The Book Review Column[1]

by Frederic Green

Department of Mathematics and Computer Science
Clark University
Worcester, MA 02465
email: fgreen@clarku.edu

There are a number of books out there being reviewed as I write this, but I would love to recruit more reviewers and get more reviews into these columns. The rewards include reading about something (at least in part) new, a published review on these esteemed pages... and, of course, a free book. If you're interested in any of the titles on the following page (which is not exhaustive; I welcome suggestions, preferably for books that have appeared within the past 3 years), please let me know!

In this column, we review the following 5 books. The first two deal with complementary aspects of "quantum," although at quite different levels.

1. **Quantum Algorithms via Linear Algebra**, by Richard J. Lipton and Kenneth W. Regan. A very accessible introduction to the basics of quantum computing, assuming only a background in linear algebra. Reviewed by Frederic Green.

2. **Quantum Information Theory**, by Mark M. Wilde. A detailed graduate-level text (developed, at length, "from the ground up") on the recently emerging field of quantum information. Reviewed by Subhayan Roy Moulick.

3. **Genome-Scale Algorithm Design (Biological sequence analysis in the era of high-throughput sequencing)** by Veli Mäkinen, Djamal Belazzougui, Fabio Cunial and Alexandru I. Tomescu. A graduate-level text about state-of-the-art data structures and algorithms for modern sequence analysis, especially those involving very large datasets, and hence requiring very efficient algorithms. Reviewed by Steven Kelk.

4. **The Mathematics of Encryption: An Elementary Introduction**, by Margaret Cozzens and Steven J. Miller. Another introduction, this one about the mathematics underlying encryption, eminently suitable for undergraduates. Reviewed by George Ledin Jr.

5. **Mathematics Everywhere** by Martin Aigner and Ehrhard Behrends (Eds.). An assortment of widely-ranging articles, presented in this volume published by the AMS, which demonstrate the omnipresence of mathematics in numerous fields. It is directed at a broad readership, but contains some real mathematics. Reviewed by S. V. Nagaraj.

BOOKS THAT NEED REVIEWERS FOR THE SIGACT NEWS COLUMN

Algorithms

1. *Distributed Systems: An algorithmic approach (second edition)*, by Ghosh
2. *Tractability: Practical approach to Hard Problems*, Edited by Bordeaux, Hamadi, Kohli
3. *Recent progress in the Boolean Domain*, Edited by Bernd Steinbach
4. *A Guide to Graph Colouring Algorithms and Applications*, by R.M.R. Lewis
5. *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, Ariel D. Procaccia, Eds.

Programming Languages

1. *Selected Papers on Computer Languages* by Donald Knuth

Miscellaneous Computer Science

1. *Algebraic Geometry Modeling in Information Theory* Edited by Edgar Moro
2. *Communication Networks: An Optimization, Control, and Stochastic Networks Perspective* by Srikant and Ying
3. *CoCo: The colorful history of Tandy's Underdog Computer* by Boisy Pitre and Bill Loguidice
4. *Introduction to Reversible Computing*, by Kalyan S. Perumalla
5. *A Short Course in Computational Geometry and Topology*, by Herbert Edelsbrunner

Computability, Complexity, Logic

1. *The Foundations of Computability Theory,* by Borut Robič
2. *Models of Computation*, by Roberto Bruni and Ugo Montanari
3. *Proof Analysis: A Contribution to Hilbert's Last Problem* by Negri and Von Plato.

Cryptography and Security

1. *Cryptography in Constant Parallel Time,* by Benny Appelbaum
2. *Secure Multiparty Computation and Secret Sharing*, Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen
3. *A Cryptography Primer: Secrets and Promises*, by Philip N. Klein

Combinatorics and Graph Theory

1. *Finite Geometry and Combinatorial Applications*, by Simeon Ball
2. *Introduction to Random Graphs*, by Alan Frieze and Michał Karoński
3. *Erdös–Ko–Rado Theorems: Algebraic Approaches*, by Christopher Godsil and Karen Meagher
4. *Words and Graphs*, by Sergey Kitaev and Vadim Lozin

Miscellaneous Mathematics and History

1. *The Magic of Math*, by Arthur Benjamin
2. *Professor Stewart's Casebook of Mathematical Mysteries* by Ian Stewart

Review of[2]
Quantum Algorithms via Linear Algebra
by Richard J. Lipton and Kenneth W. Regan
MIT Press, 2014
192 pages, Hardcover, $45.00

Review by
Frederic Green fgreen@clarku.edu
Department of Mathematics and Computer Science
Clark University, Worcester, MA

1 Background

After a little over a decade or so of relative obscurity, quantum computing ("QC") was established as a central field of theoretical computer science in the wake of Shor's breakthrough 1994 quantum algorithm for factoring [Sho]. By now QC is part of the theoretical computer science educational canon. Numerous courses all over the world have routinely included the basics of QC for well over a decade, and standard textbooks on theory of computation increasingly include a chapter or more on quantum computing.

"Quantum Algorithms via Linear Algebra" ("QALA") is a very focused, eminently accessible, and highly readable account of the "classical quantum algorithms": Shor's [Sho] and Grover's [Gro] algorithms, most prominently. The title derives from the fact that the emphasis is, first, on algorithms and, secondly, on the fact that these algorithms are all based entirely on linear algebra over a complex finite-dimensional Hilbert space. The computational model is presented in simple terms, and then applied. Very little is said, intentionally and with good reason, to motivate the model and explain *why* QC works the way it does. No Schrödinger equation, no uncertainty principle, no perturbation theory, no double-slit experiments, and indeed, no physics at all.

This is just as well, since QC doesn't need most of that machinery. More to the point, it works because quantum mechanics works, and nobody understands why. It just does! Of course, it often works in surprising ways, and Lipton and Regan are careful to point out just how very surprising the consequences of quantum mechanics can be. To paraphrase Neils Bohr, whoever isn't shocked by the contents of this book hasn't understood it. Fortunately, the authors are quite successful in helping the reader follow (and understand) *how* QC works.

