Sergey Yekhanin

Curriculum Vitae

November 29, 2018

Microsoft Research Redmond
One Microsoft Way, Redmond, WA 98052
Homepage: yekhanin.org

Office tel.: (650) 693-2941
E-mail: yekhanin@microsoft.com
Personal: yekhanin@gmail.com

Current interests: Differential privacy, Coding for DNA storage, Coding for distributed storage.

Employment

Principal Researcher	*Microsoft Research, Redmond Lab.*	*9/2018 – present*
Senior Researcher	*Microsoft Research, Redmond Lab.*	*3/2015 – 9/2018*
Researcher	*Microsoft, Windows Azure Storage.*	*9/2014 – 3/2015*
Researcher	*Microsoft Research, Silicon Valley Lab.*	*4/2008 – 9/2014*
Member of School of Mathematics	*Institute for Advanced Study, Princeton.*	*9/2007 – 4/2008*

Education

Ph.D. in Computer Science.

2003-2007 *Massachusetts Institute of Technology*, Cambridge, MA, USA.
Thesis advisor: Prof. Madhu Sudan.
Ph.D. thesis*: Locally Decodable Codes and Private Information Retrieval Schemes.*

Diploma in Computer Science, 2002

1997-2002 *Moscow State University*, Moscow, Russia.
Thesis advisor: Prof. Alexander Petrenko.
Area of research: formal methods.

Honors:

- Microsoft Privacy FY17 Technical Excellence Award, 2017.
- IEEE Communications Society & Information Theory Society Joint Paper Award, 2014.
- Invited speaker at International Congress of Mathematicians (ICM), 2014.
- Microsoft Technical Community Network (TCN) Storage Technical Award, 2013.
- Best Paper Award - USENIX Annual Technical Conference (USENIX ATC), 2012.
- ACM Doctoral Dissertation Award, 2007.
- MIT George M. Sprowls Award for the best doctoral theses in Computer Science, 2007.
- Best Paper Award - ACM Symposium on the Theory of Computing (STOC), 2007.
- Best Student Paper Award - ACM Symposium on the Theory of Computing (STOC), 2007.
- MIT Presidential Fellow, 2003.
- Diploma with Honors from Moscow State University, 2002.
- Soros fellowship in mathematics: 1999.

Industry impact:

- In early 2010s, my collaborators and I designed a new class of space efficient erasure correcting codes (local reconstruction codes) for applications in distributed storage. Our codes are currently deployed in several Microsoft products including: Microsoft Azure, Windows, and Windows Server, providing substantial monetary savings to Microsoft.
- In 2017, my collaborators and I designed locally differentially private algorithms that are now used in Windows for telemetry collection, allowing Microsoft to learn general trends in user population while protecting the privacy of individuals.
- In 2017, I have been a part of the team that set the world record by storing (and successfully retrieving) over 200MB of digital data in synthetic DNA molecules. My focus has been on design and implementation of the codec that deals with errors that occur during DNA synthesis, sequencing, and storage.

Special lectures:

- Invited plenary talk, ACM-SIAM Symp. On Discrete Algorithms (SODA), San Diego, 2015.
- Invited lecture, Intern. Congress of Mathematicians (ICM), Seoul, South Korea, 2014.
- Invited lecture, Theory of Cryptography Conference (TCC), Taormina, Italy, 2012.
- Invited lecture, Computer Science in Russia conference (CSR), St. Petersburg, Russia, 2011.

Books:

- Sergey Yekhanin, ***Locally Decodable Codes***. FnT-TCS, Now publishers, 2012.
- Sergey Yekhanin, ***Locally Decodable Codes and Private Information Retrieval Schemes***, Springer, 2010. (Winning thesis of the 2007 ACM doctoral dissertation award).

Journal papers:

1. *Random access in large-scale DNA data storage* (with Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Konstantin Makarychev, Miklos Z. Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, Christopher N. Takahashi, Sharon Newman, Hsing-Yeh Parker, Cyrus Rashtchian, Kendall Stewart, Gagan Gupta, Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze, Karin Strauss). Nature Biotechnology, vol. 36, pp. 242-248, 2018.

2. *Explicit maximally recoverable codes with locality* (with Parikshit Gopalan, Cheng Huang, Bob Jenkins). IEEE Transactions on Information Theory, vol. 60, issue 9, pp. 5245-5256, 2014.

