Preface and Introduction

S. S. RAVI
Department of Computer Science, University at Albany—SUNY, Albany, NY 12222, USA.
Email: ravi@cs.albany.edu

SANDEEP K. SHUKLA
Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA. Email: shukla@vt.edu

Overview

The purpose of this volume is to honor Professor Daniel Jay Rosenkrantz ("Dan" to his friends) for his extensive research contributions which have enriched the field of Computer Science. The volume includes reprinted forms of ten of Dan's publications in archival journals and eight contributed chapters by various researchers.

A Biographical Sketch

Dan was born on March 5, 1943 in Brooklyn, New York. He attended Evander Childs High School in Bronx, New York, before proceeding to Columbia University for his undergraduate and graduate degrees in Electrical Engineering. He received his Bachelor's degree in 1963. He was a National Science Foundation Cooperative Graduate Fellow at Columbia University and received his Master's degree in 1964. His Ph.D. thesis, completed in 1967, was supervised by Professor Stephen H. Unger. Dan spent the last year of his graduate studies at Bell Telephone Laboratories in Murray Hill, New Jersey.

After receiving his Ph.D., Dan joined the General Electric Corporate Research and Development Center (GECRD) in Schenectady, New York. During his ten years at GECRD, Dan worked with Philip M. Lewis II and Richard E. Stearns on a number of fundamental research problems in many areas including formal languages, compilers, algorithms and database systems. Their research led to numerous seminal papers, four of which are reprinted in Part I of this book. Another outcome of their collaboration is an early and highly regarded textbook on compiler design [13], which made the topic accessible to undergraduate students in Computer Science. The following quote from the contributed chapter[1] by Phil Lewis, nicely summarizes the years spent by Dan at GECRD:

> "Those were magic years for Computer Science at G.E., and Dan made important contributions to that magic."

In 1977, Dan moved to the Computer Science Department at the University at Albany–State University of New York (UAlbany) as a Full Professor. Dan was a member of the Computer Science faculty at UAlbany for 28 years before retiring in June 2005. From 1983 to 1985, when he was on leave from UAlbany, Dan worked as a Principal Computer Scientist for Phoenix Data Systems, a company which developed design automation tools for very large scale integrated (VLSI) systems. During his tenure at UAlbany, Dan supervised or co-supervised seven Ph.D. students and a large number of Master's students. He also served as the Chair of the Computer Science Department during the period 1993 to 1999.

Dan's first paper was published in *IEEE Transactions on Electronic Computers* in 1966. For more than 40 years since his first publication, Dan has published extensively in prestigious conferences and journals. As of September 2008, Dan's publication list includes more than 140 papers with 38 collaborators. These publications cover many different areas of Computer Science including formal languages, compilers, algorithms, database systems, VLSI design and testing, fault-tolerant computing, hierarchical specifications, software engineering, high performance computing, operations research, discrete dynamical systems and data mining. A beautifully concise statement which captures the nature of Dan's research contributions is the following partial quote[2] provided by Jeffrey Ullman (Computer Science Department, Stanford University):

> "His work has been a model for how one uses theoretical skills to make an impact on real problems, . . ."

Like his research record, Dan's record of service to the Computer Science community is also extensive. He was the Area Editor for Formal Languages and Models of Computation for the *Journal of the Association for Computing Machinery* (JACM) from 1981 to 1986 and then the Editor-in-Chief of the

[1] This chapter appears in Part II of this book.
[2] The full version of this quote appears on the back cover of this book.

same journal from 1986 to 1990. During his tenure as the Editor-in-Chief, three new areas (Logic in Computer Science, Computational Geometry and Deductive Systems and Equational Reasoning) were added to the journal [24]. Dan has served on the program committees of many well known conferences including IEEE Symposium on Foundations of Computer Science (FOCS) and its previous incarnation (IEEE Symposium on Switching and Automata Theory), ACM International Symposium on Principles of Database Systems (PODS), International Conference on the Management of Data (SIGMOD), IEEE Symposium on Reliable Distributed Systems (SRDS) and ACM Symposium on Programming Language Principles (POPL). He was the Chairman of the Program Committee for FOCS 1975 and also served as the Secretary of ACM SIGACT (Special Interest Group on Automata and Computability Theory) from 1977 to 1979. He was the General Chairman for PODS in 1984, 1990 and 1991 and served as a member of the PODS Executive Committee from 1990 to 1995. In addition, Dan has served on evaluation panels for the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA).

