Curriculum Vitae

Geoffrey E. Hinton

January 6, 2020

Citizenship: Canadian (also British)

Address: Department of Computer Science
University of Toronto
10 Kings College Road
Toronto, Ontario, M5S 3G4
Canada
Email: hinton@cs.toronto.edu

Higher Education and Qualifications

1967 - 1970 Cambridge University, B.A. Hons (Experimental Psychology)

1972 - 1975 Edinburgh University, PhD. in Artificial Intelligence (awarded 1978)

Professional Experience

Jan 76 - Sept 78	Research Fellow Cognitive Studies Program, Sussex University, England
Oct 78 - Sept 80	Visiting Scholar Program in Cognitive Science, University of California, San Diego
Oct 80 - Sept 82	Scientific Officer MRC Applied Psychology Unit, Cambridge, England
Jan 82 - June 82	Visiting Assistant Professor Psychology Department, University of California, San Diego
Oct 82 - June 87	Assistant Professor then Associate Professor Computer Science Department, Carnegie-Mellon University, Pittsburgh, USA
Jul 87 - June 98	Professor Computer Science Department, University of Toronto, Canada
Jul 98 - Sep 01	Founding Director of the Gatsby Computational Neuroscience Unit University College London, England
Oct 01 -	Professor Computer Science Department, University of Toronto, Canada (Became"University Professor" in 2006 then Emeritus in 2014)
Mar 13 - Sep 16	Distinguished Researcher, Google (half-time).
Oct 16 -	VP and Engineering Fellow, Google (half-time).
Jan 17 -	Chief Scientific Adviser, Vector Institute (*pro bono*)

Professional Recognition

Fellowships

2016 Honorary Foreign Member of the US National Academy of Engineering
2015 Honorary Foreign Member of the Spanish Real Academia de Ingenieria
2014 Distinguished Fellow, Canadian Institute for Advanced Research
2003 Honorary Foreign Member of the American Academy of Arts and Sciences
2003 Fellow of the Cognitive Science Society
1998 Fellow of the Royal Society
1996 Fellow of the Royal Society of Canada
1991 Fellow, Association for the Advancement of Artificial Intelligence
1987 Fellow, Canadian Institute for Advanced Research (1987-1998; 2004-2014)

Awards

2019 ACM A. M. Turing Award (jointly with Yoshua Bengio and Yann LeCun)
2019 Honda Prize
2019 Toronto Region Builder Award
2018 Companion of the Order of Canada (Canada's highest honour)
2017 BBVA Foundation Frontiers of Knowledge Award
2016 NEC C&C Award
2016 IEEE/RSE James Clerk Maxwell Gold Medal
2014 IEEE Frank Rosenblatt Medal
2012 Killam Prize in Engineering
2010 Gerhard Herzberg Canada Gold Medal
2005 IJCAI Research Excellence Award
2001 The David E. Rumelhart Prize
1998 IEEE Neural Networks Pioneer Award
1992 ITAC/NSERC award for academic excellence.
1990 IEEE Signal Processing Society Senior Award

Honorary Degrees

2013 Doctorat honorifique, University of Sherbrooke
2011 Honorary Degree of Doctor of Science, University of Sussex
2001 Honorary Degree of Doctor of Science, University of Edinburgh

Top N lists

2019 Toronto's 50 most influential people, *Toronto Life Magazine*
2018 Toronto's 50 most influential people, *Toronto Life Magazine*
2017 The Bloomberg 50, *Bloomberg Business week*
2017 The 50 most powerful people in Canadian business, *The Globe and Mail Report on Business*
2017 Toronto's 50 most influential people, *Toronto Life Magazine*
2016 The WIRED 100 - 2016's most influential people, *Wired Magazine*

Named Lectures

2014: Dertousos Lecture, MIT
2012: Killam Prize Lecture, McGill
2011: The Foundation Lecture, Royal Canadian Institute
2010: The Hans-Lukas Teuber Lecture, MIT
2010: The "Big Thinkers" lecture, Yahoo, San Jose
2010: The Rockwood Memorial Lecture, UC San Diego
2009: The Ed Posner Lecture, NIPS-09, Vancouver
2009: The Ian Howard Lecture, York University
2006: The Graham Lecture, University of Toronto
2003: The Pinkel Lecture, University of Pennsylvania
2001: The David E. Rumelhart Prize Lecture, Edinburgh
1998: The Rockwood Memorial Lecture, UC San Diego
1995: The Rockwood Memorial Lecture, UC San Diego
1993: The Herzberg Lecture, Ottawa.
1993: The Broadbent Lecture, London, UK.
1992: The Benjamin Meaker Lectures, Bristol University (5 lectures).
1991: The St Andrews Easter Lectures, St Andrews (6 lectures).
1989: The fourth annual Hebb lecture, Dalhousie University, Halifax.
1989: The Sun Annual Lectures, University of Manchester (8 lectures)
1987: The Weigand Lecture, University of Toronto
1986: The David Marr Memorial Lecture, Kings College Cambridge

