of pregnancy. This process leads to asymmetries in fingershape, especially in finger II and V.

- Volar pads appear on the fingertips and in certain areas of the palm at around the 7th week. In the following weeks the volar pads become more pronounced and slowly subside starting from the 10th week.

- At about the 11th week an unidentified stimulus initiates differential growth in the basal layer in the epidermis.
- This growth is resisted by creases and folds in the tangential direction and causes compressive stress.

- Further compressive stress arises due to the digression of the pads. This effect is most pronounced in the area above the nerve twigs innervating the volar pad, possibly due to absorption of fluids.

- The stress in the basal layer leads to buckling, the ridges form perpendicular to the greatest stress.

- Depending on the height and asymmetry of the pad, and the magnitude and distribution of normal stress, whorls, loops, arches or accidental patterns form.

- Ridges start to form at the nail furrow, the creases and the ridge anlage and spread over the fingertips and the remaining volar surface in the course of several weeks.

- This process, although initiated by buckling, is likely modulated by biological effects such as changes in cell properties in different locations of the primary ridges and nonlinear material properties.

- In the course of ridge maturation, the number of minutiae increases, possibly due to growth of the fingertip.

### 7.2 Achievements and critique

One way to estimate the quality of our work is to look at Figure 7.1 and compare the real and simulated patterns. They surely have a lot in common, such as the topology, an established wavelength and similar ridge direction. They are different in some respects that are not important because the underlying stress field is somewhat different (for instance the simulated loop appears "higher up" than the real one). And there are some differences that should be the cause of further work such as the
behavior in regions of rapid ridge direction change or the stiffness of the simulated pattern in some regions.

Even more important than the resemblance of reality and simulation is the fact that the ideas in this work integrate the mainstream ideas on fingerprints. Our model confirms that:

- Primary ridges are formed as the result of a buckling process as postulated by Kollmann, Bonnevie and Schaeuble.

- Ridges form perpendicular to the lines of greatest stress as postulated by Cummins, Hale, Mulvihill and Smith.

- Volar pad geometry influences the fingerprint pattern as observed by Whipple, Bonnevie and Babler.

- The nervous system influences this process as claimed by Bonnevie, Dell, Munger and Moore.

- Although ridges are the usual pattern, dots (hexagons) are another possibility as reported by Okajima.

- After the buckling instability has taken place and the ridge pattern is established, cell proliferations may increase the depth of the primary ridges as has been suggested by Whipple and Hale.

Further we do not rule out that other biological effects that are reported in the literature influence the postbuckling behavior and form the fingerprint pattern on our fingers, palms and soles, that we see every day.

The mathematical techniques provide the means of relating these ideas into a unified whole. For example, the linear analysis of the von Karman equations establishes the link between ridge direction and stress tensor. The weakly–nonlinear analysis strongly suggests that ridges, instead of hexagons, form in humans and most
FIGURE 7.1. Real and simulated fingerprint patterns: (a) whorl, (b) loop and (c) arch.
primates. Finally, the computer experiments provide the possibility to test how certain forces on the basal layer result in a stress pattern that in turn establishes the buckling pattern. Based on this, we were able, for the first time, to suggest specific growth forces that establish the connection between pad geometry and pattern type and are biologically motivated. These growth forces seem crucial for a more profound understanding of fingerprint formation.

To analyze these ideas we aimed for the most straightforward model that still captures the essential behavior. Therefore we chose the simplest techniques for the buckling analysis and the finite element code. These techniques seem adequate to draw the above conclusions, however they have shortcomings, some of them significant:

- Both the von Karman equations and the finite element code assume Hooke's law i.e. linear stress-strain relations. This assumption is appropriate for a first model but, in general, it is not very realistic for biological materials. Building a nonlinear model, however, is challenging because even more mechanical parameters are present that are hard to determine.

- The finite element code does not take geometric nonlinearities into account. This means that we treat all displacements as infinitesimally small. If the strain is about 10% this omission seems still justified, if it is more the assumption quickly becomes unrealistic.

- The von Karman equations use the assumption that no tangential body forces are applied. This is inconsistent with the use of tangential spring forces in the finite element code.

- In the buckling equations we assume that the resistance of the dermis underneath the basal layer and the resistance of the intermediate layer above the basal layer is given by a polynomial. This assumption is probably not realistic.
because the dermis and the intermediate layer have different spring constants. Therefore the resistance should be given by a more complex function. The same is true for the finite element code that also assumes equal normal spring constants for the dermis and the intermediate layer.

- The finite element code models the stress field by applying forces to the epidermis. This is a legitimate procedure, however, it would be better to obtain the stress field directly from the model assumptions (expanding sheet, resistance by boundary springs and stress due to geometry change). Implementing such a program would be a difficult but rewarding task.

- We assume that the stress field in the basal layer does not significantly change during the time of ridge spread over the fingertip which is certainly not the case. For example, the computations could take the change of the stress field due to buckling into account.

- The simulations of the von Karman equations lead to some effects that are not biological. An example are the jumps in the wave vector between patches of skin. It is conceivable that a model based on more realistic material properties gives better results in this respect.

### 7.3 Open questions and suggestions for further work

There are a number of interesting open questions that could be the focus of future work in this field.

