# Guest Editorial Special Issue: “From Deletion-Correction to Graph Reconstruction: In Memory of Vladimir I. Levenshtein”

THERE are few mathematicians whose contributions go beyond named conjectures and theorems: Vladimir Iosifovich Levenshtein (Владимир Иосифович Левенштейн, 1935-2017) is one such true exception. During the five decades of his active research career, he enriched combinatorics, coding, and information theory with elegant problem formulations, ingenious algorithmic solutions, and highly original proof techniques. However, his work accomplished much more—it paved the way for the creation and advancement of new scientific disciplines, such as natural language processing, metagenomics, sequence alignment, and reference-based genome assembly, as well as DNA-based data storage, to name a few. A crucial concept behind sequence alignment algorithms used in phylogeny, comparative, and cancer genomics, as well as in natural language processing is the Levenshtein (edit) distance and its extension, termed the Damerau–Levenshtein distance between strings. The Levenshtein distance equals the smallest number of insertions, deletions, or substitutions required to convert one string into another. Levenshtein introduced this metric in 1965 [item 1) in the Appendix], followed by the notion of deletion and insertion error-correcting codes that have since been used in a myriad of systems presented with synchronization errors [items 1) and 2) in the Appendix]. Levenshtein’s work also inspired the introduction of the trace reconstruction problem [items 3) and 4) in the Appendix] which has since sparked substantial interest in the field of DNA-based data storage.

Levenshtein’s early research papers strongly impacted the field of coding theory which in the early 1960s was still at a crossroad. He established that lexicodes, for a special choice of parameters, are linear codes [item 5) in the Appendix]; provided the first construction of a family of codes based on Hadamard matrices that achieve the Plotkin bound [item 6) in the Appendix]; derived upper-bounds on the size of comma-free codes and proposed explicit constructions for the same [item 7) in the Appendix]; and derived new bounds for constant-weight codes [item 8) in the Appendix]. This list is not exhaustive and only serves to illustrates the breadth of his interests and inherent understanding of what would be relevant topics in coding theory for years to come.

During the next stage of his career, Levenshtein published a number of outstanding results in extremal combinatorics, sphere packings, and discrete geometry [items 9)–11) in the Appendix] that have deeply impacted discrete mathematics and coding theory. Jointly with Kabatyansky, he established the best bounds on the largest possible packing density of spheres in Euclidean spaces, and in a follow-up work, he proved that the E8 and Leech lattice have the optimal “kissing configuration” for their given dimensions. The 1980s and 1990s also witnessed the publication of another collection of exceptional results focused on association schemes and orthogonal polynomials (Krawtchouk polynomials and what is now known as Levenshtein polynomials), made famous by the seminal work of Delsarte [items 12)–17) in the Appendix].

In the later part of his life, Levenshtein devoted himself to the study of emerging topics such as group testing, constrained coding, and graph reconstruction. The readers of this Special Issue are referred to a more comprehensive and technical account of these and other works by Levenshtein published in the *Information Theory Newsletter* in 2018.

Those of us who were influenced by the power and breadth of Levenshtein’s work, as well as those that he mentored or shared the bonds of friendship with, felt it necessary to come together to honor his life and legacy through a Special Issue of IEEE TRANSACTIONS ON INFORMATION THEORY. Despite the Covid19 pandemics, we managed to solicit seven invited contributions comprising a mix of reviews and original results and to select eighteen additional contributed papers for inclusion in the Special Issue.

The invited papers address Levenshtein’s most impactful lines of work in coding theory, combinatorics, and applications of the Levenshtein distance and string reconstruction techniques in computational biology.

The articles “Synchronization strings and codes for insertions and deletions—A survey,” by Haeupler and Shahrasbi, and “An overview of capacity results for synchronization channels,” by Cheragchi and Ribeiro, provide highly readable yet extensive reviews of the history of coding for indel error-correction, capacity results for deletion channels, and the state-of-the-art solutions in the field. The article “Trace reconstruction: Generalized and parameterized,” by Krishnamurthy *et al.*, complements the two aforementioned reviews by focusing on trace reconstruction problems and it also presents a cohort of new results extending the scope of the area of string reconstruction.