Dan has received many awards in recognition of his research contributions. He was elected a Fellow of the Association for Computing Machinery (ACM) in 1995. His fellowship citation reads as follows:

> "For pioneering contributions to formal languages, compiler design, algorithm analysis, databases, parallel and fault-tolerant computing and for exemplary ACM service including Editorship of JACM."

In 2001, Dan received the "Contributions Award" from the ACM Special Interest Group on Management of Data (SIGMOD). In 1991, he received the "Excellence in Research Award" from the University at Albany–State University of New York. Dan has been listed in *Who's Who in America* since 1990. His academic honors include Sigma Xi, Eta Kappa Nu and Tau Beta Pi.

Summary of Part I

The collection of papers which are reprinted in Part I of this book cover four decades (1967 to 2008) of Dan's research career. Many of the papers in this collection represent seminal contributions to Computer Science.

Chapter 1 ("Matrix Equations and Normal Forms for Context-Free Grammars") is the reprinted form of Dan's second journal paper, which originally appeared in JACM in 1967. This paper shows how the Greibach Normal Form [11] of a context-free grammar, where the right side of each production begins with a terminal symbol,[3] can be constructed efficiently by exploiting the

[3] The grammar may also have productions of the form $A \to \lambda$, where λ represents the null string.

correspondence between the productions of the grammar and a set of linear equations. The approach used in the paper obviates the need for generating regular expressions from a directed graph representation of the linear equations. Moreover, the paper also shows how the grammar can be modified so that each production starts and ends with a terminal symbol. A detailed discussion of the algorithm appears in the classic text on formal languages by Harrison [12, Sect. 4.9].

Chapter 2 ("Attributed Translations") is reprinted from the *Journal of Computer and System Sciences* (1974). This paper formalized the notion of attributed translation which plays a central role in the design of compilers for high-level languages. The classic text on compiler design by Aho and Ullman [1, p. 295] mentions that this paper developed one of the "classes of translation schemes for which attributed and synthesized translations can be implemented efficiently".

Chapter 3 ("An Analysis of Several Heuristics for the Traveling Salesman Problem") is an early paper on the analysis of approximation algorithms. This paper, which originally appeared in *SIAM Journal on Computing* (1977) is considered a classic and has been cited in virtually every subsequent paper on the topic. The paper established a tight performance guarantee of $\Theta(\log n)$ for the well known "near-neighbor" heuristic for the problem, where n represents the number of cities. This paper also introduced the minimum spanning tree based heuristic which provides a performance guarantee of $2(1 - 1/n)$.

Chapter 4 ("System Level Concurrency Control for Distributed Database Systems") and Chap. 5 ("Consistency and Serializability in Concurrent Database Systems") represent seminal contributions to the area of transaction processing in concurrent databases. The ideas presented in these papers are discussed in standard texts on database concurrency control [8, 22]. The first of these papers, which is reprinted from *ACM Transactions on Database Systems* (1978), presents several designs for concurrency control schemes and formally proves that the schemes are free from phenomena such as deadlocks. The second paper, which is reprinted from *SIAM Journal on Computing* (1984), shows formally that serializability is both necessary and sufficient for consistency in concurrent databases. The book by Bernstein et al. [8, p. 23] calls the conference version of this paper [25] an "influential early paper" on the topic of concurrency control.

Chapter 6 ("An Efficient Method for Representing and Transmitting Message Patterns on Multiprocessor Interconnection Networks"), reprinted from the *Journal of Parallel and Distributed Computing* (1991), discusses research done jointly by Dan and his first Ph.D. student (Philip Bernhard). This paper describes a representation (called a **mask**) for messages in multiprocessor interconnection networks and shows how a number of properties of the messages can be determined efficiently from the corresponding mask. (An example of such a property is whether the messages represented by a given mask will cause congestion.) In addition, it is shown that under this representation, the problem of partitioning a given set of messages into a minimum number of conflict-free rounds can be solved efficiently.

The paper reprinted as Chap. 7 ("Representability of Design Objects by Ancestor-Controlled Hierarchical Specifications") originally appeared in *SIAM Journal on Computing* (1992). This was coauthored with Dan's second Ph.D. student (Lin Yu). This paper developed a model called **versioned dag** (VDAG) for succinctly representing hierarchically specified design data. The paper provides a complete characterization of the expressive power of the VADG model and presents complexity results for a number of related problems.