3. *Kolmogorov width of discrete linear spaces: an approach to matrix rigidity* (with Alex Samorodnitsky, Ilya Shkredov). Computational Complexity, vol. 25, pp. 309-348, 2016.

4. *On the locality of codeword symbols in non-linear codes* (with Michael Forbes). Discrete Mathematics, vol. 324, pp. 78-84, 2014.

5. *A note on the Newton radius* (with Alex Samorodnitsky). Discrete Mathematics, vol. 312, issue 15, pp. 2392-2393, 2012.

6. *On the locality of codeword symbols* (with Parikshit Gopalan, Cheng Huang, Huseyin Simitci). IEEE Transactions on Information Theory, vol. 58, issue 11, pp. 6925-6934, 2012.

7. *Matching vector codes* (with Parikshit Gopalan, Zeev Dvir). SIAM Journal on Computing, vol. 40, issue 4, pp. 1154-1178, 2011.

8. *Sets with large additive energy and symmetric sets* (with Ilya Shkredov). Journal of Combinatorial Theory Series A., vol. 118, issue 3, pp. 1086-1093, 2011.

9. *Private information retrieval.* Communications of the ACM, vol. 53, issue 4, pp. 68-73, 2010.

10. *Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers* (with Kiran Kedlaya). SIAM Journal on Computing, vol. 38, Issue 5, pp. 1952-1969, 2009.

11. *Towards 3-query locally decodable codes of subexponential length.* Journal of the ACM, vol. 55, Issue 1, pp. 1-16, 2008.

12. *An Ω(n^{1/3}) lower bound for bilinear group based private information retrieval* (with Alexander Razborov*)*. Theory of Computing, vol. 3, pp. 221-238, 2007.

13. *A geometric approach to information theoretic private information retrieval* (with David Woodruff). SIAM Journal on Computing, vol. 37, Issue 4, pp. 1046-1056, 2007.

14. *A note on plane pointless curves*. Finite Fields and Their Applications, vol. 13, Issue 2, pp. 418-422, 2007.

15. *Long nonbinary codes exceeding the Gilbert - Varshamov bound for any fixed distance* (with Ilya Dumer). IEEE Transactions on Information Theory, vol. 50, Issue 10, pp. 2357-2362, 2004.

16. *Improved upper bound for the redundancy of fix-free codes.* IEEE Transactions on Information Theory, vol. 50, Issue 11, pp. 2815-2818, 2004.

17. *On application of the partition distance concept to a comparative analysis of psychological or sociological tests* (with Arkadii D'yachkov, Vyacheslav Rykov, David Torney). Stochastic Analysis and Applications, vol. 24, pp. 61-78, 2006.

18. *Trivial two-stage group testing for complexes using almost disjunct matrices* (with Anthony J. Macula, Vyacheslav Rykov). Discrete and Applied Mathematics, vol. 137, pp. 97-107, 2004.

19. *Evaluation of estimates for standard learning information in pattern recognition problems* (with Anna Kochetova). Computational Mathematics and Mathematical Physics, vol. 42, N3, pp. 419-423, 2002.

Conference papers:

1. *Maximally recoverable LRCs: a field size lower bound and constructions for few heavy parities* (with Sivakanth Gopi and Venkatesan Guruswami). Proceedings of ACM-SIAM symposium on discrete algorithms (SODA), 2019.

2. *Collecting telemetry data privately* (with Bolin Ding, Janardhan Kulkarni). Proceedings of advances in neural information processing systems (NIPS), pp. 3574-3583, 2017.

3. *Clustering billions of reads for DNA storage* (with Cyrus Rashtchian, Konstantin Makarychev, Miklos Racz, Siena Dumas Ang, Djordje Jevjic, Luis Ceze, Karin Strauss). Proceedings of advances in neural information processing systems (NIPS), pp. 3362-3373, 2017.

4. *Maximally recoverable codes for grid-like topologies* (with Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang). Proceedings of ACM-SIAM symposium on discrete algorithms (SODA), pp. 2092-2108, 2017.

5. *New constructions of MR and SD codes over small finite fields*. (with Guangda Hu) Proceedings of international symposium on information theory (ISIT), pp. 1591-1595, 2016.

6. *Kolmogorov width of discrete linear spaces: an approach to matrix rigidity* (with Alex Samorodnitsky, Ilya Shkredov). Proceedings of the 30th computational complexity conference (CCC), pp. 347-364, 2015.