The article “Reed–Muller codes: Theory and algorithms,” by Abbe *et al.*, covers the most recent developments regarding the computation of the weight enumerators and capacity-achieving properties of Reed–Muller codes, their connec-

Digital Object Identifier 10.1109/TIT.2021.3072555

tions to polar codes as well as accompanying algorithmic developments.

The article "On reconstruction of graphs from the multiset of subgraphs obtained by deleting $l$ vertices," by Kostochka and West, describes state-of-the-art results pertaining to Ulam's reconstruction conjecture for graphs and it also reviews Levenshtein's contributions to the field.

The articles "Levenshtein distance, sequence comparison, and biological database search," by Berger *et al.*, and "Trace reconstruction problems in computational biology," by Bhardwaj *et al.*, are coauthored by luminaries in the field of computational biology. Researchers in information and coding theory will greatly benefit from reading their abridged reviews of the fields of sequence alignment, and, in particular, dynamic programming and the BLAST toolkit, genomic pattern matching, and database search as well as emerging applications of Levenstein's distance and trace reconstruction techniques in immunogenomics and molecular data storage.

The 18 contributed papers selected for inclusion present new results on varied topics in coding for synchronization and DNA-based data storage channels, extremal coding theory, sequence design, and reconstruction, as well as Boolean functions.

Seven papers address fundamental open problems and new formulations and applications of indel error-correcting codes and trace reconstruction.

The article "Perfect multi deletion codes achieve the asymptotic optimality of code size," by Mori and Hagiwara, investigates the tightness of Levenshtein's upper bound for t-deletion codes under the assumption that perfect t-deletion binary codes exist. The main result is a proof that the bound is tight for any t and achievable by perfect t-deletion codes.

"On Levenshtein's channel and list size in information retrieval," by Junnila *et al.*, considers a trace reconstruction problem where the goal is not to uniquely reconstruct the actually transmitted string but rather to generate a list of a certain size that contains the string. The authors provide new bounds on the size of the list and improve upon existing work in this area.

The article "Multilayer codes for synchronization from deletions and insertions," by Abroshan *et al.*, considers a synchronization problem in which transmitted and received sequences are assumed to be at a small Levenshtein distance. It proposes a synchronization method based on auxiliary error-free short messages using Varshamov–Tenengoltz codes.

The article "On optimal $k$-deletion correcting codes," by Sima and Bruck, addresses the challenging problem of designing low-redundancy $k$-deletion error-correcting codes and provides explicit constructions with low encoding and decoding complexity and redundancy that complements and in certain cases improves the best known nonconstructive techniques.

The article "Covering codes using insertions or deletions," by Lenz *et al.*, considers the problem of designing covering codes under the Levenshtein distance. The authors present upper and lower bounds and show that the asymptotic rate of the resulting codes matches the best-known bounds for covering codes under the Hamming distance.

The article "Algorithms for reconstruction over single and multiple deletion channels," by Srinivasavaradhan *et al.*, investigates the problem of transmitting a single sequence over single and multiple deletion channels under the modeling assumption that perfect reconstruction of the input sequence may not be possible. Under this setup, the authors show that solving for the maximum likelihood estimate over the single deletion channel is equivalent to its relaxation, which is a continuous optimization problem.

The article "Efficient multiparty interactive coding—Part I: Oblivious insertions, deletions and substitutions," by Gelles *et al.*, addresses a problem in the area of interactive coding and provides an efficient, constant rate scheme that conducts any computation on an arbitrary network that succeeds with high probability in the presence of a predetermined number of insertion–deletion errors. This is the first such approach known for multiparty communication scheme.

Two other contributions focus on coding theory methods for DNA-based data storage, in the context of deletion and tandem error-correction.

The article "Correcting a single indel/edit for DNA-based data storage: Linear-time encoders and order-optimality," by Cai *et al.*, investigates quarternary codes that can correct either a single indel or a single edit and provide linear-time encoding algorithms.