PUBLICATIONS

Refereed Journal Papers

1. LeCun, Y., Bengio, Y. and Hinton, G. E. (2015) Deep Learning *Nature*, **521**, pp 436-444.

2. Srivastava, N., Hinton, G. E., Krizhevsky, K., Sutskever, I. and Salakhutdinov. R. (2014) Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15(1)**, pp 1929-1958

3. Sarikaya, R., Hinton, G. E. and Deoras, P. (2014) Application of Deep Belief Networks for Natural Language Understanding. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, **22.4**, pp 778-784.

4. Hinton, G. E. (2014) Where do features come from? *Cognitive Science*, **38(6)**, 1078-1101.

5. Ranzato, M., Mnih, V., Susskind, J. and Hinton, G. E. (2013) Modeling Natural Images Using Gated MRFs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **35:9**, pp 2206-2222.

6. Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, **29:6**, pp 82-97.

7. Salakhutdinov, R. R. and Hinton, G. E. (2012) An Efficient Learning Procedure for Deep Boltzmann Machines. *Neural Computation*, **24**, pp 1967-2006.

8. van der Maaten, L. J. P. and Hinton, G. E. (2012) Visualizing Non-Metric Similarities in Multiple Maps. *Machine Learning*, **87**, pp 33-55.

9. Mohamed, A., Dahl, G. and Hinton, G. E. (2012) Acoustic Modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, **20**, pp 14-22.

10. Taylor, G. W, Hinton, G. E., and Roweis, S. (2011) Two distributed-state models for generating high-dimensional time series *Journal of Machine Learning Research*, **12**, pp 863-907.

11. Hinton, G. E. and Salakhutdinov, R. (2011) Discovering Binary Codes for Fast Document Retrieval by Learning Deep Generative Models. *Topics in Cognitive Science*, **3:1**, pp 74-91.

12. Schmah, T., Yourganov, G., Zemel, R. S., Hinton, G. E., Small, S. l., and Strother, S. C. (2010) Comparing Classification Methods for Longitudinal fMRI Studies *Neural Computation*, **22**, pp 2729-2762.

13. Memisevic, R. and Hinton, G. E. (2010) Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural Computation*, **22**, pp 1473-1492.

14. Hinton, G. E. (2010) Learning to represent visual input. *Philosophical Transactions of the Royal Society, B.* **365**, pp 177-184.

15. Sutskever, I. and Hinton, G. E. (2010) Temporal Kernel Recurrent Neural Networks *Neural Networks*, **23**, pp 239-243

16. Salakhutdinov, R. and Hinton, G. E. (2009) Semantic Hashing. *International Journal of Approximate Reasoning*, **50**, pp 969-978.

17. Mnih, A., Yuecheng, Z., and Hinton, G. E. (2009) Improving a statistical language model through non-linear prediction. *NeuroComputing*, **72**, pp 1414-1418.

18. van der Maaten, L. J. P. and Hinton, G. E. (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research*, **9**(Nov) pp 2579-2605.

19. Sutskever, I. and Hinton, G. E. (2008) Deep Narrow Sigmoid Belief Networks are Universal Approximators *Neural Computation*, **20**, pp 2629-2636.

20. Hinton, G. E. (2007) Learning multiple layers of representation. *Trends in Cognitive Science*, **11**, pp 428-434.

21. Hinton, G. E. and Salakhutdinov, R. (2006) Non-linear dimensionality reduction using neural networks. *Science*, **313**, pp 504-507, July 28 2006.

22. Hinton, G. E., Osindero, S., Welling, M. and Teh, Y. (2006) Unsupervised discovery of non-linear structure using contrastive back-propagation. *Cognitive Science*, **30**, (4), pp 725-731.

23. Hinton, G. E., Osindero, S. and Teh, Y. (2006) A fast learning algorithm for deep belief nets. *Neural Computation*, **18**, pp 1527-1554.

24. Osindero, S., Welling, M. and Hinton G. E. (2006) Topographic Product Models Applied To Natural Scene Statistics. *Neural Computation*, **18**, pp 381-414.

25. Memisevic, R. and Hinton, G. E. (2005) Improving dimensionality reduction with spectral gradient descent. *Neural Networks*, **18**, pp 702-710.

