Preface

Let D be a positive non-square integer. The misnamed Pell equation is an expression of the form

$$T^2 - DU^2 = 1 , \qquad (0.1)$$

where T and U are constrained to be integers. For example, if $D = 13$, then $T = 649$ and $U = 180$ is a solution of (0.1). This very simple Diophantine equation seems to have been known to mathematicians for over 2000 years. Indeed, there is very strong evidence that it was known to Archimedes, as the Cattle Problem, attributed to him in antiquity, makes very clear. Even today, research involving this equation continues to be very active; at least 150 articles dealing with it in various contexts have appeared within the last decade. One of the main reasons for this interest is that the equation has a habit of popping up in a variety of surprising settings; it is also of great importance in solving the general second-degree Diophantine equation in two unknowns:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 .$$

Furthermore, the problem of solving (0.1) is connected to that of determining the regulator, an important invariant of a real quadratic number field, and to solving the discrete logarithm problem in such structures. Today, this latter problem is of interest to cryptographers.

Such is the interest in the Pell equation that at least three books have been devoted to it:

- H. Konen, *Geschichte der Gleichung $t^2 - Du^2 = 1$*, Leipzig, 1901.
- E. E. Whitford, *The Pell Equation*, College of the City of New York, New York, 1912
- Edward J. Barbeau, *Pell's Equation*, Springer, 2003

The first two have been out of print for a long time and are currently very difficult to find. Also, because they are quite old, they do not deal with the modern theory of the equation. Much has been learned since 1912. The last book, according to its author, "is a focused exercise book in algebra" and

is intended to motivate college students to develop an appreciation of mathematical technique. As such, it succeeds very well, but there is no attempt in the book to explore the deeper aspects of this equation, nor was that its author's intent.

It is well known that for any positive non-square integer D, (0.1) has an infinitude of solutions, which can be easily expressed in terms of the *fundamental solution* t, u, where $t, u > 0$. For example, the fundamental solution of (0.1) for $D = 7$ is $t = 8$ and $u = 3$. If $D = 1620$, then $t = 161$, but if $D = 1621$, then t is a number of 76 digits! As we will see in Chapter 13, there are even more extreme examples of this phenomenon for larger values of D. It is this puzzle of finding the fundamental solution that we refer to as the problem of solving the Pell equation. It was very likely investigated by the ancients, but it was not until the early 7th century AD that the Indian mathematician Brahmagupta discovered an ad hoc method of solving this problem. Unfortunately, his method and its more deterministic successors, which make use of the theory of continued fractions, cannot conveniently be used when D becomes large, say in excess of 15 digits.

The purpose of this book is to provide a comprehensive discussion of how to find the fundamental solution and, in particular, to describe methods for doing this that have been developed since 1972 for large values of D. As much of this material is scattered rather widely throughout the literature, this will be the first book to discuss this subject in any detail. The principal component of our enquiry will be computational techniques, but in order to derive these, it will be necessary to develop the required theory. In doing this, we will explore a great variety of different topics in number theory, some indication of which may be found by examining the table of contents.

As our approach to the Pell equation is largely computational, we assume that the reader is at least vaguely familiar with the basic precepts of measuring the computational complexity of an algorithm. We will use the terms "complexity," "time complexity," and just plain "runtime" or "time" interchangeably to describe the efficiency (the number of bit operations needed) by a particular computational technique. Thus, for example, we might say that a particular algorithm executes in time complexity $O(f(n))$, where f is some function and n is the input length, or we might say that it completes its computation in time $O(f(n))$. We may also have to measure the maximum number of bits required by an algorithm in order for it to execute. We call this the *space* or *space complexity* of the algorithm.

In order to solve (0.1) for large D, it has been discovered that it is easier first to evaluate the regulator R of the associated real quadratic number field $\mathbb{Q}(\sqrt{D})$. Nevertheless, the problem of computing R can still be very difficult, particularly when the value of the *radicand* D becomes very large $(> 10^{25})$. (Clearly, the actual value of the regulator can never be computed because it is a transcendental number; we are content to produce a rational number (often an integer) R' which is within 1 of the actual value.) The best method currently available for computing R' is Buchmann's subexponential method.

Unfortunately, the correctness of the value of R' produced by this technique is conditional on a generalized Riemann hypothesis, for which there is as yet no proof. The best unconditional algorithm (the value of R' is unconditional, not the running time) for computing the regulator of a real quadratic field is Lenstra's $O(D^{1/5+\epsilon})$ Las Vegas algorithm.

In this book we will discuss all of the above techniques and ultimately describe a rigorous method for verifying the regulator produced by the subexponential algorithm. This technique is of complexity $O(D^{1/6+\epsilon})$ and is unconditional, once we have a candidate for R'. It has been used to verify a 33-digit R' for a field with a 65-digit value of D. In addition, these methods can be extended to the problem of determining rigorously for real quadratic fields of large radicand whether or not a given ideal is principal. This, as we will point out, is of great importance in solving certain Diophantine equations. We will also describe some rather surprising applications of this material to cryptography.