- How do accidental patterns arise? BONNEVIE suggested that they form when two ridge anlagen are present, but that is the only information available. What would be the spatial relationship of the ridge anlagen to produce a certain accidental pattern? What role does the boundary play?
• Why are there so many more ulnar loops than radial loops? Referring to sym-
metries of the pad and the location of the ridge anlage seems to be part of an
answer but does not explain how these asymmetries come into being. Is there
an evolutionary advantage of ulnar loops?

• What is the role of the cushioned epidermis as described by Bonnevie? How
can it be modeled mathematically and why does cushioned epidermis on the
whole fingertip produces arches?

• How do certain aberrations in fingerprint patterns arise? Why are epidermal
ridges absent in a few individuals? How do dissociated patterns as described
by Abel form? What is the reason for the ridges-off-the-end syndrome?

• How do incipient ridges arise? They seem to be ridges that started to form but
"never quite made it". Why are they more frequent in certain individuals?

• How does the growth of the finger and the palm interact with the ridge forming
process?

• What stimulus initiates primary ridge formation in the first place? How are
cells in the upper layers of the epidermis recruited from the basal layer and how
does this process change once ridge formation starts?

• What process stops ridge formation and how is the ridge topology fixed such
that it does not change later on?

• What process governs the formation of sweat glands? A connection to primary
ridge development is certain, but how do these processes interact? What is the
genesis of secondary ridges? In what way is this process different from primary
ridge formation?

• Are the sources of stress the same in other mammal and marsupial species? Can
we identify other factors that may play a role in human fingerprint development?
• In what way do the toes differ from the fingers? The tip of the toes is broader; does this influence the kind of pattern that arises?

• The ridge pattern on the lower phalanges of the hand usually displays ridges that are more or less parallel to the flexion creases. However, frequently they are significantly curved and sometimes exhibit characteristic patterns. What causes these perturbations from straight parallel ridges?

• It has been argued in the past (by WHIPPLE) that fingerprint patterns have evolved from reptile scales. Is there a way to make the connection between these two (quite different) skin appendages mathematically?

• Similarly, there seems to be a connection between hair follicle and fingerprint formation. Both features are formed at about the same time, there are parallels in the formation process and they never both occur in the same patch of skin.

• How can the complicated patterns in certain monkey species that exhibit both ridge and hexagon features be modeled?

There are certain investigations that would greatly facilitate fingerprint modeling.

• Determination of basic material parameters of the basal layer (Young’s modulus, Poisson’s ratio), the dermis and the intermediate layers (spring constants). This would greatly help to validate the model or to discover shortcomings.

• Investigations about the nonlinear behavior of the material involved. These could lead to a more realistic model.

• Measurements of the geometry of the apical pad. Few such measurements exist. The only decent source so far are cross-sections through apical pads by BONNEVIE. Apart from that, only photographs are available, on which curvatures are hard to evaluate.
• Describing the geometry of the interdigital, thenar and hypothenar pads. Most information on these is only qualitative in the form of descriptions and photographs. More knowledge on the palmar pad's geometry is necessary for successful computer simulations of fingerprint formation on the palm.

• Photographs documenting the spread and development of the primary ridges from the time they first become visible to the time when the finger is fully covered. This way we can better understand the way minutiae form and use the results to build a more sophisticated model that adequately describes the postbuckling behavior of the basal layer.

• More embryological studies on mammals and marsupials. These could shed light on phenomena specific to certain species and put the fingerprint pattern development in humans in a broader perspective.

• The model presented in this work could be simulated — not just by computers — but by actually applying forces to a curved sheet of materials with varying elastic properties and studying the stress pattern and, if possible, the buckling behavior.
APPENDIX A

GEOMETRIC THEORY OF SHALLOW SHELLS

The purpose of this appendix is the derivation of the strain–displacement relations as they are needed for the derivation of the von Karman equations.

Let \( \Sigma \) be a surface in \( \mathbb{R}^3 \) which is given by the smooth parameterization \( r(x, y) : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3 \). Let

\[
N(x, y) = \frac{r_x \times r_y}{\|r_x \times r_y\|}
\]
denote the normal vector to the surface. Now consider a three-dimensional object that is given by the set of points \( r(x, y, z) = r(x, y) + zN(x, y) \) where \( (x, y) \in \Omega \) and \(-\frac{h}{2} \leq z \leq \frac{h}{2}\). Such an object is called a shell with constant thickness \( h \) with middle surface \( \Sigma \). We will always assume that \( h \) is small compared to the extension of \( \Sigma \). A shell is called shallow if its rise is small compared to its span. The expression \( \Sigma(z) \) denotes the surface \( r(x, y) = r(x, y) + zN(x, y) \) where \( z \) is fixed. These surfaces are called parallel surfaces. In the following discussion we will neglect terms of higher than linear order in \( h \) which is standard practice for all classical shell theories.