26. Sallans, B and Hinton, G. E. (2004) Reinforcement Learning with Factored States and Actions. *Journal of Machine Learning Research*, **5** pp 1063–1088.

27. Welling, M., Zemel, R. and Hinton, G. E. (2004) Probabilistic sequential independent components analysis. *IEEE Transactions on Neural Networks*, **15**, pp 838-849.

28. Teh, Y. W, Welling, M., Osindero, S. and Hinton G. E. (2003) Energy-Based Models for Sparse Overcomplete Representations. *Journal of Machine Learning Research*, **4**, pp 1235-1260.

29. Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G. E., and Ashburner, J. (2002) Classical and Bayesian Inference in Neuroimaging: Theory. *NeuroImage*, **16**, pp 465-483.

30. Hinton, G. E.(2002) Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, **14**, pp 1771-1800.

31. Mayraz, G. and Hinton, G. E. (2001) Recognizing hand-written digits using hierarchical products of experts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, pp 189-197.

32. Paccanaro, A., and Hinton, G. E. (2000) Learning distributed representations of concepts from relational data using linear relational embedding. *IEEE Transactions on Knowledge and Data Engineering*, **13**, 232-245.

33. Ueda, N. Nakano, R., Ghahramani, Z. and Hinton, G. E. (2000) SMEM Algorithm for Mixture Models. *Neural Computation*, **12**, 2109-2128.

34. Ghahramani, Z. and Hinton, G.E. (2000) Variational Learning for Switching State-space Models. *Neural Computation*, **12**, 831-864.

35. Ueda, N. Nakano, R., Ghahramani, Z. and Hinton, G. E. (1999) Split and Merge EM Algorithm for Improving Gaussian Mixture Density Estimates. *Journal of VLSI Signal Processing Systems*, **26**, 133-140.

36. Frey, B. J., and Hinton, G. E. (1999) Variational Learning in Non-linear Gaussian Belief Networks. *Neural Computation*, **11**, 193-214.

37. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. (1998) A comparison of statistical learning methods on the GUSTO database. *Statistics in Medicine*, **17** 2501-2508.

38. Tibshirani, R. and Hinton, G.E. (1998) Coaching variables for regression and classification. *Statistics and Computing*, **8**, 25-33.

39. de Sa, V. R. and Hinton, G. E. (1998) Cascaded Redundancy Reduction. *Network: Computation in Neural Systems*, **9**, 73-84.

40. Fels, S. S. and Hinton, G. E. (1997) Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, **8**, 977-984.

41. Hinton, G. E. and Ghahramani, Z. (1997) Generative Models for Discovering Sparse Distributed Representations. Philosophical Transactions of the Royal Society, B. **352**, 1177-1190.

42. Frey, B. J., and Hinton, G. E. (1997) Efficient stochastic source coding and an Application to a Bayesian Network Source Model. *The Computer Journal*, **40 (2)**.

43. Hinton, G. E., Dayan, P. and Revow M. (1997) Modeling the manifold of images of handwritten digits. *IEEE Transactions on Neural Networks*, **8**, 65-74.

44. Williams, C. K. I., Revow, M. and Hinton, G. E. (1997) Instantiating deformable models with a neural net. *Computer Vision and Image Understanding*. **68**, 120-126

45. Dayan, P. and Hinton, G. E. (1997) Using Expectation-Maximization for Reinforcement Learning. *Neural Computation*, **9**, 271-278.

46. Oore, S., Hinton, G. E. and Dudek, G. (1997) A mobile robot that learns its place. *Neural Computation*, **9**, 683-699.

47. Dayan, P. and Hinton, G. E. (1996) Varieties of Helmholtz Machine. *Neural Networks*, **9**, 1385-1403.

48. Revow, M., Williams, C. K. I. and Hinton, G. E. (1996) Using Generative Models for Handwritten Digit Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 592-606.

49. Dayan, P., Hinton, G. E., Neal, R., and Zemel, R. S. (1995) Helmholtz Machines. *Neural Computation*, bf 7, 1022-1037.

50. Hinton, G. E., Dayan, P., Frey, B. J. and Neal, R. (1995) The wake-sleep algorithm for self-organizing neural networks. *Science*, **268**, pp 1158-1161.

51. Zemel, R. S. and Hinton, G. E. (1995) Learning Population Codes by Minimizing Description Length *Neural Computation*, **7**, 549-564.

52. Becker, S. and Hinton, G. E. (1993) Learning mixture models of spatial coherence. *Neural Computation*, **5**, 267-277.