One of the most noticeable features of the book is its brevity. If you want to get an idea of the subject quickly, this is about as quick as you're likely to get to a rigorous understanding. Another distinctive feature is that it uses the "elementary" techniques of Grover's algorithm to get into some much more recent, and important, work on quantum walks.

The presentation avoids a number of technicalities that might stand in the way of a QC newbie. For one, only pure states are considered, and the emphasis is solidly on the computational basis. Although sometimes the notion of mixed states is mentioned, and the partial trace plays a role, this is all tangential to the main development. The model introduced is pretty much as simple as possible (and as Einstein might have said, "no simpler"). The authors also avoid Dirac's bra/ket notation almost entirely. This certainly has its advantages, as it is, for the broad populace of computer scientists and mathematicians, not universally known. It *is* common practice in QC and quantum physics, of course. Having been trained as a physicist myself, it

has been second-nature for so long it doesn't bother me in the least. But there is virtue in eliminating any possible barrier to understanding, however small it may be.

2 The Contents

Here is a run-down of the chapters, with some commentary.

1. Introduction: No beating around the bush here. The entire computational model is presented by page 5! Of course, there's some necessary elaboration to follow both here and in the following chapters, especially on the nature of unitary transformations.

2. Numbers and Strings: Very basic preliminaries, such as binary notation and asymptotic notation.

3. Basic Linear Algebra: Vectors, Hilbert space, tensor products of vectors and matrices. A bit more than "just" linear algebra, since via understanding a matrix as an adjacency matrix in the state space, a concrete introduction to Feynman's sum-over-paths formulation of quantum mechanics is given.

4. Boolean Functions, Quantum Bits, and Feasibility: A very careful and lucid, yet succinct, explanation of what it means for a computation to be feasible in terms of circuits, classical and quantum. A lot of mileage is gained through analysis of the majority function. The chapter culminates in a formal definition of what it means for a quantum computation to be feasible.

5. Special Matrices: A compendium (with relevant proofs) of feasible matrices that are instrumental in QC. This includes the Hadamard and Fourier transforms, controlled-NOT, Toffoli gates, and reflections (anticipating Grover's algorithm).

6. Tricks: Start vectors, compute-copy-uncompute, the action of the Grover oracle (elsewhere often called the phase kickback trick), and various others.

7. Phil's Algorithm: In this amusingly named chapter, let's just say that Phil is to Lipton and Regan what the cat was to Schrödinger... and is, at the same time, a much more interesting and less ill-fated character. While there is no particular well-defined "Phil's Algorithm," Phil is central to a framework for the subsequent algorithms. He lies at the heart of a lively, amusing and pretty accurate graphical picture of how entanglement, superposition and interference work within the sum-over-paths model.

8. Deutsch's Algorithm: This and the next 3 chapters constitute the mainly gentle ascent to the summit of Shor's algorithm, following what is certainly the most logical and gentle trail, and also the historical one. Deutsch's Algorithm allows you to determine the exclusive-or of the two values of a Boolean function with just one evaluation, the first hint that quantum computing has an advantage over classical. The chapter also takes advantage of the similar techniques to give an introduction to superdense coding and quantum teleportation. Phil plays an active role in this chapter to animate all the algorithms and concepts.

9. The Deutsch-Josza Algorithm: This very brief chapter gives the algorithm that determines if a Boolean function over n inputs is balanced (i.e., has an equal number of 1's and 0's) or constant, under the promise that it is one or the other. This was the first algorithm that gave an exponential improvement over classical algorithms, albeit for a somewhat artificial problem.

10. Simon's Algorithm: Another short chapter about the algorithm that finds the "hidden period" of a Boolean function, assuming it has one. In particular, it finds an s such that $f(y) = f(z)$ iff $y = z \oplus s$. Again this gives an exponential speed-up for quantum computers, and solves a more natural problem than Deutsch-Josza, especially in light of what followed.

11. Shor's Algorithm: This presents Shor's original approach to period finding, which is the only part that relies directly on quantum computation (via the Quantum Fourier Transform (QFT)). While Simon's algorithm was its direct inspiration, and the QFT generalizes the Hadamard transform as used in Simon's algorithm, Shor's algorithm was a giant leap forward. Of necessity, here the level of difficulty steps up quite a bit from the preceding chapters, but the analysis is done carefully, readably and elegantly. Furthermore, as is clearly demonstrated here, one does not need a great deal of mathematical background to follow its main features. The only part that is left out, understandably, are the details of how to approximate the QFT, which is here assumed to be exactly feasible. The final (classical) step in finding the period here uses integer linear programming. The reader is also given the option to understand this step in terms of the original proof using continued fractions.

12. Factoring Integers: This describes the number theoretic background by which factoring *classically* reduces to period finding, and thus proves the result that factoring can be done in (probabilistic) quantum polynomial time.

13. Grover's Algorithm: Grover's algorithm determines, with high probability, if there exists a string x (among exponentially, say $N = 2^n$, many) that obeys any feasible property in time $O(\sqrt{N})$. This is a pretty standard treatment of Grover's algorithm, building on tools that were developed in Chapters 5 and 6. The geometrical picture of Grover's algorithm is nicely described in the text. I missed not seeing the standard figure illustrating that picture (by now well-known, but maybe not to the readers of this book); it would be a welcome addition. The chapter also includes a (relatively) lengthy and technical section on the important subject of approximate counting of the number of solutions to the search problem.

14. Quantum Walks: This is an excellent elementary introduction to random walks, beginning with classical random walks and a very nice description of how quantum random walks emerge from the classical idea. There is solid motivation underlying the unitary operators describing quantum random walks, and the unexpected consequences are cogently explained. Quantum walks diffuse much more rapidly, and interference tends to destroy many of the possible classical outcomes.

