

# Andrew McGregor

---

CONTACT INFORMATION	Microsoft Research 1065 L'Avenida Mountain View, CA, 94043.	<i>Phone:</i> (267) 432-6308 <i>E-mail:</i> amcgreg@microsoft.com <i>Web:</i> talk.ucsd.edu/andrewm
RESEARCH INTERESTS	Algorithms for processing massive data sets and data streams; computing with noisy or incomplete data; clustering; communication complexity; coding and information theory. My work appears in theoretical computer science, database, data mining, coding theory, and machine learning publications.	
EDUCATION	<b>University of Pennsylvania</b> , Philadelphia, USA <span style="float: right;"><b>2001 – 2006</b></span> Ph.D., "Processing Data Streams" advised by Sampath Kannan, 2007 M.S.E., Computer Science, 2002	
	<b>University of Cambridge</b> , Cambridge, UK <span style="float: right;"><b>1997 – 2001</b></span> Certificate of Advanced Study in Mathematics (Distinction), 2001 B.A., Mathematics (1st Class), 2000	
EMPLOYMENT	<b>Microsoft Research</b> , Mountain View, USA <span style="float: right;"><b>May 2008 – Present</b></span> Postdoctoral researcher in MSR-SVC.	
	<b>University of California</b> , San Diego, USA <span style="float: right;"><b>October 2006 – May 2008</b></span> Postdoctoral researcher in the Information Theory and Applications Center working on clustering, algorithmic issues in coding theory, and processing massive data sets and streams. Co-ordinating an interdisciplinary seminar between the CSE and ECE departments.	
	<b>Bell Labs, Lucent Technologies</b> , Murray Hill, USA <span style="float: right;"><b>Summer 2002, 2004, 2005</b></span> Three summer internships in the Fundamental Mathematics Group solving combinatorial optimization problems related to optical network design and developing new bounds for the error rate of optimal codes. Mentored by A. Barg and B. Shepherd.	
	<b>DIMACS</b> , New Brunswick, USA <span style="float: right;"><b>Summer 2003</b></span> Long term visitor (ten weeks) developing efficient list-decoding algorithms for concatenated codes. Hosted by A. Barg.	
	<b>University of Glasgow</b> , Glasgow, UK <span style="float: right;"><b>Summer 1997, 1999</b></span> Summer internship in the Department of Computer Science developing and analyzing computer models of bacteria evolution (1999). Summer internship in the Department of Mechanical Engineering developing computer models of musical instruments (1997).	
	<b>Memorial University</b> , St. John's, Canada <span style="float: right;"><b>Summer 1998</b></span> Summer internship in the Multimedia and Communications Laboratory developing image registration algorithms for augmented-reality applications.	

BOOKS AND  
THESES

- [B1] A. McGregor and S. Muthukrishnan. Streams 2.0: New developments in data stream processing. Book in Preparation.
- [B2] A. McGregor. Processing Data Streams. Ph.D. Thesis, University of Pennsylvania, 2007. Advisor: Sampath Kannan

JOURNAL  
PUBLICATIONS

- [J1] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. *Submitted to ACM Transactions in Algorithms*, 2007.
- [J2] A. McGregor and O. Milenkovic. On the hardness of approximating stopping and trapping sets in LDPC codes. *Submitted to IEEE Trans. Inform. Theory*, 2008.
- [J3] S. Guha and A. McGregor. Stream order and order statistics: Quantile estimation in random-order streams. *SIAM Journal of Computing* (under revision), 2008.
- [J4] T. S. Jayram, A. McGregor, S. Muthukrishnan, and E. Vee. Estimating statistical aggregates on probabilistic data streams. *ACM Transactions on Database Systems* (under revision), 2008.
- [J5] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. Graph distances in the data stream model. *SIAM Journal of Computing* (to appear), 2008.
- [J6] S. Guha, A. McGregor, and S. Venkatasubramanian. Sub-linear estimation of entropy and information distances. *ACM Transactions in Algorithms* (to appear), 2008.
- [J7] S. Guha, P. Indyk, and A. McGregor. Sketching information divergences. *Journal of Machine Learning*, 72(1-2):5–19, 2008.
- [J8] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. *Theoretical Computer Science*, 348(2-3):207–216, 2005.
- [J9] A. Barg and A. McGregor. Distance distribution of binary codes and the error probability of decoding. *IEEE Trans. Inform. Theory*, 51(12):4237– 4246, 2005.