7. *Extending memory lifetime by reviving dead blocks* (with Rodolfo Azevedo, John D. Davis, Karin Strauss, Parikshit Gopalan, Mark Manasse). Proceedings of the 40-th international symposium on computer architecture (ISCA), pp. 452-463, 2013.

8. *Erasure coding in Windows Azure Storage* (with Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li). Proceedings of the 2012 USENIX annual technical conference (USENIX ATC), 2012.

9. *Noisy interpolation of sparse polynomials, and applications* (with Shubhangi Saraf). Proceedings of the 26th computational complexity conference (CCC), pp. 86-92, 2011.

10. *Locally decodable codes: a brief survey*. Proceedings of the 3rd international workshop on coding and cryptography (IWCC), pp. 173-282, 2011.

11. *High-rate codes with sublinear-time unique decoding* (with Swastik Kopparty, Shubhangi Saraf). Proceedings of the 43rd ACM symposium on the theory of computing (STOC), pp.167-176, 2011.

12. *Pan-private streaming algorithms* (with Cynthia Dwork, Moni Naor, Toni Pitassi, Guy Rothblum). Proceedings of the 1st symposium on innovations in computer science (ICS), pp. 66-80, 2010.

13. *New efficient attacks on statistical disclosure control mechanisms* (with Cynthia Dwork). Proceedings of the 28th international cryptology conference (Crypto), pp. 469-480, 2008.

14. *Deterministic approximation algorithms for the nearest codeword problem* (with Noga Alon, Rina Panigrahy). Proceedings of the 13th international workshop on randomization and computation (Random), pp. 339-351, 2009.

15. *Detecting rational points on hypersurfaces over finite fields* (with Swastik Kopparty). Proceedings of the 23rd computational complexity conference (CCC), pp. 311-320, 2008.

16. *Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs* (with Nicholas Harvey, Mihai Patrascu, Yonggang Wen, Vincent W. S. Chan). Proceedings of the 26th IEEE conference on computer communications (INFOCOM), 2007.

17. *On the hardness of matrix completion* (with Nicholas Harvey, David Karger). Proceedings of ACM-SIAM symposium on discrete algorithms (SODA), pp.1103-1111, 2006.

18. *Secure biometrics via syndromes* (with Emin Martinian, Jonathan S. Yedidia). Proceedings of the Allerton conference on communication, control, and computing, 2005.

19. *Upper bound on the rate of superimposed (s,l)-codes based on Engel's inequality* (with Arkadii D'yachkov, Pavel Vilenkin). Proceedings of international conference on algebraic and combinatorial coding theory (ACCT), pp. 95-99, 2002.

20. *Sufficient conditions of existence of fix-free codes*. Proceedings of international symposium on information theory (ISIT), p. 284, 2001.

21. *Cover-free families and superimposed codes: constructions, bounds and applications to cryptography and group testing* (with Arkadii D'yachkov, Vladimir Lebedev, Pavel Vilenkin). Proceedings of international symposium on information theory (ISIT), p. 117, 2001.

22. *New results in the theory of superimposed codes* (with Arkadii D'yachkov, Anthony Macula, David Torney, Pavel Vilenkin). Proceedings of international conference on algebraic and combinatorial coding theory (ACCT), pp. 126-136, 2000.

23. *Some new constructions of optimal superimposed designs*. Proceedings of international conference on algebraic and combinatorial coding theory (ACCT), pp. 232-235, 1998.

Patents:

- *Biometric based user authentication and data encryption* (with Anthony Vetro, Jonathan S Yedidia, Emin Martinian). US Patent 7,620,818, granted November 17, 2009.
- *High rate locally decodable codes* (with Swastik Koppaty, Shubhangi Saraf). US Patent 9,141,679, granted September 22, 2015.
- *Cloud data storage using redundant encoding* (with Parikshit Gopalan, Cheng Huang, Huseyin Simitci). US Patent 8,621,330, granted December 31, 2013.
- *Writing memory blocks using codewords* (with John D. Davis, Parikshit Gopalan, Mark Manasse, Karin Strauss). US Patent 8,972,649, granted March 3, 2015.
- *Local erasure codes for data storage* (with Dennis Fetterly, Parikshit Gopalan, Cheng Huang, Bob Jenkins, Jin Li). US Patent 9,600,365, granted March 21, 2017.
- *Message storage in memory blocks using codewords* (with John D. Davis, Parikshit Gopalan, Mark Manasse, Karin Strauss). US Patent 9,280,417, granted March 8, 2016.
- *Self-identifying memory errors* (with John D. Davis, Parikshit Gopalan, Mark Manasse, Karin Strauss). US Patent. Granted 2017.
- *Erasure coding across multiple zones* (with Brad Calder, Parikshit Gopalan, Cheng Huang, Jin Li, Aaron Ogus, Huseyin Simitci). US Patent 9,378,084, granted June 28, 2016.