15. Quantum Walk Search Algorithms: This and the previous chapter are more advanced, and seem like an unusual choice of topic in an introductory text. However, they do a great job leveraging the work that was already done on Grover's algorithm. Furthermore, this chapter deals with work that is much more recent than Shor's and Grover's algorithms, within the past decade [MNRS], and places Grover's algorithm within a wider framework. The generic algorithm for a quantum walk search on a graph leads in a natural way not only to Grover's algorithm but also quantum algorithms that give a quadratic speedup relative to classical ones (like Grover) for a number of problems: element distinctness, subgraph triangle incidence, playing chess, and evaluating certain Boolean formulas. There are more lengthy and advanced treatments of this material in the textbook of Moore and Mertens [MM] (with a quite different emphasis) and the monograph by Portugal [Por] on quantum walks, but this is the only treatment known to me at this accessible a level.

16. Quantum Computation and BQP: It is a wise choice, which benefits the novice reader, to defer the introduction of complexity classes until one has witnessed the power of quantum algorithms, and hence gained a decent motivation for defining them. The chapter gives a definition of BQP, proves the universality of a standard set of gates for BQP computations, and characterizes acceptance amplitudes in terms of the difference between positive and negative roots of polynomials. This in turn sets the stage for bounding BQP from above by PP.

17. Beyond: This chapter takes up some of the broader issues on the subject, and includes some hints for further study, and a thought-provoking discussion about the nature of quantum computing and where it gets its power.

There are numerous exercises at the ends of chapters.

3 Further Comments

As mentioned above, this book really gets on with it, neither getting embroiled in the irrelevant technical details of quantum mechanics, nor in knotty interpretive problems of quantum theory. The latter is a fascinating subject in itself, but there is no need to grapple with it in understanding quantum algorithms. And anyway, those very algorithms actually lend insight to how these foundational issues might finally be resolved. If you want to understand interpretations of quantum mechanics, start with quantum algorithms! David Deutsch, one of the founders, regarded quantum computing as a vindication of the many-worlds interpretation of quantum mechanics. There are many detractors, but this illustrates how the deceptively easily understood world of quantum algorithms was motivated by, and can lead to, very deep questions.

The text is very reader-friendly, facilitated by bite-size chapters. As is only fitting, most of the learning gets done in the exercises. A number of these exercises are instrumental in fully understanding later more advanced developments in the text.

QALA serves a different function than existing dedicated QC textbooks that immediately come to this reviewer's mind. For example the widely-regarded classic Nielsen and Chuang [NC] is likely remain required reading for all who desire a broad and deep mastery of the subject, but is geared towards a graduate or research audience. Mermin's book [Mer] is narrower and much more succinct than [NC], but assumes more of the reader than QALA. Interestingly (and unavoidably, due to the earlier dates of those publications) neither of those texts covers quantum walks.

The level of QALA is appropriate for undergraduate computer science or math majors or minors. I easily imagine directing a readings course based on it, or teaching a full or partial semester-long course based on the book. It could also serve as great supplemental reading for a course, graduate or undergraduate, which may be based on a text that does *not* get into QC. One could cover the first 12 chapters (up to and including Shor's factoring algorithm) in a matter of weeks, given sufficient background and interest on the part of the students. For researchers who want a quick introduction to the field, it is perfect.

References

[Gro] L. Grover, Fast quantum algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 212-219, 1996.

[MNRS] F. Magniez, A. Nayak, A. Roland, and M. Santha, search via quantum walk. In *Proceedings, 39th Annual ACM Symposium on Theory of Computing (STOC '07)*, pp. 575-584, 2007.

[Mer] N. David Mermin, Quantum Computer Science. Cambridge University Press, 2007.

[MM] C. Moore and S. Mertens, The Nature of Computation. Oxford University Press, 2011.

[NC] M. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, 2000.

[Por] R. Portugal, Quantum Walks and Algorithms. Springer-Verlag, 2013.

[Sho] P. Shor, Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual IEEE Symposium on the Foundations of Computer Science (FOCS '94)*, pp. 124-134, 1994.

Review of[3]
Quantum Information Theory
by Mark M. Wilde
Cambridge University Press, 2013
672 pages, Hardcover, $79.00

.

Review by
Subhayan Roy Moulick subhayan@acm.org
Department of Physical Sciences
Indian Institute of Science Education and Research, Kolkata, India

1 Introduction

Over the last decade, quantum information theory has become one of the frontier areas of study among mathematicians, computer scientists, physicists, and engineers. This is because of the profound implications in communication and computational tasks that quantum physics offers, and that challenge our current classical conventions. Quantum information theory regards quantum states as a new form of information, and seeks to understand what quantum information is good for – things that classical computers cannot do. Mark Wilde's book aims to foster the very same and discusses the role of quantum information for communication tasks.

While there are several books and concrete references that have been written in the area over the years, this book offers a unique approach and focuses on quantum information theory alone. It includes the post-millennium results, which most others do not convey. The book is divided into six sections and, through twenty-five chapters, it introduces the subject and builds up a solid foundation with rigorous (but approachable) mathematics.

2 Summary

The first section gives an introduction to (quantum) Shannon theory and mentions the essentials and the history and development of the subject. The second chapter on classical Shannon theory gives a necessary brief introduction to the same.

The second section is where quantum theory makes its real appearance. Chapter 3 discusses the noiseless quantum theory and talks about qubits, gate operators, and measurements. The uncertainty principle, composite systems and their evolution, along with the famous no-cloning theorem, and the star of quantum theory - entanglement - is covered here. The chapter finishes with an extension to qudits, or higher dimension systems, which is a rather straightforward extension to include. This is followed by the famous Schmidt decomposition and CHSH games, all have well presented here. However, quantum systems are not always perfect, so say nothing of noiseless. This is not a matter of philosophy, but a hard fact when dealing with quantum systems. Thus Chapter 4 gives the noisy quantum theory, along the same lines as the preceding chapter. Ensembles, POVM, the partial trace, classical-quantum states, and very important examples of noisy evolutions as channels are new, among other things. The key takeaway from this chapter is the ensemble viewpoint which the author demonstrates beautifully, and which is to be a key tool for further study

and research. The last chapter of this section discusses quantum theory from a radically different viewpoint, through the purification theorem. Loosely, this has to do with the idea that the uncertainty in quantum systems is due to their being entangled with other systems to which we do not have access. It talks about the purified quantum theory, and isometric extensions of the evolution of a system, quantum instruments, and measurement. This is one of those chapters that covers topics most books do not cover in their full glory.