A.1 Classical differential geometry of surfaces

Distances and angles on the middle surface can be found using the first fundamental form

\[
I = dr \cdot dr = Edx^2 + 2F dx dy + G dy^2
\]

where

\[
E = r_x \cdot r_x, \quad F = r_x \cdot r_y, \quad G = r_y \cdot r_y.
\]

The second fundamental form measures how much the surface deviates locally from its tangent plane

\[
II = -dN = edx^2 + 2f dx dy + gdy^2
\]
where
\[ e = -r_{,x} \cdot N_x = r_{,xx} \cdot N, \quad f = -r_{,x} \cdot N_{,y} = r_{,xy} \cdot N, \quad g = -r_{,y} \cdot N_{,y} = r_{,yy} \cdot N. \]

Of special interest are the second order derivatives of \( r \). It can be shown [54] that they are given by the following expressions
\[
\begin{align*}
    r_{,xx} &= \Gamma_1^{11} r_{,x} + \Gamma_1^{22} r_{,y} + eN \\
    r_{,xy} &= \Gamma_1^{12} r_{,x} + \Gamma_2^{22} r_{,y} + fN \\
    r_{,yy} &= \Gamma_2^{12} r_{,x} + \Gamma_2^{22} r_{,y} + gN
\end{align*}
\]
(A.1)

where \( \Gamma_{jk}^i \) are the Christoffel symbols and can be determined as follows
\[
\begin{pmatrix}
    \Gamma_1^{11} \\
    \Gamma_1^{12} \\
    \Gamma_2^{22} \\
    \Gamma_2^{11} \\
    \Gamma_2^{12} \\
    \Gamma_2^{22}
\end{pmatrix} = \frac{1}{EG - F^2}
\begin{pmatrix}
    \frac{1}{2}E_xG - FF_x - \frac{1}{2}F E_y \\
    \frac{1}{2}E_yG - \frac{1}{2}F G_x \\
    -\frac{1}{2}FG_y + GF_y - \frac{1}{2}GG_x \\
    -\frac{1}{2}G_xF + EF_x - \frac{1}{2}EE_y \\
    \frac{1}{2}EG_x - \frac{1}{2}F E_y \\
    -FF_y + \frac{1}{2}FG_x + \frac{1}{2}EG_y
\end{pmatrix}
\]

For shallow shells it is possible to simplify these relations significantly. Because their rise is small compared to their span we may treat the middle surface as the graph of a function \( \phi(x, y) \) with small derivatives \( \phi_x \) and \( \phi_y \). Then we obtain the following approximations for the coefficients of the fundamental forms:
\[
\begin{align*}
    E &= 1 + \phi_x^2 \approx 1 \\
    F &= \phi_x \phi_y \approx 0 \\
    G &= 1 + \phi_y^2 \approx 1. \\
    e &= \phi_{,xx} \\
    f &= \phi_{,xy} \\
    g &= \phi_{,yy}.
\end{align*}
\]

Here we neglect nonlinear terms in the slopes of \( \phi \). Therefore, we see that a shallow shell is effectively treated as a flat surface (after all the metric is just like the one of a plane) but on the other hand we study the effects of deformation on the second
fundamental form that retains curvature effects. With these assumptions all the Christoffel symbols vanish. If we additionally assume that \( x \) and \( y \) are coordinates of largest and smallest curvature we have \( f = 0 \). Furthermore, the remaining coefficients of the second fundamental form can now be given in terms of the principal radii of curvature:

\[
e = -\frac{1}{R_x}, \quad g = -\frac{1}{R_y}.
\]

We will now derive formulas for the metric on the parallel surfaces \( \Sigma(z) \) of a shallow shell. We wish to compute the metric

\[
I(z) = E(z)dx^2 + 2F(z)dx\,dy + G(z)dy^2.
\]

By definition we have

\[
E(z) = (r + zN)_x \cdot (r + zN)_x = r_{xx} + 2zr_{xx} \cdot N_x + z^2N_x \cdot N_x \approx 1 + 2ze.
\]

Similarly,

\[
G(z) \approx 1 + 2zg.
\]

Further we find

\[
F(z) = (r + zN)_x \cdot (r + zN)_y = r_{xy} + z(r_{xx} \cdot N_y + r_{xy} \cdot N_x + z^2N_x \cdot N_x) \approx 0.
\]

We obtain

\[
I(z) = (1 + 2ze)dx^2 + (1 + 2zg)dy^2.
\]

A.2 Deformation of a shallow shell

We consider the deformation of the middle surface along the local unit tangent vectors and the normal vector as follows

\[
r' = r + \Delta = r + ut_x + vt_y + wN
\]
where $t_x = \frac{r_x}{\|r_x\|}$ and $t_y = \frac{r_y}{\|r_y\|}$. Then the deformed shell is represented by the points $r'+zN'$ where again $-\frac{h}{2} \leq z \leq \frac{h}{2}$. After deformation the first fundamental form will have changed. Let us write it as follows

$$I'(z) = E(z)(1 + \epsilon_z(z))^2dx^2 + \epsilon_{xy}(z)dx dy + G(1 + \epsilon_y(z))^2dy^2.$$ 

where the quantities $\epsilon_z$, $\epsilon_{xy}$ and $\epsilon_y$ are called the strains. We again consider shallow shells. Note that in this case $\|r_x\| = 1$ and $\|r_y\| = 1$ on the middle surface. At first we find the strains on the middle surface, that is $\epsilon_z(0)$, $\epsilon_{xy}(0)$ and $\epsilon_y(0)$. We have

$$r'_x = r_x + u_x t_x + v_x t_y + (w_x + u_e)N$$

$$r'_y = r_y + u_y t_x + v_y t_y + (w_y + v_e)N.$$ 

where (see A.1 and [54])

$$t_{x,x} = r_{xx} = \epsilon N$$

$$t_{x,y} = r_{xy} = 0$$

$$N_{xx} = -\epsilon t_x.$$ 

This gives us

$$r'_x = (1 + u_x - we)t_x + v_x t_y + (w_x + u_e)N$$

$$r'_y = (1 + u_y - wg)t_y + u_y t_x + (w_y + v_g)N.$$ 

Now we can find the coefficients of the first fundamental form

$$E' = r'_x \cdot r'_x = (1 + u_x - we)^2 + v_x^2 + (w_x + u_e)^2$$

$$F' = r'_x \cdot r'_y = (1 + u_x - we)u_y + (1 + v_y - wg)v_x + (w_x + u_e)(w_y + v_g)$$