Chapter 8 ("The Complexity of Processing Hierarchical Specifications"), reprinted from *SIAM Journal on Computing* (1993), represents another important contribution to the topic of hierarchically specified objects. In that paper, it is shown that any hierarchically specified acyclic circuit can be simulated deterministically in space that is linear in the size of the representation, even when the description is not explicitly acyclic. This result settled an open problem due to Lengauer [21]. Further, it is shown that the problem of simulating a hierarchically specified acyclic monotone circuit is PSPACE-complete and that the simulation of any hierarchically specified acyclic circuit (even if it is not monotone) can be carried out in deterministic time $2^{O(\sqrt{n})}$, where n represents the size of the description. In addition, it is shown that the simulation problem for hierarchically specified cyclic circuits is EXPSPACE-complete.

Chapter 9 ("Approximation Algorithms for Degree-Constrained Minimum-Cost Network-Design Problems") is reprinted from *Algorithmica* (2001). This paper is one of the first to consider the idea of approximating several objectives simultaneously. The definition of multiobjective approximation introduced in the conference version of this paper [23] is widely used today. The paper considered several variants of the problem of constructing spanning trees that minimize two objectives, namely the total cost and the maximum node degree. Since the publication of this paper, the problem and its variants have been studied by many researchers (see for example [2] and the references cited therein).

Chapter 10 ("Efficient Algorithms for Segmentation of Item-Set Time Series"), reprinted from *Data Mining and Knowledge Discovery* (2008), represents Dan's joint work with another Ph.D. student (Parvathi Chundi). This paper considers the problem of mining a special form of time series data, called **item-set** time series. In such data sets, the information stored for each time instant is a group of items rather than a single item. For example, in a software repository, the data stored for each time instant may be the names of files that were changed at that time or the names of people who modified those files.

One way to extract useful patterns from such data sets is to first divide them into appropriate segments. The paper defines the notion of an optimal segmentation under different objectives and presents polynomial time algorithms that construct such segmentations. The paper also includes extensive experimental results obtained by applying the algorithms to several data sets.

Summary of Part II

Part II of this book contains eight chapters contributed by various researchers. Like Dan's research record, these chapters also cover a wide variety of areas.

Chapter 11 ("Structure Trees and Subproblem Independence") by Richard E. Stearns and Harry B. Hunt III shows that many constraint satisfaction problems (such as different versions of the Boolean Satisfiability problem, their counting versions, several graph theoretic problems, etc.) can be captured using a very general framework called the **sum-of-products** form. This framework enables one to formalize the notion of subproblem independence, which has important implications on the time needed to solve the corresponding problem. The chapter shows that when instances of the problem have bounded treewidth, one can readily obtain an efficient algorithm from the general framework. The topic of developing efficient algorithms for treewidth-bounded problem instances has been of interest to Dan since the 1990's, and he has published many papers on that topic (e.g. [6, 7, 19]).

The next two chapters are on transaction processing in concurrent databases. As mentioned earlier, this is a topic to which Dan has made several seminal contributions. Chapter 12 ("An Optimistic Concurrency Control Protocol for Replicated Databases") by Yuri Breitbart, Henry Korth and Abraham Silberschatz discusses a concurrency control protocol that guarantees serializability and freedom from deadlock for multi-site transactions. An important characteristic of the protocol is that it does not rely on any special properties of the database systems running at the individual sites. Further, the new protocol reduces the communication overhead needed to achieve serializability and deadlock freedom. Breitbart et al. compare their approach with the approach presented in one of Dan's papers on the topic [10].

Chapter 13 ("SNAPSHOT Isolation: Why do Some People Call it SERIALIZABLE?") by Philip M. Lewis is based on his popular lecture entitled "Why Does Oracle Make Fun of Theoreticians?". He points out that several commercial database systems use a relaxed consistency requirement called **snapshot isolation** (instead of serializability) for running concurrent transactions. It is known that snapshot isolation can result in nonserializable schedules. However, users of such systems do not complain about getting incorrect results from their transactions. The chapter offers a possible explanation for this phenomenon: those transactions are based on certain design patterns for which snapshopt isolation is sufficient to produce correct results. The chapter also presents an example of such a design pattern.

Chapter 14 ("A Richer Understanding of the Complexity of Election Systems") by Piotr Faliszewski, Edith Hemaspaandra, Lane Hemaspaandra and Jörg Rothe presents a detailed survey of recent results on the complexity of various election systems. They point out that in addition to political elections, the topic of voting arises naturally in a number of other contexts such as spam detection, web search engines, etc. They consider the effect of control, manipulation and bribery on the complexity of the underlying election problems. The chapter provides the necessary background on various election systems and an outline of the techniques used to establish the complexity results. The chapter also includes an extensive list of references. Dan has studied the complexity of problems arising in many different areas since the mid 1970's and has published extensively on the topic (e.g. [9, 14–18]).