Section 3 talks about the unit quantum protocols. Chapter 6 defines what it means to be a unit non-local quantum resource. Following that there are introductions to the protocols for entanglement distribution, superdense coding, and finally the very interesting, quantum teleportation. The arguments given here about their optimality using resource inequalities are interesting. In addition, again there is a straightforward extension to higher dimensions which is appended to the end of the chapter. Chapter 7 discusses coherent protocols through the implementation of the coherent bit channel, coherent dense coding, and coherent teleportation. A fundamental result, derived in this chapter, is the coherent communication identity. The protocols discussed in Chapter 6 are revisited in Chapter 8, to obtain the unit resource capacity region, and show that the unit resource achievable region consists of all linear combinations of the three protocols. Furthermore, it is shown that that is the best that one can do with the unit resources. From this, the direct and converse theorems are also derived.

Section 4 is where things get very interesting. It is devoted to introducing the mathematical toolkit required to study quantum Shannon theory. Chapter 9 is about distance measures, which allow comparison of quantum states and quantum channels. It discusses trace distance and fidelity from the very basics to the operational interpretations. It also talks about the relationship between the two. Gentle measurements, fidelity of noisy quantum channels and the Hilbert-Schmidt distance measure conclude the chapter. Information theory, right from the very first principles – information as a surprisal of an event to the ideas of entropy and information inequalities – are beautifully given in Chapter 10. This is perhaps my favorite chapter in the book, due to its condensed presentation without any compromise of rigor. Chapter 11 extends the ideas of classical information developed in the previous chapter to the quantum domain, along with discussions on the non-classical ideas of entropy being negative in conditional quantum entropy. This is quite profound in a sense, that we see entropy as the expected value of information content, which is always positive in the classical regime. The operational interpretation of conditional quantum entropy is interesting. However while not covered in the first edition, it is anticipated in the second edition of the book. I was quite impressed to read about continuity of quantum entropy, and the subsection on the uncertainty principle in the presence of quantum memories. While the preceding two chapters considered static entropic quantities, Chapter 12 is about the six dynamic entropic quantities for classical and quantum channels. These include mutual information of a classical and quantum channel, private information of a classical wiretap channel and quantum channel, Holevo information and coherent information of a quantum channel. The chapters that follow in this section are about classical typicality and are centered around the asymptotic equipartition property. These chapters build towards an asymptotic theory of information through elaborate discussions on weak and strong (joint and conditional) typicality. The application of typical sequences in compression and conditional typicality to Shannon's channel capacity theorem is also well explained here. In the same spirit, typicality in the asymptotic quantum domain, in the i.i.d. setting, is described in Chapter 14, in addition to definitions of quantum information source and typical subspace measurement. The last two chapters of this section cover the packing lemma and cover lemma. These two are in a sense opposite to each other. These are rather important and new results which have been developed recently, which is again quite unique to the book.

Section 5 consists of two chapters which study compression of information and entanglement concentration. Chapter 17 studies the quantum data compression theorem and the achievable rates, and proves both

the direct coding theorem and the converse theorem for quantum data compression. It ends with an example and a brief discussion of other variations. The next chapter discusses how one can manipulate the "quantity" of entanglement using LOCC, through entanglement concentration. It starts out with a very clear example, and then formally defines entanglement concentration and proves it. The sections on common randomness compression and comparison with Schumacher compression are well worth the time.

The final section of the book deals with noisy quantum theory. This is where one encounters the gems of quantum Shannon theory. The Holevo-Schumacher-Westmoreland (HSW) theorem characterizes the classical capacity of a certain class of channels and is seen in Chapter 19. The classical capacity of the Hadamard channel, depolarizing channel, and superadditivity of Holevo information are also studied here. The next chapter in this section studies entanglement-assisted classical communication and asks how shared entanglement could be useful in transmitting classical information over a noisy channel. Following a detailed example, the chapter introduces the entanglement-assisted classical capacity theorem due to Bennett-Shor-Smolin-Thapliyal. I enjoyed the discussion on how (quantum) feedback does not increase (entangled assisted) classical capacity. It finally ends with a rather nice discussion of how to compute the entanglement-assisted classical communication with limited entanglement for the quantum erasure and amplitude damping channel. Chapter 20 discusses coherent communication with noisy resources. Chapter 22 and 23 consider the nature of information transmission over private channels, proves the private classical capacity theorem due to Devetak-Cai-Winter-Yeung, and the quantum capacity theorem, along with good examples. Discussions on quantum capacities and entanglement distillation conclude the chapter. Finally, the channel coding theorems discussed throughout the book are unified in Chapter 24, with the quantum dynamic capacity theorem, before concluding the book with yet another chapter containing the summary and outlook.

3 Opinion

My overall impression of the book is quite favorable. While at times I felt some parts were a bit tedious, I would still say the book does a phenomenal job of introducing, developing and nurturing a mathematical sense of quantum information processing. I also quite liked the coherent picture it paints. The steady presentation of the material from the ground up is done very well. While it may be useful for the reader already to have taken a course, or at least have had some exposure to quantum physics or information theory to fully appreciate the material, I am confident a motivated reader can do without. In a nutshell, this is an essential reference for students and researchers who work in the area or are trying to understand what it is that quantum information theorists study. Wilde, as mentioned in his book, beautifully illustrates "the ultimate capability of noisy physical systems, governed by the laws of quantum mechanics, to preserve information and correlations" through this book. I would strongly recommend it to anyone who plans to continue working in the field of quantum information.