$$G' = r'_y \cdot r'_y = (1 + v_y - wg)^2 + v_y^2 + (w_y + v_g)^2.$$ 

As usual we retain linear terms and neglect nonlinear ones, however, this time there is an exception. Because we want to use the strains for the analysis of buckling we anticipate that the slopes $w_x$ and $w_y$ will be large enough that their squares cannot be neglected. Therefore we arrive at the following approximation

$$E' \approx 1 + 2u_x - 2we + w_x^2$$

$$F' \approx u_y + v_x + w_x w_y$$

$$G' \approx 1 + 2v_y - 2wg + w_y^2.$$
Using the relation $\sqrt{1+x} = 1 + \frac{x}{2} + O(x^2)$ we now arrive at an approximation for the strains of the middle surface:

\[
\begin{align*}
\varepsilon_x(0) &= u_x - we - \frac{w_y^2}{2} \\
\varepsilon_{xy}(0) &= u_y + v_x + w_xw_y \\
\varepsilon_y(0) &= v_y - wg + \frac{w_x^2}{2}.
\end{align*}
\]

To find the displacement $\Delta(z)$ of a point away from the middle surface we use Kirchhoff’s hypothesis. According to the hypothesis lines normal to the middle surface are again normal after the deformation. That means we have

\[
\begin{align*}
r(z) &= r(0) + zN, \\
r'(z) &= r'(0) + zN'.
\end{align*}
\]

Subtracting these equations gives us

\[
\Delta(z) = r'(z) - r(z) = r'(0) - r(0) + z(N' - N) = \Delta(0) + z(N' - N).
\]

We now find $N'$ according to

\[
N' = r'_{,x} \times r'_{,y} = (-ue - w_x)t_x + (-vg - w_y)t_y + (1 + u_x - we + v_y - wg)N.
\]

In most shell theory quantities of order $h : R$ are neglected. Therefore we drop the terms $zue$ and $zvg$. Thus we obtain, neglecting more terms of order $h : R$

\[
r'(z) = r'(0) + zN' \\
= r'(0) + z(-w_xt_x - w_yt_y + (1 + u_x - we + v_y - wg)N).
\]

This gives us, neglecting nonlinear terms,

\[
\begin{align*}
r'_{,x}(z) &= r'_{,x}(0) + z(-w_xt_x - w_yt_y + (1 + u_x - we + v_y - wg)_xN) \\
&= r'_{,x}(0) + z(-w_yt_x - w_xt_y + (1 + u_x - we + v_y - wg)_yN).
\end{align*}
\]

Now we can find the coefficients of the first fundamental form, again neglecting nonlinear terms,

\[
\begin{align*}
E'(z) &= E'(0) - 2zw_{xx} \\
F'(z) &= F'(0) - 2zw_{xy} \\
G'(z) &= G'(0) - 2zw_{yy}.
\end{align*}
\]
This finally gives us the desired expression for the strain–displacement relations on the parallel surfaces

\[
\epsilon_x(z) = \epsilon_x(0) - zw_{xx} \\
\epsilon_{xy}(z) = \epsilon_{xy}(0) - 2zw_{xy} \\
\epsilon_y(z) = \epsilon_y(0) - zw_{yy} .
\]
APPENDIX B

NUMERICAL ALGORITHMS

B.1 Finite elements

The method of finite elements has proven to be the method of choice for problems of complex geometries. Because the nontrivial geometry of the finger is essential for fingerprint development finite elements were used to determine the stress field that is formed as the result of certain growth forces.

There are several different shell elements to choose from. See [52, 107, 108] for an overview. We chose the theory described in chapter 8 in [108] and chapter 6 in [52]. This approach treats shells as three-dimensional body where one dimension (the thickness) is much smaller than the others. The approach does not assume Kirchhoff's hypothesis (normals remain normal after deformation) but the more realistic Reissner-Mindlin assumption (normals remain straight lines but are not necessarily normal anymore) that takes transverse shear phenomena into account and gives good results, even for moderately thick shells.

Throughout the calculation we use simple, 4-node, bilinear, quadratic elements. The program could easily be extended to higher order elements. Indeed, the internal structure already supports 8-node serendipity elements and 9-node Lagrange elements. However, not all features support these elements and it appears that, for our purposes, no significant improvement of the results is achieved.

In the following, we will outline the most important concepts of the calculations and refer to the literature for details. Let us consider a single quadrilateral face and introduce local coordinates $(\xi, \eta, \zeta)$ that vary from $-1$ to $1$ over the face (see Figure
Using these local coordinates, the face can be parameterized by

\[
\begin{pmatrix}
x \\
y \\
z
\end{pmatrix} = \sum_{i=1}^{4} N_i(\xi, \eta) \begin{pmatrix}
x_i \\
y_i \\
z_i
\end{pmatrix} + \frac{1}{2} \zeta h v_{3i}^{\text{mid}}.
\]

Here the index \(i\) refers to each corner point of the face. \(N_i\) is a bilinear function that is 1 at node \(i\) but vanishes on the other nodes. The index \(\text{mid}\) refers to the midplane of the shell and \(v_{3i}\) is a unit vector perpendicular to the shell at node \(i\).