The next two chapters deal with different forms of approximation algorithms for NP-hard problems, a topic in which Dan has had an active interest since the 1970's. Chapter 15 ("Fully Dynamic Bin Packing") by Zoran Ivković and Errol L. Lloyd considers the **fully dynamic** version of the online bin packing problem. In that version, requests to insert and delete new items arrive one at a time, and the online algorithm is required to maintain a packing with a small number of bins. To process a request, the online algorithm is allowed to do a limited amount of repacking. The performance of such an algorithm is measured by its **competitive ratio**, which is the worst-case ratio of the number of bins used by the algorithm to the minimum number of bins used by an optimal offline algorithm. The authors consider several variants of the problem and establish lower and upper bounds on achievable competitive ratios.

Chapter 16 ("Online Job Admission") by Sven O. Krumke, Rob van Stee and Stephen Westphal addresses an online scheduling problem. In this problem, each job has a release time and execution time which are revealed to the online algorithm only when the job arrives. A time horizon T is specified and the jobs must be scheduled up to time T in a nonpreemptive fashion on a given number of processors. As each job arrives, the online algorithm must make a decision whether to accept or reject the job without any knowledge of the future jobs. The goal is to accept a subset of jobs such that the total execution time of the accepted jobs is close to the maximum possible value obtainable using an optimal offline algorithm. The authors first present a lower bound on the achievable competitive ratio. Then they present deterministic and randomized algorithms which achieve competitive ratios that are close to the lower bound.

Chapter 17 ("A Survey of Graph Algorithms Under Extended Streaming Models of Computation") by Thomas C. O'Connell summarizes many known results on graph algorithms when the input to the algorithm is in the form of a data stream; thus, an algorithm can only make one pass over the entire input. In addition, there is a restriction on how much of the stream data can be stored by the algorithm. Using ideas from communication complexity [20], many natural graph problems have been shown to be inherently difficult under this streaming model. Therefore, researchers have proposed extensions of the streaming model under which one can solve some of the graph problems. The chapter provides descriptions of the various extensions and outlines known algorithms for several graph problems (e.g. finding connected components, computing shortest paths) under those models. Thomas O'Connell was a Ph.D. student in the Computer Science Department at UAlbany when Dan served as the Chair of the department.

Chapter 18 ("Interactions Among Human Behavior, Social Networks and Societal Infrastructures: A Case Study in Computational Epidemiology") was contributed by a group of researchers (Christopher Barrett, Keith Bisset, Jiangzhuo Chen, Stephen Eubank, Bryan Lewis, V. S. Anil Kumar, Madhav Marathe and Henning Mortveit) at the Network Dynamics and Simulation Science Laboratory (NDSSL), which is a part of the Virginia Bioinformatics Institute and Virginia Tech. This chapter gives an overview of the ongoing research on large scale simulations and computational epidemiology at NDSSL. The issues addressed by this research are extremely important in practice. To address those issues, ideas and techniques from a number of different fields (e.g. Computer Science, Mathematics, Biology, Sociology) are needed. The focus of the chapter is on some questions that are of interest to Computer Science researchers. Dan has been involved in joint research with several members of the NDSSL group for many years, and this collaboration has led to a number of publications (e.g. [3–7]).

Acknowledgments

We thank all the authors who contributed chapters for the second part of this book. We also thank ACM, SIAM and Elsevier for granting permission to reprint Dan's journal papers for the first part of this book. We thank Springer for publishing this book. In particular, we are grateful to Mark DeJongh and Cindy Zitter of Springer for their help and patience.

We owe a special thanks to Veronica Lee, who created latex source files for many of the papers that appear in Part I of this book. Without her help, it would have taken us even longer to finish this book.

One of the editors (Ravi) thanks Nandaki Systems and Glomantra e-Services (P) Limited (both located in Bangalore, India) for allowing him to use their networking facilities to work on this book while he was visiting India.

Last but not the least, we express our sincere thanks to Dan, who has been an exceptional colleague and mentor. We have benefitted significantly from his extensive knowledge, tremendous enthusiasm for research, exceptional problem solving skills and high ethical standards. It has been our pleasure and privilege to work with him. We hope to continue to work with him for many years to come.

References

[1] A. V. Aho and J. D. Ullman. *Principles of Compiler Design*. Addison–Wesley, Reading, 1977.

[2] N. Bansal, R. Khandekar, and V. Nagarajan. Additive guarantees for degree bounded directed network design. In *Proc. 40th annual ACM Symp. Theory of Computing*, pages 769–778, May 2008.