Review of[4]
Genome-Scale Algorithm Design
(Biological sequence analysis in the era of high-throughput sequencing)
by Veli Mäkinen, Djamal Belazzougui,
Fabio Cunial and Alexandru I. Tomescu
Cambridge University Press, 2015
Hardcover, 391 pages, $64.99

Review by

Steven Kelk (steven.kelk@maastrichtuniversity.nl**)**
Department of Data Science and Knowledge Engineering (DKE)
Maastricht University (UM), Netherlands

1 Overview

The growing interest in "big data" means that algorithm designers are more than ever confronted with the distinction between theoretically efficient algorithms and practically efficient algorithms. Of course, some datasets are now so large that even linear-time exact algorithms are hopelessly inadequate, and in such cases a transition to a different algorithmic paradigm (such as algorithms with sublinear time and/or space complexity) is inevitable. However, there are many datasets arising in practice that, while large, can still conceivably be tackled by aggressively optimized polynomial-time algorithms. The volumes of data generated in bioinformatics often fall into this "large but not hopelessly large" category and in recent decades this has kept quite a few algorithmic people busy, in particular those working on strings and data structures for efficient manipulation of strings (because for many species the genome can be modelled as a string of DNA symbols).

This recent, research-level book describes state-of-the-art algorithmic techniques and data structures for a whole range of problems in bioinformatics directly and indirectly related to genome sequencing. As I point out later in the review, it is not the kind of book you should read if you consider an algorithm with running time $O(n^2)$ to be efficient, because the strings tackled in this book are typically assumed to contain millions or billions of DNA symbols. At this scale attaining $O(n)$ running time and space complexity, or getting as close as possible (by painstakingly shaving off sublinear terms) is critical – and this is what this book is all about.

2 Summary of Contents

The book spans 16 chapters, divided into 5 sections. With a view to making the book as self-contained as possible the first section (encompassing approximately 70 pages) establishes basic definitions and concepts. For the theorists there is the obligatory crash course in bioinformatics, but for the rest this section is algorithmic, with introductory chapters on algorithmic complexity analysis, graphs and network flows. The network flow chapter is particularly elegant, highlighting the wide array of combinatorial problems that can be solved within this framework. Network flow takes some time to appear but it is used extensively in the last chapters of the book.

As continued throughout the book each chapter is concluded with challenging, Ph.D.-level exercises and an overview of relevant literature. The literature overviews are excellent, helping the reader to appreciate the wider context of the (often very recent) material presented in the chapter. The most striking chapter within the first section is the very short introduction to data structures. I must confess, I am indeed someone who stops optimizing algorithms at $O(n^2)$ and in terms of data structures I haven't ventured much further than the material I learned during my bachelor degree, so for me this chapter was somewhat intimidating, discussing topics such as bitvector rank and select operations, wavelet trees and range queries.

The second section of the book, spanning approximately 50 pages, is also introductory, focusing heavily on alignments and, to a lesser extent, Hidden Markov Models. Alignment is the process by which multiple strings of DNA are rendered (biologically) comparable by inserting gaps and deleting symbols. Theorists will recognize that this is a very close relative of the Edit Distance problem, which recently attracted attention because of the proof that subquadratic-time algorithms are not possible unless the Strong Exponential Time Hypothesis fails. Here practical speed-ups for Edit Distance are discussed (e.g. Myers' bitparallel algorithm) and closely related problems (such as Longest Common Subsequence) also receive some attention. The discussion on Edit Distance gradually shifts into a discussion on alignment as it is used in bioinformatics. This practical discussion is slightly less technical than other parts of the book, in part because it concerns well-established material, and in part because alignment is NP-hard: the book tends to back away from problems that are NP-hard. This is entirely understandable given the focus of the book.

The third and fourth sections of the book, on genome-scale index structures and genome-scale algorithms respectively, constitute in some sense its algorithmic core.

The third section (spanning 80 pages) starts with a rigorous and technical description of efficient algorithms for construction and manipulation of suffix trees (and arrays). Suffix trees are a classical data structure in which suffixes of a string are encoded as root-to-leaf paths in a rooted tree. As the authors demonstrate they are flexible and powerful data structures and can be used for enumeration of maximal repeats, maximal unique matches and so on. The material on suffix trees is presented as a natural lead-up to the chapter on Burrows-Wheeler (BWT) indexes. The Burrows-Wheeler index (based on the Burrows-Wheeler Transform) can be viewed as a space-efficient variant of the suffix tree that can nevertheless support a number of useful navigational primitives efficiently. Specifically, they can be stored in $O(n \log \alpha)$ bits where n is the length of the string and α is the size of its alphabet (which in bioinformatics is often 4). This saves (at least) a $O(\log n)$ factor with respect to suffix trees. The material on the BWT index is necessarily technical, extending into the bidirectional BWT index and several variants which operate on trees and (directed acyclic) graphs rather than strings. The BWT index is used repeatedly in later chapters and, as the authors themselves state, it is the foundation for much of the new and/or state-of-the-art material presented in the book. Judging by the literature overview it is also a research specialization of at least one of the authors.

The fourth section (spanning 100 pages) on genome-scale algorithms consists of 4 chapters. The first of these, Chapter 10, is on the topic of read alignment. Roughly speaking the idea is that the sequencing process generates many short strings (corresponding to substrings of the sequenced genome) that subsequently need to be located within a longer reference string (which is a previously-assembled genome of a closely related organism). The crucial point here is that the match will in many places not be perfect, since the two genomes are not completely identical (and there are other complications to take into account, such as the fact that DNA actually consists of two complementary strings wound together). This necessitates the generalization of the earlier presented data structures, which mainly support exact pattern matching, to approximate pattern matching. This generalization is described for suffix trees and then for the BWT index. A number of variations of the read alignment problem are discussed which take the specifics of modern sequencing technology into account, such as the deliberate generation of reads with gaps in the middle.