The displacements in the global directions are now assumed to have the following form

\[
\begin{pmatrix}
u \\
v \\
w
\end{pmatrix} = \sum_{i=1}^{4} N_i(\xi, \eta) \begin{pmatrix}
u_i \\
v_i \\
w_i
\end{pmatrix} + \frac{1}{2} \zeta h [v_{1i}, -v_{2i}] \begin{pmatrix}
\alpha_i \\
\beta_i
\end{pmatrix}.
\]

The vectors \(v_{1i}\) and \(v_{2i}\) are vectors tangential to the midsurface. Hence, the displacements in a face can be described by the displacement at the node point, \((u_i, v_i, w_i)\), and the rotations around \(v_{1i}\) and \(v_{2i}\), \((\alpha_i, \beta_i)\), which add up to 5 degrees of freedom.

Let us denote the displacement at any point in the element by \(\mathbf{u}\) and the degrees of freedom per element by \(\alpha^e\). Here we have

\[
\alpha^e = \begin{pmatrix}
\alpha_1^e \\
\vdots \\
\alpha_4^e
\end{pmatrix} \quad \text{with} \quad \alpha_i^e = \begin{pmatrix}
u_i \\
v_i \\
w_i \\
\alpha_i \\
\beta_i
\end{pmatrix}.
\]
Then we have the relation

\[ u = Na^e \]

where \( N = [N_1, \ldots, N_4] \) are functions of position.

As the next step we find an expression for the strains. This is most easily accomplished for a local coordinate system using \((v_1, v_2, v_3)\) as the coordinate axes. We denote quantities in this coordinate system with a bar. As usual, we do not consider the normal strain, \( \bar{\epsilon}_z \), but retain all shear strains; that means we have \( \bar{\epsilon} = (\bar{\epsilon}_x, \bar{\epsilon}_y, \bar{\epsilon}_{xy}, \bar{\epsilon}_{yz}, \bar{\epsilon}_{xz}) \). The evaluation of \( \bar{\epsilon} \) can be accomplished by using a linear differentiation operator that determines the strains from the displacements as follows

\[
\begin{bmatrix}
\bar{u}_x \\
\bar{v}_y \\
\bar{u}_y + \bar{u}_z \\
\bar{v}_z + \bar{u}_v \\
\bar{w}_z + \bar{u}_z
\end{bmatrix}
= S
\begin{bmatrix}
\bar{u} \\
\bar{v} \\
\bar{w}
\end{bmatrix}
\]

where subscripts denote derivatives. Therefore we obtain the strains simply as

\[ \epsilon = SNa^e = B a^e. \]

Because we use linear material behavior we can obtain the stresses by multiplying the strains with a matrix \( D \) as follows:

\[ \sigma = D(\epsilon - \epsilon_0). \]

Here \( \epsilon_0 \) is called initial strain and can be used to model expansion effects.

Using either the principle of virtual work or the minimization of elastic energy one arrives at the following equation for the unknown \( a^e \)

\[ K^e a^e = f^e \]

where

\[ K^e = \int_{V^e} B^T D B \, dV \]
is called the stiffness matrix and

\[ f^e = b^e + q^e + \int_{V^e} B^T D e_0 \, dV. \]

Here \( b^e \) are body forces acting on each node element and \( q^e \) are internal node forces.

Such systems of equations can be found for each element. To solve the global problem they can be all added up and yield the system of equation

\[ K a = f \]

where \( a \) is the vector of all degrees of freedoms,

\[ [K_{ij}] = \sum_e [K^e_{ij}] \]

and

\[ f = b + \int_V B^T D e_0 \, dV \]

\[ [f_i] = \sum_e [f^e_i]. \]

Note that the internal node forces \( q^e \) drop out after summing up all element contributions.

Through a series of coordinate transformations the integrals necessary to compute \( K^e \) and \( f^e \) are considered in terms of the element coordinates \((\xi, \eta, \zeta)\). The integration is performed numerically by a Gaussian quadrature. To maintain the rate of convergence one would expect that \( 2 \times 2 \) quadrature (i.e. the quadrature uses the function values at 4 points inside the element) gives acceptable results. However, this is not the case due to so-called locking phenomena [107]. The obtained displacement are orders of magnitude too small and therefore worthless.

This problem has been avoided by only integrating over one point \((1 \times 1\) integration), which paradoxically often gives much better results. This procedure is called underintegration. Unfortunately, it is often plagued by the occurrence of zero energy
modes that destroy the solution. Usually underintegration is used only for certain terms which avoids both locking and the zero energy modes.

Boundary conditions can be imposed by modifying the stiffness matrix so that certain values are enforced for boundary displacements. These displacements are prescribed in terms of the local coordinate system. An alternative way of ensuring that the stiffness matrix is nonsingular is the attachment of spring forces at some or every node. This is easily achieved by increasing the diagonal elements of the stiffness matrix.

It is well-known that the approximation of the strains and stresses is relatively poor at the nodes itself and much better results can be achieved by using certain interior points. To avoid this problem the SPR (Superconvergent Patch Recovery) method described in [107] was used. Here the strains are determined at the interior points and the node stresses are recovered by a least square approximation.