CONFERENCE  
PUBLICATIONS

- [C1] K. Chaudhuri and A. McGregor. Finding metric structure in information theoretic clustering. *Conference on Learning Theory*, 2008.
- [C2] S. Guha and A. McGregor. Tight multi-pass stream lower bounds via pass elimination. *International Colloquium on Automata, Languages and Programming*, pages 760–772, 2008.
- [C3] G. Cormode and A. McGregor. Approximation algorithms for clustering uncertain data. *ACM Symposium on Principles of Database Systems*, pages 191–200, 2008.
- [C4] A. Chakrabarti, G. Cormode, and A. McGregor. Robust lower bounds for communication and stream computation. *ACM Symposium on Theory of Computing*, pages 641–650, 2008.
- [C5] S. Angelov, K. Kunal, and A. McGregor. Sorting and selection with random costs. *Latin American Theoretical Informatics Symposium*, pages 48–59, 2008.
- [C6] P. Indyk and A. McGregor. Declaring independence via the sketching of sketches. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 737–745, 2008.

- [C7] A. McGregor and O. Milenkovic. On the hardness of approximating stopping and trapping sets in LDPC codes. In *IEEE Information Theory Workshop*, pages 248–253, 2007.
- [C8] S. Guha and A. McGregor. Lower bounds for quantile estimation in random-order and multi-pass streaming. In *International Colloquium on Automata, Languages and Programming*, pages 704–715, 2007.
- [C9] M. Chu, S. Kannan, and A. McGregor. Checking and spot-checking of heaps. In *International Colloquium on Automata, Languages and Programming*, pages 728–739, 2007.
- [C10] S. Guha, P. Indyk, and A. McGregor. Sketching information divergences. In *Conference on Learning Theory*, pages 424–438, 2007. **Invited to Special Issue of the Journal of Machine Learning.**
- [C11] T. S. Jayram, A. McGregor, S. Muthukrishnan, and E. Vee. Estimating statistical aggregates on probabilistic data streams. In *ACM Symposium on Principles of Database Systems*, pages 243–252, 2007. **Invited to Special Issue of ACM Transactions on Database Systems.**
- [C12] S. Guha and A. McGregor. Space-efficient sampling. In *AISTATS*, pages 169–176, 2007.
- [C13] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 328–335, 2007.
- [C14] A. McGregor and B. Shepherd. Island hopping and path coloring with applications to WDM network design. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 864–873, 2007. **SIGACT “Significant papers on new areas published in proceedings 2007” (1 of 13).**
- [C15] S. Guha, A. McGregor, and S. Venkatasubramanian. Streaming and sublinear approximation of entropy and information distances. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 733–742, 2006.
- [C16] D. Agarwal, A. McGregor, J. M. Phillips, S. Venkatasubramanian, and Z. Zhu. Spatial scan statistics: approximations and performance study. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 24–33, 2006.
- [C17] S. Guha and A. McGregor. Approximate quantiles and the order of the stream. In *ACM Symposium on Principles of Database Systems*, pages 273–279, 2006.
- [C18] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. Graph distances in the streaming model: the value of space. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 745–754, 2005.
- [C19] S. Kannan and A. McGregor. More on reconstructing strings from random traces: Insertions and deletions. In *IEEE International Symposium on Information Theory*, pages 297–301, 2005.
- [C20] B. Harb, S. Kannan, and A. McGregor. Approximating the best-fit tree under  $l_p$  norms. In *APPROX-RANDOM*, pages 123–133, 2005.
- [C21] A. McGregor. Finding graph matchings in data streams. In *APPROX-RANDOM*, pages 170–181, 2005.
- [C22] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. In *International Colloquium on Automata, Languages and*

*Programming*, pages 531–543, 2004. **Invited to Special Issue of Theoretical Computer Science.**

[C23] A. McGregor. A problem in scheduling: Your time starts now. . . . In *FUN with Algorithms*, pages 34–40, 2004.

[C24] A. Barg and A. McGregor. List decoding of concatenated codes: improved performance estimates. In *IEEE International Symposium on Information Theory*, page 419, 2004.

[C25] T. Batu, S. Kannan, S. Khanna, and A. McGregor. Reconstructing strings from random traces. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 910–918, 2004.

[C26] A. Barg and A. McGregor. More on the reliability function of the binary symmetric channel. In *IEEE International Symposium on Information Theory*, page 115, 2003.

[C27] A. Barg and A. McGregor. Distance distribution of binary codes and the error probability of decoding. In *Workshop in Coding and Cryptography*, pages 51–61, 2003.

[C28] A. McGregor, E. Miranda, and P. Gawthrop. Physical modeling of musical instruments using bond graphs. In *Brazilian Symposium on Computer Music*, 1999.

TEACHING  
EXPERIENCE

**University of Pennsylvania**, Philadelphia, USA **2002-2003**  
Teaching assistant for CSE 334 Advanced Algorithms and CIS 502 Algorithms.