53. Nowlan. S. J. and Hinton, G. E. (1993) A soft decision-directed LMS algorithm for blind equalization. *IEEE Transactions on Communications*, **41**, 275-279.

54. Fels, S. S. and Hinton, G. E. (1992) Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, **3**.

55. Becker, S. and Hinton, G. E. (1992) A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, **355:6356**, 161-163.

56. Nowlan. S. J. and Hinton, G. E. (1992) Simplifying neural networks by soft weight sharing. *Neural Computation*, **4**, 173-193.

57. Jacobs, R., Jordan, M. I., Nowlan. S. J. and Hinton, G. E. (1991) Adaptive mixtures of local experts. *Neural Computation*, **3**, 79-87.

58. Hinton, G. E. and Shallice, T. (1991) Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* **98**, 74-95.

59. Hinton, G. E. (1990) Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, **46**, 47-75.

60. Hinton, G. E. and Nowlan, S. J. (1990) The bootstrap Widrow-Hoff rule as a cluster-formation algorithm. *Neural Computation*, **2**, 355-362.

61. Lang, K., Waibel, A. and Hinton, G. E. (1990) A Time-Delay Neural Network Architecture for Isolated Word Recognition. *Neural Networks*, **3**, 23-43.

62. Hinton, G. E. (1989) Connectionist learning procedures. *Artificial Intelligence*, **40**, 185-234.

63. Waibel, A. Hanazawa, T. Hinton, G. Shikano, K. and Lang, K. (1989) Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Acoustics Speech and Signal Processing*, **37**, 328-339.

64. Hinton, G. E. (1989) Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation*, **1**, 143-150.

65. Touretzky, D. S. and Hinton, G. E. (1988) A distributed connectionist production system. *Cognitive Science*, **12**, 423-466.

66. Hinton, G. E. and Parsons, L. A. (1988) Scene-based and viewer-centered representations for comparing shapes. *Cognition*, **30**, 1–35.

67. Hinton, G. E. (1987) The horizontal-vertical delusion. *Perception*, **16**.

68. Plaut, D. C. and Hinton, G. E. (1987) Learning sets of filters using back-propagation. *Computer Speech and Language*, **2**, 35–61.

69. Hinton, G. E. and Nowlan, S. J. (1987) How learning can guide evolution. *Complex Systems*, **1**, 495–502.

70. Fahlman, S. E. and Hinton, G. E. (1987) Connectionist architectures for Artificial Intelligence. *IEEE Computer*, **20**, 100–109.

71. Sejnowski, T. J., Kienker, P. K., and Hinton, G. E. (1986) Learning symmetry groups with hidden units: Beyond the perceptron. *Physica D*, **22**, 260–275.

72. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature*, **323**, 533–536.

73. Kienker, P. K., Sejnowski, T. J., Hinton, G. E., and Schumacher, L. E. (1986) Separating figure from ground with a parallel network. *Perception*, **15**, 197–216.

74. Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985) A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**, 147–169.

75. Hutchins, E. L. and Hinton, G. E. (1984) Why the islands move. *Perception*, **13**, 629–632.

76. Hinton, G. E. (1984) Parallel computations for controlling an arm. *The Journal of Motor Behavior*, **16**, 171–194.

77. Ballard, D. H., Hinton, G. E., and Sejnowski, T. J. (1983) Parallel visual computation. *Nature*, **306**, 21–26.

78. Hinton, G. E. (1979) Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, **3**, 231-250.

79. Hinton, G. E. (1978) Respectively reconsidered. *Pragmatics Microfiche*, May issue.

Refereed Conference Papers

80. Qin, Y., Frosst, N., Sabour, S., Raffel, C., Cottrell, C. and Hinton, G. (2020) Detecting and Diagnosing Adversarial Images with Class-Conditional Capsule Reconstructions *ICLR-2020*

81. Kosiorek, A. R., Sabour, S., Teh, Y. W. and Hinton, G. E. (2019) Stacked Capsule Autoencoders *Advances in Neural Information Processing Systems 32*

82. Zhang, M., Lucas, J., Ba, J., and Hinton, G. E. (2019) Lookahead Optimizer: k steps forward, 1 step back *Advances in Neural Information Processing Systems 32*

83. Muller, R., Kornblith, S. and Hinton G. (2019) When Does Label Smoothing Help? *Advances in Neural Information Processing Systems 32*

84. Kornblith, S., Norouzi, M., Lee, H. and Hinton, G. (2019) Similarity of neural network representations revisited *ICML-2019*