**B.2 The spectral scheme**

For the solution of the von Karman equations a simple spectral method was used. We start with equations 4.21 without the curvature terms

\[
\kappa w_t + D \nabla^4 w - N_x w_{xx} - N_y w_{yy} - 2N_{xy} w_{xy} - [f, w] + \gamma w + \alpha w^2 + \beta w^3 = 0
\]

\[
\frac{1}{Eh} \nabla^4 f + \frac{1}{2} [w, w] = 0. \tag{B.1}
\]

Let us denote

\[
\lambda(f, w) = -N_x w_{xx} - N_y w_{yy} - 2N_{xy} w_{xy} - [f, w] + \alpha w^2 + \beta w^3.
\]

This expression contains all the nonlinear and position dependent terms for which the Fourier transform cannot easily be found. For all the other components, derivatives in Fourier space can easily be found using multiplication by powers of the wave vector. The wave numbers with smallest modulus should be used for this purpose (see [88]).
Let us denote the Fourier transform by a hat and let \( k = (l, m) \) and \( k^2 = k^2 \). Then we have
\[
\kappa \hat{w}_t + Dk^4 \hat{w} + \lambda (\hat{f}, \hat{w}) + \gamma \hat{w} = 0
\]
\[
\frac{k^4}{Eh} \hat{f} + \frac{1}{2} [\hat{w}, \hat{w}] = 0.
\]
Let us assume that we have obtained an approximation of \( \hat{w} \) and \( \hat{f} \) at step \( n \) which we call \( \hat{w}_n \) and \( \hat{f}_n \), respectively. Then \( \hat{w} \) is updated using the first equation according to
\[
\hat{w}_{n+1} = \frac{\kappa \hat{w}_n - \Delta t \lambda (\hat{f}_n, \hat{w}_n)}{\kappa + \Delta t (Dk^4 + \gamma)}.
\]
Then we can update the second derivatives of \( f \) using the second equation
\[
(\hat{f}_{xx})_{n+1} = \frac{Ehl^2}{2k^4} [\hat{w}_{n+1}, \hat{w}_{n+1}]
\]
\[
(\hat{f}_{yy})_{n+1} = \frac{Ehm^2}{2k^4} [\hat{w}_{n+1}, \hat{w}_{n+1}]
\]
\[
(\hat{f}_{xy})_{n+1} = \frac{Ehlm}{2k^4} [\hat{w}_{n+1}, \hat{w}_{n+1}].
\]
Note that this procedure implies that the derivatives of \( f \) vanish if \( [w, w] = 0 \). This scheme is easily implemented, especially using software packages that provide a fast Fourier scheme. In our case FFTW ("the fastest Fourier transform in the west") was used. As the initial conditions random values with small amplitudes were used. Also, after every step a small random disturbance is added in order to prevent the extinction of the solution in subcritical domains.

The scheme has the disadvantage that, due to its explicit nature, instabilities can occur if the resolution is not small enough and we are high above the critical buckling stress.

### B.3 shells.cpp

The program shells.cpp was written to explore the effects of certain body forces, spring forces and expansion rates on the stress distribution in a curved surface. Then the stress field obtained is used to simulate the buckling process.
Different geometries can be used. They are provided in separate files that define all the necessary quantities. Each of these files provides lists of elements and nodes, the local coordinate systems and the forces acting on the nodes. The following files currently exist

- `bump.cpp` – Gaussian shaped bump.
- `cylinder.cpp` – cylinder geometry, useful for testing purposes.
- `finger.cpp` – flat version of a finger geometry, now obsolete.
- `finger3d.cpp` – curved version of a finger geometry.
- `plane.cpp` – planar rectangle.

The only nontrivial geometry is `finger3d`. It is obtained by considering a domain in spherical coordinates (the fingertip) above a domain in cylindrical coordinates. We define the surface as a map from a rectangular region $\Omega$. Let $(u, v) \in \Omega$. Then the surface $S$ is defined as the image of the map $r : \Omega \to S$. The map is defined for $v < 0$ as

$$
\begin{align*}
    r(u, v) &= \left( \begin{array}{c}
        \rho(u, v) \cos(u/r_0) \\
        v \\
        \rho(u, v) \sin(u/r_0)
    \end{array} \right)
\end{align*}
$$

and for $v > 0$ as

$$
\begin{align*}
    r(u, v) &= \left( \begin{array}{c}
        \rho(u, v) \cos(u/r_0) \cos(v/r_0) \\
        \rho(u, v) \sin(v/r_0) \\
        \rho(u, v) \sin(u/r_0) \cos(v/r_0)
    \end{array} \right).
\end{align*}
$$

Here $r_0$ is a parameter that should be chosen close to the average of the radius of pad curvature. Further we have a function $\rho(u, v)$ that denotes the distance of the surface from the $y$-axis in the cylinder part and the distance from the origin in the spherical part of the surface. This construction ensures a smooth (at least once differentiable) surface for all smooth $\rho(u, v)$. 
The pad geometry can now be specified by choosing a certain function \( \rho(u, v) \). In our simulations it is given by the following expression

\[
\rho(u, v) = \left[ c_1 + \left( c_2 e^{-\frac{(v-c_9)^2}{c_4}} + c_5 \right) \sin(u/\tau_0) + \left( c_6 e^{-\frac{(v-c_9)^2}{c_8}} + c_9 \right) \sin(2u/\tau_0) \\
+ \left( c_{10} e^{-\frac{(v-c_9)^2}{c_{12}}} + c_{13} \right) \sin(3u/\tau_0) \\
+ \frac{1}{\sqrt{\frac{\cos^2(v/\tau_0)}{c_2^2} + \frac{\sin^2(v/\tau_0)}{c_4^2}}} \cdot (1 - \theta) \right] \cdot \theta
\]

where

\[
\theta = \frac{\tanh(-c_{15}(v - c_{18})) + 1}{2}.
\]

Along a horizontal cross-section the function is defined by the first four terms of a sine series in the cylinder part. Along the length of the cylinder, the coefficients of the sine series are varied according to a Gaussian. In the spherical part, a smooth transition into an ellipsoid is achieved. Using this paradigm many different shapes can be achieved, although the many parameters (16!) are certainly a drawback of the method.