Chapter 11, on genome analysis and comparison, tackles the problem that, although the alignment operation is polynomial-time solvable on two strings, entire genomes are simply too long to compare this way, comprising often billions of symbols. For this reason genomes are often compared by building summary statistics around the frequencies of locally occurring substructures. The chapter starts by describing how the BTW index can be utilized to enumerate all maximal repeats, or maximal exact matches, or maximal unique matches. It then moves onto the topic of string kernels. The goal here is to map the two strings into vectors, over which a similarity measure can be computed, without actually constructing the vectors explicitly. There is quite some attention for k-mer kernels (based on the frequencies of all length-k substrings) and substring kernels (based on the frequencies of arbitrary length substrings) and how to compute these space-efficiently using the BWT index. There then follows a statistical variant (substring kernels with Markovian correction) and an adaptation in which it is only necessary to compute an indexing structure for one of the genomes when certain auxiliary statistics for the other string are already available. The chapter closes with a few pages briefly introducing the notion of compression distance. This is a natural note upon which to end the chapter, since Chapter 12 concerns genome compression, focussing on progressively more time and space efficient algorithms for Lempel-Ziv parsing (based on the Burrows-Wheeler transform and range minimum query data structures), before moving onto bit-optimal Lempel-Ziv compression. As the authors note this last topic is a nice vehicle for demonstrating the combined use of many of the data structures discussed earlier in the book.

Chapter 13, on fragment assembly, focusses on the problem of merging strings into longer contiguous blocks. Much of the theory in this area rests on the idea of assembly graphs which, broadly speaking, are directed graphs in which vertices represent substrings and arcs denote that the two substrings at the endpoints overlap in some fashion. The most famous (but heavily stylized) example of this involves transforming the assembly problem into an Eulerian path problem on a *de Bruijn* graph, with arcs denoting that the suffix of one endpoint is the prefix of the other, but this is just used for illustration; the material in this chapter goes somewhat further towards making this model more realistic. It then moves onto the scaffolding problem in which previously assembled strings, known as contigs, are partitioned into groups such that the contigs in each group form a connected component (in some well-defined, but highly sequencing specific, sense).

The final section of the book, on applications, is approximately 80 pages long, divided into chapters on genomics, transcriptomics and metagenomics. The chapters on genomics and transcriptomics cover a variety of different problems and their algorithmic solutions. The material in these two chapters is somewhat lighter than in earlier chapters because there is less reliance on sophisticated data structure construction and manipulation. Indeed, many of the problems presented here can be tackled using dynamic programming, matchings, network flow and other standard techniques. (This is not to say that the reductions are trivial - on the contrary, they are often highly elegant and constitute great examples of how these staples of combinatorial optimization can be used in unexpected contexts). The final chapter, on metagomics, looks at some of the combinatorial issues arising when data is drawn not from a single genome but from a "soup" of different organisms mingled together.

3 Conclusion

This is a very well written, research-level book that covers a wide range of algorithmic topics related to sequence analysis, focussing primarily on time/space-efficient algorithms and data structures for genome-length strings. In essence it is a comprehensive research-level reference for advanced string algorithms in bioinformatics, merging many established and recent results together into a coherent whole and occasionally adding some new results of its own. The proofs, woven into a classical motivation, model, lemma,

theorem structure are rigorous, convincing and to a large extent self-contained. The data structures involved, while complex, are fundamental and lightweight in the sense that - as far as I can tell - there is no attempt to hide enormous constants in the big-O notation, or to leverage existing results that do just that. Indeed, the book makes a very authoritative impression not just on the algorithmic side but also on the applied/biological side: it's clear that, although the authors keep the biology to a minimum, they also understand this side of the story very well indeed. This last point is emphasized by the supporting website (`http://www.genome-scale.info`) which incorporates a number of excellent resources, including links to software packages where many of the algorithms covered in the book have been implemented. The literature overviews at the end of each chapter are also insightful, and can almost be read independently of the more technical material.

The only critical point I have about this book is not really a crticism of the book *per se*, but rather a word of warning to potential readers. As alluded to earlier in the review, this will not be an easy book to read if you are the type of person who already considers $O(n^2)$ to be efficient and/or who recoils in horror at the idea of expending energy to shave logarithmic factors off running times. Similarly, given its dominant focus on algorithms with (near-)linear time/space complexity, it's not going to appeal to algorithmic people who are mainly interested in the interface between P and NP. As you might have guessed I am in this last group so, yes, I found it pretty tough going to read this book. However, I stand by my positive review because I recognize its overall quality and I can see that for the right audience this is timely and important material. I also think that in the era of "big data" books like this have an increased importance, for the following reason. Many young students of computer science are so immersed in the "big data" culture that they start to see any problem instance that does not fit on an exam sheet as a nail to be hit with their MapReduce or machine learning hammer. This book is an important reminder that, with enough algorithmic effort, exact algorithms can be used to solve really quite unexpectedly large problem instances.

Review of[5]
The Mathematics of Encryption An Elementary Introduction
by Margaret Cozzens and Steven J. Miller
AMS, 2013
Mathematical World Volume 29
Softcover, xvii + 333 pages, $49.00

Review by
George Ledin Jr ledin@sonoma.edu
Computer Science Department
Sonoma State University, Rohnert Park, CA

This is a marvelous book. I highly recommend it.

If you are not acquainted with cryptology, this book is a great way to start. The authors provide explanations and examples and each chapter concludes with problems for solution. If you are familiar with cryptology but are not a practitioner, this book is an excellent overview of what you may already know and a solid introduction to topics that may be new to you. If you teach cryptology this book will serve as an outstanding textbook for upper division undergraduate students majoring in mathematics or computer science. Graduate students will get much value from using the book as a baseline.

By agreeing to review this gem of a book I should discuss what's in it but could be omitted or done differently, or what's not in it and should or shouldn't have been included. With a few exceptions I will refrain from doing so, because the choices made by the authors were theirs to make, and despite my comments and suggestions these choices were good choices. This 300-page book is just the right size.

Lets begin with the title, encryption. This is a term that under the umbrella of cryptology usually includes, however briefly or comprehensively, encoding and enciphering, and their inverses, decoding, deciphering, and, of course, decryption. Cryptanalysis, the process of breaking encryption, is also essential.

Recent news headlines have been focusing on encryption, melding all such variations. Journalists writing about strong encryption, back doors, and other arcane concepts, can be forgiven for not defining related ideas, such as cryptanalysis. The authors of this book, quite rightly, decided to abstain from getting mired by such semantic superficialities, which is as it should be their prerogative. Besides, if news media report confrontations, such as Apple vs. FBI, couching them as encryption battles, we should welcome the attention that presents us with opportunities to examine emerging technologies.