The article "Error-correcting codes for noisy duplication channels," by Tang and Farnoud, considers codes for storage in living cells and addresses noise models termed uniform tandem-duplications. It addresses a special choice of duplication parameters and performs recovery of what is known as the duplication root of the stored codeword in an asymptotically optimal manner.

Three contributed papers address coding paradigms centered around Boolean functions and specialized Boolean functions, their constructions, and applications.

The article "How to construct mutually orthogonal complementary sets with non-power-of-two lengths?" by Wu *et al.*, considers the problem of designing mutually orthogonal complementary sets based on generalized Boolean functions for certain practical OFDM systems which require nonstandard codelengths.

The article "Cyclic bent functions and their applications in sequences," by Abdukhalikov *et al.*, establishes a link between quadratic cyclic bent functions and a special type of prequasifields, and presents a class of quadratic cyclic bent functions from the Kantor–Williams prequasifields. It then proposes the use of cyclic bent functions to construct families of optimal sequences for use in CDMA communication.

The article "Further study of 2-to-1 mappings over $F_{2^n}$," by Li *et al.*, addresses the problem of characterizing 2-to-1 mappings over finite fields with characteristics equal to two. The main result is a complete characterization of 2-to-1 polynomials of degree five over fields with characteristics equal to two.

The remaining seven contributions to the Special Issue address problems pertaining to a variety of topics related to Levenshtein's research.

The article "New construction of optimal type-II binary Z-complementary pairs," by Gu *et al.*, provides constructions of several optimal binary sequences with good correlation properties and a new design method for binary sequences with flexible parameter choices.

The article “A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres,” by Kirshner and Samorodnitsky, derives a new, powerful inequality for $l_p$ norms of polynomials over the Hamming space, and shows that it is asymptotically nearly tight for Krawtchouk polynomials. It also establishes a number of new properties of Krawtchouk polynomials.

The article “Treeplication: An erasure code for distributed full recovery under the random multiset channel,” by Gandelman and Cassuto, considers a novel coding framework for distributed storage in which data is recovered in a distributed manner. The proposed construction uses a tree-like redundancy structure, and has a full-recovery communication-cost comparable to that of replication.

The article “Multichannel conflict-avoiding codes of weights three and four,” by Lo *et al.*, addresses the topic of conflict-avoiding codes, introduced by Levenshtein as a single-channel transmission scheme for a multiple-access collision channel without feedback. The authors extend this paradigm for multichannel models and describe upper bounds on the number of codewords and optimal constructions for multichannel conflict-avoiding codes of weights three and four.

The article “Universal bounds for size and energy of codes of given minimum and maximum distances,” by Boyvalenkov *et al.*, establishes Levenshtein-type upper bounds on the cardinality of codes with given minimum and maximum distances, as well as universal lower bounds on the potential energy of codes with a given maximum distance and cardinality. In their derivations, the authors use Levenshtein-type quadrature formulas to establish the distance distributions of codes that meet the derived upper bounds.

The article “On multifold packings of radius-1 balls in Hamming graphs,” by Krotov and Potapov, is concerned with $\lambda$-fold $r$-packings in Hamming spaces, i.e., codes with the property that balls of radius $r$ with centers from the code cover each point of the space not more than $\lambda$ times. The main results of the paper pertain to upper bounds on the maximum size of twofold 1-packings and derivations of new properties of 1-perfect unitrades.

We sincerely hope that this Special Issue and its invited and contributed papers will renew the interest in some of Levenshtein’s lesser known work and maintain the broad fascination with his pioneering research for decades to come.