Using this description, local coordinate systems (required for the finite element code) can be found easily. In the cylindrical part an orthonormal set of vectors (one normal to the surface \( n \), two tangential to the surface \( t_1 \) and \( t_2 \)) is found as follows

\[
\begin{align*}
n &= \frac{r_u \times r_v}{\|r_u \times r_v\|}, \\
t_1 &= \frac{r_u}{\|r_u\|}, \\
t_2 &= n \times t_1.
\end{align*}
\]

The spherical part is trickier, here we use the construction

\[
\begin{align*}
n &= \frac{r_u \times r_v}{\|r_u \times r_v\|}, \\
t_1 &= n \times \left( \cos(v/\tau_0)i - \sin(v/\tau_0)j \right), \\
t_2 &= n \times t_1.
\end{align*}
\]

This construction ensures that the local coordinate systems do not change rapidly, not even in the vicinity of the north pole in the spherical part.

This construction is also useful for projecting the surface into the plane, we simply use the inverse map \( r^{-1} : S \rightarrow \Omega \) for that purpose. Furthermore, the stress field on
the surface is imposed on the plane by identifying $t_1$ with the $x$-direction and $t_2$ with the $y$-direction.

The program provides different means by which stress can be applied to the shell. Most forces are specified as some kind of densities to ensure comparable results when the resolution is changed. The following variables are most important for specifying the problem

- $nodecoord$ – Matrix containing the coordinated of all nodes
- $enodes$ – Matrix containing the node numbers for each element
- $bndnodes$ – vector containing of all boundary nodes.
- $constraint$ – vector of constraints for the boundary nodes. The specified numbers are interpreted in their binary expansion.
- $load$ – matrix containing the load density for each element in the three local directions.
- $bodyforce$ – matrix containing forces for each node in the three local directions.
- $expansion$ – matrix containing the expansion rates for each nodes in the two local tangential directions.
- $normspring$ – scalar defining the spring strength in the normal direction.
- $tangspring$ – vector defining for each node the spring strength in the tangential directions.
- $specspring$ – matrix defining springs at certain points. Each row defines the node number, the spring direction in local coordinates and the spring strength.
- $expspring$ – vector that specifies for each boundary node springs that resist expansion. It is given as a density along the boundary.
- `expoffset` - vector that modifies the spring direction of `expspring`.

- `bndforce` - vector that defines forces acting on the boundary, it is given as a density.

- `springcorrect` - boolean, if true the springs are attached not directly at the node but at the place the node would have been if expansion had not been resisted, essential for obtaining meaningful pictures if expansion effects are used.

### B.4 Tests

The fairly complex finite element code was tested using examples with known analytical solution.

1. **Plate Bending**

   We consider a quadratical plate with side length $L = 10$ m and thickness $h = 0.1$ m which is subjected to a uniform load of $q = 1$ N/m². We choose $E = 10^6$ N/m² and $\mu = 0.4$. In Table B.1 we represent our numerical results for the deflection $w_c$ in the center of the plate for different grids. $N$ denotes the number of elements in $x$ and $y$ direction. These results are contrasted with the exact solution for thin plates

   $$w_c = \frac{12\zeta(1 - \mu^2)qL^4}{Eh^3}$$

   where $\zeta = 0.00127$ for the clamped case and $\zeta = 0.00406$ for the simply supported case (see [108]). The agreement between numerics and theory is strong with errors less than 1%.

2. **Beam Bending**

   We consider the classic problem of a slender cantilever (see Figure B.2) with rectangular cross-section that is bent by a force $P$ at its end. The deflection in $y$-direction is denoted by $v$. The following formulas for displacements and stresses are well-known
TABLE B.1. Computed and exact thin plate solution for the center deflection of a uniformly loaded plate.

<table>
<thead>
<tr>
<th>$N$</th>
<th>clamped case</th>
<th>simply supported case</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.1056</td>
<td>0.3682</td>
</tr>
<tr>
<td>10</td>
<td>0.1270</td>
<td>0.4139</td>
</tr>
<tr>
<td>20</td>
<td>0.1276</td>
<td>0.4125</td>
</tr>
<tr>
<td>30</td>
<td>0.1277</td>
<td>0.4126</td>
</tr>
<tr>
<td>exact</td>
<td>0.1280</td>
<td>0.4092</td>
</tr>
</tbody>
</table>

**FIGURE B.2.** A cantilever of length $L$ with rectangular cross-section with dimensions $b$ and $c$ is bent by a force $P$ at its end.

(see [14]):

$$v(L, y) = \frac{PL^3}{3EI}$$

$$\sigma_x(x, y) = -\frac{P(L - x)y}{I}$$

$$\sigma_{xy}(x, y) = \frac{3P}{2bc} \left(1 - \frac{4y^2}{b^2}\right).$$

Here $I = \frac{b^3c}{12}$ is the moment of inertia.