85. Hinton, G. E., Sabour, S. and Frosst, N. (2018) Matrix Capsules with EM Routing *ICLR-2018*

86. Kiros, J. R., Chan, W. and Hinton, G. E. (2018) Illustrative Language Understanding: Large-Scale Visual Grounding with Image Search *ACL-2018*

87. Anil, R., Pereyra, G., Passos, A., Ormandi, R., Dahl, G. and Hinton, G. E. (2018) Large scale distributed neural network training through online distillation *ICLR-2018*

88. Guan, M. Y., Gulshan, V., Dai, A. M. and Hinton, G. E. (2018) Who Said What: Modeling Individual Labelers Improves Classification *AAAI-2018*

89. Sabour, S., Frosst, N. and Hinton, G. E. (2017) Dynamic Routing between Capsules *NIPS-2017*

90. Frosst, N. and Hinton, G. E. (2017) Distilling a Neural Network Into a Soft Decision Tree. Preprint at arXiv:1711.09784

91. Pereyra, G., Tucker, T., Chorowski, J., Kaiser, L. and Hinton, G. E. (2017) Regularizing neural networks by penalizing confident output distributions. Preprint at arXiv:1701.06548

92. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., and Dean, J. (2017) Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *NIPS-2017*, Preprint at arXiv:1701.06538

93. Ba, J. L., Hinton, G. E., Mnih, V., Leibo, J. Z. and Ionescu, C. (2016) Using Fast Weights to Attend to the Recent Past. *NIPS-2016*, Preprint at arXiv:1610.06258v2

94. Ba, J. L., Kiros, J. R. and Hinton, G. E. (2016) Layer normalization. *Deep Learning Symposium, NIPS-2016*, Preprint at arXiv:1607.06450

95. Ali Eslami, S. M., Nicolas Heess, N., Theophane Weber, T., Tassa, Y., Szepesvari, D., Kavukcuoglu, K. and Hinton, G. E. (2016) Attend, Infer, Repeat: Fast Scene Understanding with Generative Models. *NIPS-2016*, Preprint at arXiv:1603.08575v3

96. Hinton, G. E., Vinyals, O., and Dean, J. (2015) Distilling the knowledge in a neural network. *Workshop on Deep Learning, NIPS-2014*, Preprint at arXiv:1503.02531

97. Jaitly, N., Vanhoucke, V. and Hinton, G. E. (2014) Autoregressive product of multi-frame predictions can improve the accuracy of hybrid models. *Fifteenth Annual Conference of the International Speech Communication Association.*

98. Jaitly, N., and Hinton, G. E. (2013) Vocal Tract Length Perturbation (VTLP) improves speech recognition. *Proc. ICML Workshop on Deep Learning for Audio,Speech and Language Processing*, Atlanta, USA.

99. Srivastava, N., Salakhutdinov, R. R. and Hinton, G. E. (2013) Modeling Documents with a Deep Boltzmann Machine. *Uncertainty in Artificial Intelligence (UAI 2013)*

100. Graves, A., Mohamed, A. and Hinton, G. E. (2013) Speech Recognition with Deep Recurrent Neural Networks. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

101. Dahl, G. E., Sainath, T. N. and Hinton, G. E. (2013) Improving Deep Neural Networks for LVCSR Using Rectified Linear Units and Dropout. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

102. Zeiler, M. D., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q.V., Nguyen, P., Senior, A., Vanhoucke, V., Dean, J. and Hinton, G. E. (2013) On Rectified Linear Units for Speech Processing. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

103. Deng, L., Hinton, G. E. and Kingsbury, B. (2013) New types of deep neural network learning for speech recognition and related applications: An overview *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

104. Sutskever, I., Martens, J., Dahl, G. and Hinton, G. E. (2013) On the importance of momentum and initialization in deep learning. *International Conference on Machine Learning*, Atlanta, USA

105. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2013) Tensor Analyzers. *International Conference on Machine Learning*, Atlanta, USA

106. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. (2012) Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580

107. Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing 25*, MIT Press, Cambridge, MA

108. Salakhutdinov, R. R. and Hinton, G. E. (2012) A Better Way to Pretrain Deep Boltzmann Machines. *Advances in Neural Information Processing 25*, MIT Press, Cambridge, MA

109. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Deep Lambertian Networks. *International Conference on Machine Learning*,

110. Mnih, V. and Hinton, G. E. (2012) Learning to Label Aerial Images from Noisy Data. *International Conference on Machine Learning*,

111. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Deep Mixtures of Factor Analysers. *International Conference on Machine Learning*,

112. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Robust Boltzmann Machines for Recognition and Denoising. *IEEE Conference on Computer Vision and Pattern Recognition*,

113. Mohamed,A., Hinton, G. E. and Penn, G. (2012) Understanding how Deep Belief Networks perform acoustic modelling *ICASSP 2012*, Kyoto.