ALEXANDER BARG, *Guest Editor*
University of Maryland
College Park, MD 20742 USA

LARA DOLECEK, *Guest Editor*
University of California at Los Angeles
Los Angeles, CA 90095 USA

RYAN GABRYS, *Guest Editor*
Naval Information Warfare Center
San Diego, CA 92110 USA

GYULA O. H. KATONA, *Guest Editor*
Renyi Institute
1145 Budapest, Hungary

JÁNOS KÖRNER, *Guest Editor*
Department of Computer Science
La Sapienza University of Roma
00198 Rome, Italy

ANDREW MCGREGOR, *Guest Editor*
College of Information and Computer Sciences
University of Massachusetts
Amherst, MA 01003-9264 USA

OLGICA MILENKOVIC, *Guest Editor-in-Chief*
University of Illinois at Urbana–Champaign
Urbana, IL 61801 USA

SIHEM MESNAGER, *Guest Editor*
Department of Mathematics
University of Paris VIII
F-93526 Saint-Denis, France

Laboratory Geometry, Analysis and Applications (LAGA)
University of Paris XIII Sorbonne Paris Nord, CNRS
UMR 7539
F-93430 Villetaneuse, France

Telecom Paris
Polytechnic Institute
91120 Palaiseau, France

GILLES ZÉMOR, *Guest Editor*
Institut de Mathématiques de Bordeaux Université de Bordeaux
33400 Talence, France

## APPENDIX
## RELATED WORK

1) V. I. Levenshtein, “Binary codes capable of correcting spurious insertions and deletions of ones,” (in Russian), *Problemy Peredachi Informatsii*, vol. 10, no. 8, pp. 12–25, 1965.
2) V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals,” (in Russian), *Doklady Akademii Nauk SSSR*, vol. 163, no. 4, pp. 845–848, 1965.
3) V. I. Levenshtein, “Efficient reconstruction of sequences,” *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 2–22, Jan. 2001.
4) V. I. Levenshtein, “Efficient reconstruction of sequences from their subsequences or supersequences,” *J. Combinat. Theory A*, vol. 93, no. 2, pp. 310–332, Feb. 2001.
5) V. I. Levenshtein, “A class of systematic codes,” *Doklady Akademii Nauk SSSR*, vol. 131, no. 5, pp. 1011–1014, 1960.
6) V. I. Levenshtein, “The application of Hadamard matrices in coding theory,” *Problemy Kibernetiki*, vol. 5, pp. 125–136, 1964.
7) V. I. Levenshtein, “Combinatorial problems motivated by comma-free codes,” *J. Combinat. Des.*, vol. 12, no. 3, pp. 184–196, 2004.
8) V. I. Levenshtein, “On upper bounds for codes with codewords of a fixed weight,” *Problemy Peredachi Informatsii*, vol. 7, no. 4, pp. 3–12, 1971.
9) G. A. Kabatyansky and V. I. Levenshtein, “Bounds for packings on the sphere and in the space,” *Problemy Peredachi Informatsii*, vol. 14, no. 1, pp. 3–25, 1978.
10) V. I. Levenshtein, “On the choice of polynomials for obtaining bounds in packing problems,” in *Proc. 7th All-Union Conf. Coding Inform. Theory*, 1978, pp. 103–108.
11) V. I. Levenshtein, “On bounds for packings in N-dimensional Euclidean space,” *Doklady Akademii Nauk SSSR*, vol. 245, no. 6, pp. 1299–1303, 1979.
12) V. I. Levenshtein, “Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces,” *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1303–1321, Sep. 1995.
13) V. I. Levenshtein, “Designs as maximum codes in polynomial metric spaces,” *Acta Applicandae Mathematicae*, vol. 29, nos. 1–2, pp. 1–82, Nov. 1992.
14) V. Levenshtein, “On designs in compact metric spaces and a universal bound on their size,” *Discrete Math.*, vol. 192, nos. 1–3, pp. 251–271, Oct. 1998.
15) V. Levenshtein, “Equivalence of Delsarte’s bounds for codes and designs in symmetric association schemes, and some applications,” *Discrete Math.*, vols. 197–198, nos. 1–3, pp. 515–536, Feb. 1999.
16) V. Levenshtein, “Universal bounds for codes and designs,” in *Handbook Coding Theory*, vol. 1, V. Pless and W. C. Huffman, Eds. Amsterdam, The Netherland: Elsevier, 1998, pp. 499–648.
17) P. Delsarte and V. I. Levenshtein, “Association schemes and coding theory,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2477–2504, Oct. 1998.