For our tests we chose $L = 10$ m, $b = 1$ m, $c = 0.1$ m, $E = 10^6$ N/m$^2$ and $P = 10$ N. The calculation results are given in Table B.2. The agreement between theory and numerics is quite good (errors are less than 4%) for the finer grids. The coarser grids do not produce as good results which simply points out that our shell elements are not very efficient for beam calculations.
TABLE B.2. Computed and exact solutions for deflections and stresses of a loaded cantilever.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$v(L,0)$</th>
<th>$\sigma_x(L/2,0)$</th>
<th>$\sigma_{xy}(L/2,0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.1636</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>10</td>
<td>0.2935</td>
<td>2208</td>
<td>135.2</td>
</tr>
<tr>
<td>20</td>
<td>0.3675</td>
<td>2753</td>
<td>145.4</td>
</tr>
<tr>
<td>30</td>
<td>0.3859</td>
<td>2885</td>
<td>147.9</td>
</tr>
<tr>
<td>exact</td>
<td>0.4000</td>
<td>3000</td>
<td>150.0</td>
</tr>
</tbody>
</table>

TABLE B.3. Computed and exact solutions for deflections and stresses of a loaded cylinder.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$w$</th>
<th>$\sigma_\theta$</th>
<th>$\sigma_{nx}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-0.001806</td>
<td>-3451</td>
<td>355.8</td>
</tr>
<tr>
<td>10</td>
<td>-0.002133</td>
<td>-4416</td>
<td>446.1</td>
</tr>
<tr>
<td>20</td>
<td>-0.002231</td>
<td>-4606</td>
<td>487.7</td>
</tr>
<tr>
<td>30</td>
<td>-0.002250</td>
<td>-4588</td>
<td>495.5</td>
</tr>
<tr>
<td>exact</td>
<td>-0.002227</td>
<td>-4455</td>
<td>500</td>
</tr>
</tbody>
</table>

3. Cylinder Compression

We consider a cylinder of radius $R$ with $x$ being the axial coordinate, $\theta$ the circumferential coordinate and $n$ the normal coordinate. We consider a line load of strength $P$ on the curve $x = 0$. If the cylinder is large enough then we have the following formulas for deflections $w$ and stresses $\sigma_\theta$, $\sigma_{nx}$ at $x = 0$:

$$w = -\frac{P}{8k^3D},$$

$$\sigma_\theta = \frac{P}{2h},$$

$$\sigma_{nx} = -\frac{PRk}{2h},$$

where

$$k^4 = \frac{3(1-\mu^2)}{R^2h^2}$$

and $R$ is the radius of the cylinder (see [39] for details).

We used $\mu = 0.45$, $E = 10^6$, $R = 5\text{ m}$, $h = 0.1\text{ m}$ and $P = 10\text{ N}$. $N$ denotes the number of elements in $x$ and $\theta$ direction. The calculation results are shown in
Table B.3. There is very good agreement between numerics and theory with errors less than 3%.

B.5 Visualization

The visualization of the 3d data was an important and crucial task. It was accomplished using TRIVIZ (triangular visualization) written by Heiko Gerdes. It is based on OpenGL libraries. Using a C++ interface, graphics can be produced by adding points, edges, triangles and quadrilaterals to a data structure which is then stored as a *.tri file. Different perspectives and lighting effects can also be used.

The *.tri files can be viewed by TRIVIZ. Using the mouse the graphic can be rotated in all directions. The function values of the different quantities on the shell are represented by a color or grey scale. A legend is attached at the side of the graphic.

TRIVIZ has not only been essential for displaying the results but also for debugging purposes. During a program run of shells.cpp the following files are produced.

- *mesh.tri* – displays the finite element mesh.
- *bndforce.tri* – displays the direction of *bndforce*.
- *bodyforceu.tri, bodyforcev.tri, bodyforcen.tri* – shows the three components of *bodyforce*.
- *expansionx.tri, expansiony.tri* – the expansion rates in the two tangential directions.
- *expoffset.tri* – shows *expoffset*.
- *expspring.tri* – displays *expspring*.
- *forcedir.tri* – displays the direction of all forces added up.
• \textit{forceu.tri, forcev.tri, forcen.tri} – shows the strength of all forces added up in the local directions.

• \textit{loadu.tri, loadv.tri, loadn.tri} – the components of load in the local directions.

• \textit{maxstress.tri} – displays the direction of the smallest stress and displays the magnitude of the largest stress. Essential for predicting the buckling pattern.

• \textit{udisp.tri, vdisp.tri, ndisp.tri} – the local displacements.

• \textit{xdisp.tri, ydisp.tri, zdisp.tri} – the global displacements.

• \textit{rotu.tri, rotv.tri} – the rotations along the \(u\) and \(v\) axis.

• \textit{strainu.tri, strainv.tri, strainuv.tri, strainun.tri, strainvn.tri} – the components of the strain tensor.

• \textit{stressu.tri, stressv.tri, stressuv.tri, stressun.tri, stressvn.tri} – the components of the stress tensor.

• \textit{v1.tri, v2.tri, v3.tri} – the directions of the local coordinate system.

• \textit{stressratio.tri} – the ratio of the principal stresses.

The results of the von Karman equations do not need to be visualized using TRIVIZ because the data is just two-dimensional. The variables \(f\) and \(w\) and their second derivative are simply normalized and stored in a JPEG file.
REFERENCES