Most of these techniques rely on estimations of certain irrational quantities; thus, in order to establish our results rigorously, it is essential that we have provable upper bounds on the errors that result from our use of these approximations. We provide a complete discussion of this and the associated algorithms, but, unfortunately, certain aspects of this are necessarily very technical and, frankly, rather wearisome. In order to facilitate a relatively smooth flow of this material, we have relegated the greater portion of the more tedious minutiae required by this investigation to an appendix.

We use different modes of presentation of the algorithms discussed in this book. The most formal of these include a name, such as NUCOMP or WNEAR, and a detailed listing of the pseudocode for the algorithm. This is usually provided for the basic algorithms, which are frequently employed in the latter part of the book. Several of these can be found in the aforementioned appendix. Some of the other algorithms, which in this formal format would be far too long, are described in pseudocode, which is less detailed. This is particularly the case for the index-calculus techniques described in Chapter 13. Finally, we sometimes simply describe certain processes rather informally as a simple sequence of steps which involve the use of the more formally presented algorithms that have been described previously. For example, this is the case for the technique of rigorously verifying the value of R' mentioned in Chapter 15.

We wish to emphasize here that this book is not intended to be used as a textbook; its focus is much too narrow, and although we do include a number of examples, we provide no exercises. It could, however, be used as supplementary reading for students enrolled in a second course in number theory. The intended primary audience is number theorists, both professional and amateur, and students, but as we discuss a number of cryptographic applications of the material that we develop in the book, a possible secondary readership would be that of mathematical cryptographers at about the same level as the primary readership. The subject matter should be accessible to anyone with

an undergraduate knowledge of elementary number theory, abstract algebra, and analysis. We have provided many references and notes for those who may wish to follow up on various topics, but in spite of the size of the Reference section, we must point out that it should not be regarded as complete. We have mostly included citations to work which is relevant to our theme of deriving methods for solving (0.1), and we sincerely hope that we have not though ignorance or inadvertence omitted any important contributions.

We had two principal objectives in writing this book. One was to provide a relatively gentle introduction for senior undergraduates, and others with the same level of preparedness, to the delights of algebraic number theory through the medium of a mathematical object that has fascinated people since the time of Archimedes. Our other goal was to detail the enormous progress that has been made, since Shanks' discovery in 1972 of what he termed the *infrastructure* of an ideal class, on the development of efficient algorithms for performing arithmetic in quadratic number fields. What we are able to do today is most remarkable; it certainly surprises us.

Acknowledgements

copy of his book with Ulrich Vollmer entitled *Binary Quadratic Forms: An Algorithmic Approach.* As this book covers in part some of the material of this volume, we are most appreciative of this gesture. It allowed us to produce a more focused work which presents some of the same material from a different point of view and which we hope will be seen as complementary to his. We extend our considerable thanks also to John P. Robertson for his eagerness in asking to be involved as a proofreader for this book from the very beginning and for his enormous competence in carrying out this duty. His efforts in this regard have resulted in a much better book. What errors remain, and it is inevitable that there will be some, are totally the responsibility of the authors.

To our former student, Reg Sawilla, we wish to express our thanks for the considerable programming effort that has gone into testing the many algorithms that appear in this work. We are also most indebted to our current student, Alan Silvester, for completing with great competence the enormous task of entering and editing this work on the computer. This should be seen in the light of our almost completely indecipherable handwriting, particularly that of the second author.

There are also a number of institutions that we wish to acknowledge. First and foremost of these is the Alberta Informatics Circle of Research Excellence (iCORE). Their generous support of our efforts has provided us with precious time and several much needed opportunities to interact with many scholars in the preparation of this work. We also thank Alberta Ingenuity and the Canadian Foundation for Innovation (CFI) for providing us with the funds needed to acquire the computing machinery that was used so often in producing the results and testing the routines in this work. In this connection, we want to thank Marc Wrubleski for his considerable efforts in keeping the machines working efficiently. Of course, we are also most grateful to the National Science and Engineering Research Council (NSERC) for their continued support of our research. We are also indebted to the libraries of several universities for providing us access to their collections. In no particular order these are the University of Calgary, the University of Toronto, the University of Sydney, Australian National University, and the University of Illinois at Urbana-Champaign. The second author would like to express his gratitude to the Fields Institute and to the Department of Computing Sciences at Macquarie University for providing him sanctuaries in which much of the work needed for his contribution to this work could be undertaken free of the inevitable interruptions that occur in his own institution.

Finally, we wish to acknowledge the contribution of our respective families to this project. During the time needed to complete this work, we have not been as attentive to them as we should have been, and we deeply appreciate this sacrifice on their part.

Calgary, AB,   *Michael J. Jacobson, Jr.*
June, 2008   *Hugh C. Williams*