Each of the twelve chapters of Cozzens and Millers books packs a wallop. The first three chapters present basic ideas in their historical contexts. Although there is no mathematics, the first chapter demands to be read more than once. Enigma and Purple are mentioned in Chapter 1 and discussed more thoroughly in Chapter 3. Rotor machines are intriguing in themselves – potentially educational toys – but Lorenz (the "undecipherable Enigma") and Fialka (the Soviet version) are omitted. Also curious is the lack of reference to the standard history book, Codebreakers, by David Kahn.

Encryption in a narrower sense pertains to converting plaintext (cleartext) into ciphertext and recovering the plaintext from a given ciphertext. Cozzens and Miller do a splendid job on these fronts. But they also introduce hash functions, error detecting codes, linear feedback shift registers, digital signatures, key exchange, lattices, quantum cryptography, and other topics, making their book a true resource in a much wider sense. And steganography to boot.

Here are some comments dealing with what Cozzens and Miller do not cover. These are quibbles that do not affect their accomplishment.

The prehistory of cryptology starts with substitution and transposition. Substitution is illustrated by Caesar's, which can be generalized as the affine cipher. Transposition relies on scytale or colorful and easy to visualize "railfence" and other ciphers. The mathematics of affine ciphers is reedy (and adequately dealt with in this book) while transposition can be relegated to the Sunday comics.

Cozzens and Miller omit Playfair, but this is no great loss. The idea is to flatten the frequency distribution, an idea nicely presented in Chapters 4 and 5.

An entertaining combination of affine and frequency ciphers is the n^{th} appearance alphabet. For example, the second appearance alphabet based on Abraham Lincoln's Gettysburg Address is `o r s e n a u f g h t c i v d b y p l w m k j q x z`. In the limit, as $n \to \infty$, this yields to single-letter frequency analysis. Guessing the underlying text adds a steganographic flavor to the pursuit.

Single-letter keystreams are not limited to the standard Vigenère tableau; the state space of these 26×26 tables is less than $(26!)^2$, which is large enough. Double-letter keystreams can pick any one of nearly $(676!)^2$ obviously much larger tables.

Kasiski's test helps determine keyword length but it is based on the assumption that the ciphertexts are monoalphabetic. For monoalphabetic ciphertexts letter frequency information would ordinarily be sufficient to get at their plaintexts. Friedman found a clever way to determine the likelihood that a ciphertext was not monoalphabetic.

What could be more secure than the one-time pad? In section 5.4 the authors delve on the concept of perfect security. (Check the closing credits of Swordfish, a movie starring Halle Berry, to see mentioned the mysterious Vernam cipher.)

By section 5.5 the authors expect readers to have learned various ways to break several ciphers as well as how to perform attacks, such as known-plaintext and ciphertext-only. Next stop: NSA.

Symmetric encryption reappeared as DES, about one hundred years after Playfair. Starting with Lucifer, the authors give their readers a quick chronicle of how symmetric encryption works. For example, plaintext is chunked into 64-bit blocks, while the master key is reduced to 56-bit ($28 + 28$) subkeys.

When mathematicians (as opposed to computer scientists) teach encryption, symmetric key methods, such as DES, Triple-DES, DES-X, AES, SMS4, etc., suffer by either being almost completely ignored or by focusing on parts like the number of rounds, or emphasizing the pretty ideas, such as irreducible polynomials as moduli.

Nevertheless, it is okay. Cozzens and Miller safely bypass these brambles. Symmetric key encryption is mathematically messy and to do it justice would require significantly increasing the size of this book. And they do an excellent job with areas that are mathematically elegant. Fermat's little theorem is a cogent example. This failed primality test would be splendid were it not for pseudoprimes and, especially, those monsters known as Carmichael numbers. Yes, read all about it in Chapter 7.

Asymmetric (public) key encryption proved to be catnip, mathematically speaking, in the form of RSA. The authors describe this method in all its gracefulness. They also show how other pivotal ideas, such as signatures and key exchange (key management) were spawned from RSA.

Cozzens and Miller acquaint their readers with cryptographically useful hashing methods. The IACR (International Association for Cryptographic Research) ramped up research aimed at finding a new secure hash algorithm after the 2004 event at their annual meeting at UCSB, where the previous standard proved insufficiently collision resistant.

S-boxes (not the DES kind) are permutations. AES and SMS4 are examples of 16×16 S-boxes. Desirable features of a cryptologically well-designed S-box include high order (or power – the exponent that transforms a given permutation into the identity), mid-center permutation number (number of inversions), none or few fixed points, zero or near-zero correlation coefficient, and in some sense random length and

number of cycles. The hidden gremlin in this list is the fact that Franklin's 16×16 magic square meets all of these criteria.

The richest and to date most promising cryptological development is that of elliptic curve cryptography (ECC). ECC supersedes RSA, in that ECC relies on discrete logs while RSA relies on factoring. Cozzens and Miller must have done a lot of soul searching to decide to leave ECC out, which has a lot of fertile mathematics, but to keep in other topics, such as Hill. But then homomorphic encryption is missing, too.

I understand, and agree with, their decision. In practical terms giving ECC its due would significantly increase the size of this beautiful, compact book. Urging them to write a follow-up book, with a modified title A More Advanced Introduction? Whom am I kidding? This book was no weekend affair. All of us ought to be grateful that they did such an excellent job that whoever wants to learn ECC can do so mastering the book's solid foundation from which to proceed.

Well done!

Review of[6]
Mathematics Everywhere
Martin Aigner and Ehrhard Behrends (Eds.)
American Mathematical Society, 2010
330 pages, Softcover, $52

Review by
S. V. Nagaraj, svnagaraj@acm.org
VIT University, Chennai Campus, India

1 Introduction

Mathematics is a very enchanting branch of science. It has many applications in real life although we may not be fully aware of its role. This book is a collection of essays on the applications of mathematics by experts in various disciplines. It is based on lectures given at a popular hall in Berlin. This publication by the AMS is an English translation of a book originally in German. The first edition of the German language book appeared in the year 2000, the second in 2002, and the third in 2008. The translation from German to English was done by Philip G. Spain. The ISBN of the book is 978-0-8218-4349-9.