- *Erasure coding across multiple zones and sub-zones* (with Brad Calder, Parikshit Gopalan, Cheng Huang, Jin Li, Aaron Ogus, Huseyin Simitci). US Patent 9,244,761, grant Jan. 26, 2016.
- *Extended lifetime memory* (with John D. Davis, Parikshit Gopalan, Mark Manasse, Karin Strauss). US Patent 9,442,799, granted September 17, 2016.
- *Flexible erasure coding schemes with rich structure of local protection groups* (with Brad Calder, Parikshit Gopalan, Cheng Huang, Aaron Ogus, Huseyin Simitci). Patent filed 2015.
- *Erasure coding of data within a group of storage units based on connection characteristics* (with Robert Jenkins, Edmund Nightingale, Cheng Huang, Parikshit Gopalan, Alexander Shamis). US Patent 9,983,959, granted May 29, 2018.
- *Trace reconstruction from noisy polynucleotide sequencer reads* (with Luis Ceze, Siena Dumas Ang, Parikshit Gopalan, Nebojsa Jojic, Miklos Racz, Karin Strauss). Patent filed 2016.
- *Efficient clustering of noisy polynucleotide sequence reads* (with Luis Ceze, Siena Dumas Ang, Ravi Kannan, Konstantin Makarychev, Cyrus Rashtchian, Karin Strauss). Patent filed 2016.
- *Primer design for retrieval of stored polynucleotides* (with Luis Ceze, Yuan-Jyue Chen, Siena Dumas Ang, Karin Strauss). Patent filed 2017.
- *Collection of sensitive data – such as software usage data or other telemetry data – over repeated collection cycles in satisfaction of privacy guarantees* (with Joshua Allen, Bolin Ding, Alex Meade, Janardhan Kulkarni). Patent filed 2017.
- *Hardware protection for differential privacy* (with Joshua Allen, Josh Benaloh, Melissa Chase, Bolin Ding, Janardhan Kulkarni, Jay Lorch, Harsha Nori, Olya Ohrimenko, Srinath Setty). Patent filed 2017.
- *Trace reconstruction from reads with indeterminant errors* (with Miklos Racz). Patent filed 2018.

Teaching Experience:

- *Codes with local decoding*, JTG / IEEE IT summer school, IIT Bombay, May 2017.
- *Codes with local decoding*, Higher School of Economics, mini-course, Moscow, Dec. 2016.
- *Codes with local decoding procedures*, Swedish Summer School in CS, Djuronaset, July 2015.
- *Locally decodable codes*, mini-course, Saint Petersburg computer science club, June 2012.
- *Introduction to Algorithms*, MIT, 2006. (Teaching Assistant).
- *Advanced Algorithms*, MIT, 2005. (Teaching Assistant).
- *Advanced Complexity Theory*, MIT, 2004. (Teaching Assistant).

Program committees:

- 31st Computational complexity conference (CCC), May 2016, Tokyo, Japan.
- 14th Intern. workshop on randomization & comp. (RANDOM), June 2010, Barcelona, Spain.
- 5th Intern. computer science symposium in Russia (CSR), June 2010, Kazan, Russia.
- 24th Computational complexity conference (CCC), July 2009, Paris, France.
- 3rd Intern. computer science symposium in Russia (CSR), June 2008, Moscow, Russia.

Interns mentored:

- *Sankeerth Rao,* UCSD, Summer 2018.
- *Sivakanth Gopi*, Princeton, Fall 2016.