114. Jaitly, N. and Hinton, G. E. (2011) A new way to learn acoustic events. *Advances in Neural Information Processing Systems 24, Deep Learning workshop*, Grenada, Spain.

115. Mnih, V., Larochelle, H. and Hinton, G. (2011) Conditional Restricted Boltzmann Machines for Structured Output Prediction *Uncertainty in Artificial Intelligence, 2011.*

116. Hinton, G.E., Krizhevsky, A. and Wang, S. (2011) Transforming Auto-encoders. *ICANN-11: International Conference on Artificial Neural Networks*, Helsinki.

117. Suskever, I., Martens, J. and Hinton, G. E. (2011) Generating Text with Recurrent Neural Networks. *Proc. 28th International Conference on Machine Learning*, Seattle.

118. Ranzato, M., Susskind, J., Mnih, V. and Hinton, G. (2011) On deep generative models with applications to recognition. *IEEE Conference on Computer Vision and Pattern Recognition*

119. Susskind,J., Memisevic, R., Hinton, G. and Pollefeys, M. (2011) Modeling the joint density of two images under a variety of transformations. *IEEE Conference on Computer Vision and Pattern Recognition*

120. Hinton, G. E., Krizhevsky, A. and Wang, S. (2011) Transforming Auto-encoders. In T. Honkela *et. al.* (Eds.): *ICANN 2011, Part I*, LNCS 6791, pp. 44-51.

121. Jaitly, N. and Hinton, G. E. (2011) Learning a better Representation of Speech Sound Waves using Restricted Boltzmann Machines. *ICASSP 2011*, Prague.

122. Mohamed,A., Sainath, T., Dahl, G., Ramabhadran, B., Hinton, G. and Picheny, M. (2011) Deep Belief Networks using Discriminative Features for Phone Recognition. *ICASSP 2011*, Prague.

123. Sarikaya, R., Hinton, G. and Ramabhadran, B. (2011) Deep Belief Nets for Natural Language Call-Routing. *ICASSP 2011*, Prague.

124. Krizhevsky, A. and Hinton, G.E. (2011) Using Very Deep Autoencoders for Content-Based Image Retrieval In *European Symposium on Artificial Neural Networks ESANN-2011)*, Bruges, Belgium.

125. Deng, L., Seltzer, M., Yu, D., Acero, A., Mohamed A., and Hinton, G. E. (2010) Binary Coding of Speech Spectrograms Using a Deep Auto-encoder. *Interspeech 2010*, Makuhari, Chiba, Japan.

126. Ranzato, M., Mnih, V., and Hinton, G. E. (2010) How to generate realistic images using gated MRF's. *Advances in Neural Information Processing Systems 23.*

127. Dahl, G., Ranzato, M., Mohamed, A., Hinton, G. E. (2010) Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine. *Advances in Neural Information Processing Systems 23.*

128. Larochelle, H. and Hinton, G. E. (2010) Learning to combine foveal glimpses with a third-order Boltzmann machine. *Advances in Neural Information Processing Systems 23.*

129. Memisevic, R., Zach, C., Hinton, G. E. and Pollefeys M. (2010) Gated Softmax Classification. *Advances in Neural Information Processing Systems 23.*

130. Ranzato, M. and Hinton, G. E. (2010) Modeling pixel means and covariances using factored third-order Boltzmann machines. *IEEE Conference on Computer Vision and Pattern Recognition.*

131. Taylor, G., Sigal, L., Fleet, D. and Hinton, G. E. (2010) Dynamic binary latent variable models for 3D human pose tracking. *IEEE Conference on Computer Vision and Pattern Recognition.*

132. Nair, V. and Hinton, G. E. (2010) Rectified linear units improve restricted Boltzmann machines. *Proc. 27th International Conference on Machine Learning*, Israel.

133. Ranzato, M., Krizhevsky, A. and Hinton, G. E. (2010) Factored 3-way restricted Boltzmann machines for modeling natural images. *Proc. Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia.

134. Mnih, V. and Hinton, G. E. (2010) Learning to detect roads in high-resolution aerial images. *To appear in European Conference on Computer Vision.*

135. Mohamed, A. R. and Hinton, G. E. (2010) Phone recognition using restricted Boltzmann machines. *ICASSP-2010*

136. Mohamed, A. R., Dahl, G. and Hinton, G. E. (2009) Deep belief networks for phone recognition. *NIPS 22 workshop on deep learning for speech recognition*

137. Salakhutdinov, R. and Hinton, G. E. (2009) Replicated Softmax: An Undirected Topic Model. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1607-1614.