2 Summary

The book has five parts and is made up of twenty one chapters.

The first part is a prologue. It has just one chapter about how mathematics becomes a cult. It actually portrays the hopes of the author of that chapter that this should indeed happen.

The second part with seven chapters focuses on case studies.

Music lovers know that transferring music to a compact disc is better than that to a traditional vinyl disc. However, few may know that it is mathematics that really enables this. The interesting chapter on the mathematics of the compact disc looks at error-correcting codes and describes Hamming codes and Reed-Solomon codes.

Many may not believe a claim that math can play a vital role in liver surgery. The authors of the chapter on image processing and imaging for operation planning in liver surgery highlight the growing importance of computer supported radiology, where mathematical methods play a key role.

Graph theory provides many interesting problems which have practical significance. One such problem is the Hamiltonian circuit problem where we have to determine if a graph has a closed tour that visits every node in the graph exactly once. The authors introduce other engrossing problems such as the traveling salesman problem, the problem of finding shortest paths in graphs, and the duty scheduling problem of bus

drivers. The authors conclude that mathematics has made public transport cheaper and more efficient. They introduce the reader to the concepts of algorithms and complexity.

The remaining case studies are on topics as varied as spontaneous pattern formation and Turing's instability, design of new materials by employing mathematics, discrete tomography, and reflections.

The third part is entitled "Current Topics." It has seven chapters.

The role of mathematics in financial markets is highlighted by focusing on stochastic processes.

Electronic money has now become a reality. However, laypeople may not be aware of the role of mathematics, in particular cryptography, for securing electronic payment systems.

Here is an example of some of the detail in Chapter 11: Sphere packing was studied around 1611 by Johannes Kepler, who came up with a conjecture. The Kepler conjecture is about sphere packing in three-dimensional Euclidean space. It says that no arrangement of equally sized spheres filling space has a greater average density than that of the cubic close packing (face-centered cubic) and hexagonal close packing arrangements. The density of these arrangements is around 74.05%, more precisely $\pi/\sqrt{18}$. Many mathematicians including Gauss worked on this conjecture which was purportedly settled by T. C. Hales using computer power. Hales' proof is a proof by exhaustion involving the checking of many individual cases using complex computer calculations. The main difficulty with the proof is that it used massive computation viz. about 100,000 linear optimization problems, each formulated in about 100 to 200 variables, with 1000 to 2000 constraints. This only indicates that it is difficult to check the correctness of the proof and of the computer programs behind them. Though Hales' main work was finally published in the reputable journal Annals of Mathematics in 2005, nevertheless, computer assisted proofs are often looked at with skepticism. The discerning reader may recollect Appel and Haken's computer assisted proof of the four color theorem. Nonetheless, these days computers have become indispensable in proving theorems and settling conjectures.

The chapter on quantum computers talks about why prime numbers are important in cryptography and how to factor large numbers using quantum computers. The following chapter discusses some of the complicated mathematical ideas that lead to the solution of the 300 year old problem of Fermat known popularly as Fermat's Last Theorem. Fermat had conjectured that there are no non-zero integers a, b, c for which $a^n + b^n = c^n$ if n is greater than 2. Fermat claimed that he had a proof for this for which the margin of his notebook was too narrow. The chapter includes a high-level description of the connections between Fermat's Last Theorem, the theory of elliptic curves, and modular forms, which enabled Wiles and Taylor to finally prove the theorem. The chapter on a short history of the Nash equilibrium introduces the reader to key concepts in game theory. The next chapter is on mathematics for the study of climate change. More specifically, it discusses the application of numerical analysis and asymptotic analysis to climate change.

The fourth part is given the title "The Central Theme," and has five chapters.

The chapter on prime numbers, secret codes and the boundaries of computability mentions the interesting applications of prime numbers to modern cryptography. The chapter on knots presents absorbing discussions about the mathematics of knots. The next chapter studies soap bubbles from a mathematical perspective.

Poincaré conjectured that every closed simply connected three-dimensional space must be topologically equivalent to a three-dimensional sphere. This conjecture stimulated the minds of many mathematicians and was ultimately settled by Perelman. The last chapter of this part discusses mathematics and chance by focussing on probability theory.

The fifth part is the epilogue. It has just one chapter which discusses the many applications of mathematics in a multi-media civilization. By multi-media civilization the author of that chapter refers to the world in which computers, the Internet, email, search engines etc. are widely used.

3 Opinion

The book demonstrates that mathematics is all around us, although we may be unaware of its omnipresence. It portrays that math can be pure thought but at the same time a very practically applicable science. The book presents some well-known applications of mathematics in everyday life but also some of those that are not so familiar. Many recent developments in mathematics are covered in the book. Such developments include the solution to Fermat's Last Theorem. The book covers a wide range of topics that illustrate that mathematics is indeed everywhere even in places where we never expected it. New developments in technology illustrate that new ways of applying mathematics are being found.

The chapters in the book have been authored by notable mathematicians. The writing style in the chapters varies, but serves to convey the theme of the book, that mathematics has many real-life applications and it is everywhere. Although the book is based on lectures to popularize mathematics among the general public, some chapters are heavy on mathematics. Thus the book has a lot of real mathematics in it and for a proper understanding the reader would require some mathematical sophistication. Many are put off by mathematics at a young age and develop an aversion to it. This book serves to dispel such feelings by showing that mathematics is indeed apprehensible and can be fun. It takes the reader on a fascinating journey.

It is an uncomfortable truth that not all mathematicians are good evangelists of mathematics and its applicability to theory and practice. Thus, it is necessary to make laypeople aware of the power of mathematics. The latest applications of mathematics need to be presented to the general public, something precisely done by this book. The eager reader will benefit from the references at the end of chapters. The book is attractively printed on high quality paper, and includes many illustrations in color. The book will be useful for students and teachers of mathematical sciences and the mathematically-inclined layperson.