- *Cyrus Rashtchian*, University of Washington, Summer 2016.
- *Govinda Kamath*, Stanford, Summer 2016.
- *Guangda Hu*, Princeton, Fall 2015.
- *Carol Wang*, CMU, Fall 2014.
- *Michael Forbes*, MIT, Summer 2012.
- *Swastik Kopparty*, MIT, Summer 2010.
- *Klim Efremenko*, Weizmann Institute of Science, Summer 2009.

Invited Talks

- EPFL, IC colloquium, Lausanne, November 2017.
- NIST (Nat. Inst. Stand. and Tech.), mini-symp. on molecular based memories, Gaithersburg, Dec. 2016.
- Yandex, Computer Science seminar, Moscow, Russia, December 2016.
- Institute for Information Transmission Problems, Coding Theory seminar, Moscow, December 2016.
- Math., theor. physics, and data science (Y. Sinai and G. Margulis anniv.), Moscow, Russia, July 2016.
- ICERM workshop on algorithmic coding theory, Providence, June 2016.
- SIAM conference on discrete mathematics, Atlanta, June 2016.
- Workshop on Algorithms in Comm. Compl., Property Testing and Comb., Moscow, Russia, April 2016.
- Higher School of Economics, TCS workshop, Moscow, Russia, April 2016.
- University of Washington, Theory seminar, November 2015.
- Information Theory Workshop (ITW), Jeju Island, South Korea, October 2015.
- California Institute of Technology, CS theory seminar, May 2015.
- Simons Institute for the Theory of Computing, Information theory seminar, May 2015.
- DIMACS workshop on coding-theoretic methods for network security, April 2015.
- Rutgers University, CS Theory seminar, April 2015.
- Institute for Advanced Study, Theoretical computer science seminar, April 2015.
- Algebra, Codes, and Networks conference, Bordeaux, France, June 2014.
- Microsoft Research Redmond, Theory day, March 2014.
- New York University, CS Theory seminar, October 2013.
- Princeton University, Computer science departmental colloquium, October 2013.
- Rutgers University, CS Theory seminar, October 2013.
- Allerton Conference on Communication, Control, and Computing, October 2013.
- Institute for Information Transmission Problems, Moscow, September 2013.
- Eighth Int. Conf. on Computability, Complexity and Randomness, Moscow, Russia, September 2013.
- Microsoft Research Redmond, Theory seminar, March 2013.
- Session on Discrete Geom. and Algebraic Comb., National AMS meeting, San Diego, January 2013.
- Session on Advances in Coding Theory, National AMS meeting, Boston, January 2012.
- International Workshop on Coding and Cryptography, Qingdao, China, May 2011.
- University of Washington, Theory seminar, March 2011.
- UCLA, Institute for Pure and Applied Mathematics, March 2011.
- Hebrew University, Theory seminar, February 2011.
- Allerton Conference on Communication, Control, and Computing, October 2010.
- SIAM conference on discrete mathematics, June 2010.
- University of Toronto, Computer science departmental colloquium, April 2010.
- University of Waterloo, Theory seminar, April 2010.
- Institute for Information Transmission Problems, Moscow, February 2010.
- Moscow State University, Number theory seminar, February 2010.
- University of Chicago, Theory seminar, February 2010.
- Microsoft Research New England, TCS seminar, January 2010.
- Dagstuhl seminar on algebraic methods in computational complexity, September 2009.

- Berkeley, Theory Lunch, February 2009.
- Session on Recent Trends in Coding Theory, National AMS meeting, Washington D.C., January 2009.
- Microsoft Research India Lab., Bangalore, December 2008.
- Institute for Information Transmission Problems, Coding Theory Seminar, Moscow, June 2008.
- DIMACS, Privacy workshop, February 2008.
- Carnegie Mellon University, Database privacy workshop, November 2007.
- Rutgers University, Theoretical Computer Science Seminar, October 2007.
- Georgia Inst. of Tech., Algorithms and Randomness Center, Colloquium, September 2007.
- Oberwolfach Workshop on Complexity Theory, June 2007.
- Berkeley, Theory Lunch, April 2007.
- Harvard, Theory of Computation Seminar, April 2007.
- University of Connecticut, Departmental Colloquium, March 2007.
- Northeastern University, College of Comp. and Inform. Science Colloquia, February 2007.
- Institute for Advanced Study, Theoretical Computer Science Seminar, Princeton, November 2006.
- UCLA, Institute for Pure and Applied Mathematics, October 2006.
- Allerton Conference on Communication, Control, and Computing, October 2004.