[3] C. L. Barrett, H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. Analysis problems for sequential dynamical systems and communicating state machines. In *MFCS*, pages 159–172, 2001.

[4] C. L. Barrett, H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. Predecessor and permutation existence problems for sequential dynamical systems. In *DMCS*, pages 69–80, 2003.

[5] C. L. Barrett, H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. Complexity of reachability problems for finite discrete dynamical systems. *J. Comput. Syst. Sci.*, 72(8):1317–1345, 2006.

[6] C. L. Barrett, H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, R. E. Stearns, and M. Thakur. Computational aspects of analyzing social network dynamics. In *IJCAI*, pages 2268–2273, Jan. 2007.

[7] C. L. Barrett, H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, R. E. Stearns, and M. Thakur. Predecessor existence problems for finite discrete dynamical systems. *Theor. Comput. Sci.*, 386(1-2):3–37, 2007.

[8] P. Bernstein, V. Hadzilacos, and N. Goodman. *Concurrency Control and Recovery in Database Systems*. Addison–Wesley, Reading, 1987.

[9] P. A. Bloniarz, H. B. Hunt III, and D. J. Rosenkrantz. Algebraic structures with hard equivalence and minimization problems. *J. ACM*, 31(4):879–904, 1984.

[10] P. Chundi, D. Rosenkrantz, and S. S. Ravi. Deferred updates and data placement in distributed databases. In *Proc. International Conference on Data Engineering*, pages 469–476, New Orleans, LA, Feb.–Mar. 1996.

[11] S. Greibach. A new normal-flow theorem for context-free phrase structure grammars. *J. ACM*, 12(1):42–52, Jan. 1965.

[12] M. A. Harrison. *Introduction to Formal Language Theory*. Addison–Wesley, Reading, 1978.

[13] P. M. Lewis II, D. J. Rosenkrantz, and R. E. Stearns. *Compiler Design Theory*. Addison–Wesley–Longman, Boston, 1976.

[14] H. B. Hunt III and D. J. Rosenkrantz. On equivalence and containment problems for formal languages. *J. ACM*, 24(3):387–396, 1977.

[15] H. B. Hunt III and D. J. Rosenkrantz. The complexity of testing predicate locks. In *SIGMOD Conference*, pages 127–133, May 1979.

[16] H. B. Hunt III and D. J. Rosenkrantz. The complexity of monadic recursion schemes: Executability problems, nesting depth, and applications. *Theor. Comput. Sci.*, 27:3–38, 1983.

[17] H. B. Hunt III and D. J. Rosenkrantz. The complexity of monadic recursion schemes: Exponential time bounds. *J. Comput. Syst. Sci.*, 28(3):395–419, 1984.

[18] H. B. Hunt III, D. J. Rosenkrantz, and P. A. Bloniarz. On the computational complexity of algebra on lattices. *SIAM J. Comput.*, 16(1):129–148, 1987.

[19] H. B. Hunt III, M. V. Marathe, V. Radhakrishnan, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. NC-approximation schemes for NP- and PSPACE-Hard problems for geometric graphs. *J. Algorithms*, 26(2):238–274, 1998.

[20] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, New York, 1996.

[21] T. Lengauer. Exploiting hierarchy in VLSI design. In F. Makedon et al., editor, *Proc. Aegean Workshop on Computing*. Lecture Notes in Computer Science, volume 227, pages 180–193. Springer, Berlin, 1986.

[22] C. Papadimitriou. *The Theory of Database Concurrency Control*. Computer Science Press, Rockville, 1986.

[23] R. Ravi, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and H. B. Hunt III. Many birds with one stone: multi-objective approximation algorithms. In *Proc. 25th Annual ACM symposium on Theory of Computing*, pages 438–447, May 1993.

[24] D. J. Rosenkrantz. JACM 1986–1990. *J. ACM*, 50(1):18, 2003.

[25] R. E. Stearns, P. M. Lewis II, and D. J. Rosenkrantz. Concurrency controls for database systems. In *Proc. 17th Annual Symp. on Foundations of Comptr. Sci.*, pages 19–32, Houston, Texas, Oct. 1976.

Springer

http://www.springer.com/978-1-4020-9687-7

Fundamental Problems in Computing
Essays in Honor of Professor Daniel J. Rosenkrantz
Ravi, S.S.; Shukla, S.K. (Eds.)
2009, XXII, 516 p., Hardcover
ISBN: 978-1-4020-9687-7