138. Nair, V. and Hinton, G. E. (2009) 3-D Object recognition with deep belief nets. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1339-1347.

139. Palatucci, M, Pomerleau, D. A., Hinton, G. E. and Mitchell, T. (2009) Zero-Shot Learning with Semantic Output Codes. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1410-1418.

140. Heess, N., Williams, C. K. I. and Hinton, G. E. (2009) Learning generative texture models with extended Fields-of-Experts. In *Proc. British Machine Vision Conf.*.

141. Taylor, G. W. and Hinton, G. E. (2009) Products of Hidden Markov Models: It Takes N¿1 to Tango. *Proc. of the 25th Conference on Uncertainty in Artificial Intelligence*.

142. Taylor, G. W. and Hinton, G. E. (2009) Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style. *Proc. 26th International Conference on Machine Learning*, pp 1025-1032. Omnipress, Montreal, Quebec.

143. Tieleman, T. and Hinton, G. E. (2009) Using Fast Weights to Improve Persistent Contrastive Divergence. *Proc. 26th International Conference on Machine Learning*, pp 1033-1040. Omnipress, Montreal, Quebec.

144. Zeiler, M.D., Taylor, G.W., Troje, N.F. and Hinton, G.E. (2009) Modeling pigeon behaviour using a Conditional Restricted Boltzmann Machine. In *European Symposium on Artificial Neural Networks ESANN-2009)*.

145. Salakhutdinov, R. and Hinton, G. E. (2009) Deep Boltzmann Machines. In D. van Dyk and M. Welling (Eds.), *Proc. Twelfth International Conference on Artificial Intelligence and Statistics*, JMLR: W&CP 5, pp 448-455, Clearwater Beach, Florida, April 2009.

146. Mnih, A. and Hinton, G. E. (2009) A Scalable Hierarchical Distributed Language Model. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA
PLEASE NOTE: In 2009 NIPS changed from publishing in the year after the conference to publishing in the same year. So both NIPS22 and NIPS 21 were published in 2009

147. Nair, V. and Hinton, G. E. (2009) Implicit Mixtures of Restricted Boltzmann Machines. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

148. Sutskever, I. and Hinton, G. E. (2009) Using matrices to model symbolic relationships. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

149. Sutskever, I., Hinton, G. E. and Taylor, G. W. (2009) The Recurrent Temporal Restricted Boltzmann Machine. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

150. Schmah, T., Hinton, G. E., Zemel, R., Small, S. and Strother, S. (2009) Generative versus Discriminative Training of RBM's for classification of fMRI images. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

151. Nair, V., Susskind, J., and Hinton, G.E. (2008) Analysis-by-Synthesis by Learning to Invert Generative Black Boxes. *ICANN-08: International conference on Artificial Neural Networks*, Prague.

152. Yuecheng, Z, Mnih, A, and Hinton, G (2008) Improving a statistical language model by modulating the effects of context words. 16^{th} *European Symposium on Artificial Neural Networks*, pages 493–498.

153. Osindero, S. and Hinton, G. E. (2008) Modeling image patches with a directed hierarchy of Markov random fields. *Advances in Neural Information Processing Systems 20*, J.C. Platt and D. Koller and Y. Singer and S. Roweis (eds.), MIT Press, Cambridge, MA

154. Salakhutdinov, R. and Hinton, G. E. (2008) Using Deep Belief Nets to Learn Covariance Kernels for Gaussian Processes. *Advances in Neural Information Processing Systems 20*, J.C. Platt and D. Koller and Y. Singer and S. Roweis (eds.), MIT Press, Cambridge, MA

155. Salakhutdinov R. R, and Hinton, G. E. (2007) Semantic Hashing. *Proceedings of the SIGIR Workshop on Information Retrieval and Applications of Graphical Models*, Amsterdam.

156. Memisevic R. F. and Hinton, G. E. (2007) Unsupervised learning of image transformations. *IEEE Conference on Computer Vision and Pattern Recognition* Pages: 508-515.

157. Mnih, A. and Hinton, G. E. (2007) Three New Graphical Models for Statistical Language Modelling *International Conference on Machine Learning*, Corvallis, Oregon.

158. Salakhutdinov, R., Mnih, A. and Hinton, G. E. (2007) Restricted Boltzmann Machines for Collaborative Filtering *International Conference on Machine Learning*, Corvallis, Oregon.