DEPARTMENT  
SERVICE

Organizer and Co-Creator of the ITA Seminar (2007-Present)  
Organizer of the Algorithms and Complexity Seminar (2003-06)  
Graduate Student Representative at Faculty Meetings (2003-04)  
Organizer and Co-Creator of the Bi-weekly Departmental Social (2002-04)  
Organizer of the Theory Lunch (2001-02)  
President of the Archimedean, University of Cambridge Maths Society (2000-2001)

INVITED TALKS

Invited Sessions and Workshops:  
NIPS Workshop “Representations & Inference on Distributions” with P. Indyk (2007)  
Allerton Conference on Communication, Control, and Computing (2007)  
UCSD, Information Theory and Applications Workshop (2007)  
AMS & MAA Joint Mathematics Meeting (2007)  
IIT Kanpur Workshop on Algorithms for Data Streams (2006)  
Bertinoro Workshop on Space-Conscious Algorithms (2006)  
Dagstuhl Seminar 05291 on Sublinear Algorithms (2005)  
DIMACS Workshop on Algebraic Coding Theory and Information Theory (2003)

Seminars (Universities):

King’s College London, Massachusetts Institute of Technology, McGill University, Purdue University, University of California (Merced & San Diego), University of Cambridge, University of Illinois (Urbana Champaign), University of Maryland, University of Massachusetts (Amherst), University of Michigan, Université Paris Sud, University of Pennsylvania, and University of Utah.

Seminars (Industry):

Alcatel-Lucent Bell Labs, AT&T Labs, Google NYC, IBM Almaden, Microsoft SVC, and Yahoo! Research.

COMMUNITY  
INVOLVEMENT

Organizer:

DIMACS/DyDAn Workshop on Streaming, Coding, and Compressive Sensing: Unifying Theory and Common Applications to Sparse Signal/Data Analysis and Processing (2009)

Program Committees:

17th ACM Conference on Information and Knowledge Management (2008)  
4th IEEE/ACM International Conference on Distributed Computing in Sensor Systems (2008)

Conference Reviewing:

COLT, ESA, FOCS, ICALP, ISAAC, ISIT, MFCS, PODC, PODS, RANDOM, SODA, and STACS.

Journal Reviewing:

SIAM Journal on Computing, Journal of Algorithms, IEEE Transactions on Information Theory, IEEE Transactions on Knowledge and Data Engineering, Theory of Computing Systems, Discrete Applied Mathematics, and ACM Transactions on Algorithms.

Collation and Editing of "Open Problems in Data Streams and Related Topics":

[www.cse.iitk.ac.in/users/sganguly/data-stream-probs.pdf](http://www.cse.iitk.ac.in/users/sganguly/data-stream-probs.pdf)

HONORS AND  
AWARDS

Special Issue of Journal of Machine Learning for Best Papers from COLT (2007)  
Special Issue of Transactions on Database Systems for Best Papers from PODS (2007)  
SIGACT's "Significant papers on new areas" (1 of 13 from 2007)  
IEEE Travel Award for ISIT (2003, 2004, 2005)  
Special Issue of Theoretical Computer Science for Best Papers from ICALP (2004)  
IBM Research Travel Award for SODA (2003)  
St. Andrew's Society of the State of New York Scholarship (2001)  
Brown Prize in Pure Mathematics (2001)  
Gonville and Caius Senior Scholarship for Mathematics (2000)  
Bodey Prize in Applied Mathematics (1999)  
Paton Taylor Traveling Scholarship (1998)  
Gonville and Caius Junior Scholarship for Mathematics (1998)  
Daily Telegraph & BAAS Young Science Writer of the Year (1998)

REFERENCES

**Alexander Barg**

Dept. of Electrical & Computer Eng.,  
University of Maryland,  
College Park, MD 20742  
Email: [abarg@ieee.org](mailto:abarg@ieee.org)

**Sampath Kannan**

Dept. of Computer & Info. Science,  
University of Pennsylvania,  
Philadelphia, PA 19104  
Email: [kannan@cis.upenn.edu](mailto:kannan@cis.upenn.edu)

**Sudipto Guha**

Dept. of Computer & Info. Science,  
University of Pennsylvania,  
Philadelphia, PA 19104  
Email: [sudipto@cis.upenn.edu](mailto:sudipto@cis.upenn.edu)

**S. Muthukrishnan**

Research Scientist,  
Google Inc.,  
New York, NY 10011  
Email: [muthu@google.com](mailto:muthu@google.com)

**Piotr Indyk**

Computer Science & A.I. Lab,  
Massachusetts Institute of Technology,  
Cambridge, MA 02139  
Email: [indyk@theory.lcs.mit.edu](mailto:indyk@theory.lcs.mit.edu)

**Bruce Shepherd**

Dept. of Mathematics & Statistics,  
McGill University,  
Montreal, Canada H3A 2K6  
Email: [bruce.shepherd@mcgill.ca](mailto:bruce.shepherd@mcgill.ca)