159. Salakhutdinov R.R, and Hinton, G. E. (2007) Learning a non-linear embedding by preserving class neighbourhood structure. (Meila, M. and Shen, X. eds), pp 409-416, *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, The Society for AI and Statistics, Puerto Rico.

160. Sutskever, I. and Hinton, G. E. (2007) Learning multilevel distributed representations for high-dimensional sequences. (Meila, M. and Shen, X. eds), *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, pp 544-551, The Society for AI and Statistics, Puerto Rico.

161. Cook, J. A., Sutskever, I., Mnih, A. and Hinton , G. E. (2007) Visualizing similarity data with a mixture of maps. (Meila, M. and Shen, X. eds), *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, pp 65-72, The Society for AI and Statistics, Puerto Rico.

162. Taylor, G. W., Hinton, G. E. and Roweis, S. (2007) Modeling human motion using binary latent variables. *Advances in Neural Information Processing Systems* **19** MIT Press, Cambridge, MA

163. Hinton, G. E. and Nair, V. (2006) Inferring motor programs from images of handwritten digits. *Advances in Neural Information Processing Systems* **18** pp 515-522, MIT Press, Cambridge, MA

164. Memisevic, R. and Hinton, G. E. (2005) Embedding via clustering: Using spectral information to guide dimensionality reduction. *IEEE International Joint Conference on Neural Networks (IJCNN 2005)* Pages: 3198-3203

165. Mnih, A. and Hinton. G. E. (2005) Learning Unreliable Constraints using Contrastive Divergence. *IJCNN 2005*

166. Carreira-Perpignan, M. A. and Hinton. G. E. (2005) On Contrastive Divergence Learning. *Artificial Intelligence and Statistics, 2005*, Barbados.

167. Hinton, G. E., Osindero, S. and Bao, K. (2005) Learning Causally Linked Markov Random Fields. *Artificial Intelligence and Statistics, 2005*, Barbados.

168. Welling, M,, Rosen-Zvi, M. and Hinton, G. E. (2005) Exponential Family Harmoniums with an Application to Information Retrieval. *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

169. Memisevic, R. and Hinton, G. E. (2005) Multiple Relational Embedding. *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

170. Goldberger, J., Roweis, S., Salakhutdinov, R and Hinton, G. E. (2005) Neighborhood Components Analysis *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

171. Bishop, C. M. Svensen, M. and Hinton, G. E. (2004) Distinguishing Text from Graphics in On-line Handwritten Ink. In Kimura, F. and Fujisawa, H. (eds.), *Proceedings Ninth International Workshop on Frontiers in Handwriting Recognition, IWFHR-9*, Tokyo, Japan, pp. 142147.

172. Hinton, G. E., Welling, M. and Mnih, A. (2004) Wormholes Improve Contrastive Divergence. *Advances in Neural Information Processing Systems* **16** pages 417-424. MIT Press, Cambridge, MA

173. Welling, M., Zemel, R. S., and Hinton, G. E. (2003) Efficient parametric projection pursuit density estimation. In *UAI-2003: 19th Conference on Uncertainty in Artificial Intelligence.*

174. Welling, M., Hinton, G. E. and Osindero, S. (2003) Learning Sparse Topographic Representations with Products of Student-t Distributions. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

175. Welling, M., Zemel, R. and Hinton, G. E. (2003) Self-Supervised Boosting. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

176. Hinton, G. E. and Roweis, S. (2003) Stochastic Neighbor Embedding. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

177. Welling, M. and Hinton, G. E. (2002) A New Learning Algorithm for Mean Field Boltzmann Machines. *International Joint Conference on Neural Networks*, Madrid.

178. Oore, S., Terzopoulos, D. and Hinton, G. E. (2002) Local Physical Models for Interactive Character Animation. *Eurographics 2002*, **21**, Blackwell Publishers, Oxford.

179. Oore, S., Terzopoulos, D. and Hinton, G. E. (2002) A Desktop Input Device and Interface for Interactive 3D Character Animation. *Graphics Interface, to appear*

180. Roweis, S., Saul, L. and Hinton, G. E. (2002) Global Coordination of Local Linear Models *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

181. Paccanaro, A., and Hinton, G. E. (2002) Learning Hierarchical Structures with Linear Relational Embedding. *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

182. Brown, A. D. and Hinton, G. E. (2002) Relative Density Nets: A New Way to Combine Backpropagation with HMM's. *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

183. Paccanaro, A. and Hinton, G. E. (2001) Learning distributed representations of relational data using linear relational embedding. Proceedings of the 12th Italian Workshop on Neural Nets. WIRN VIETRI-2001.

184. Brown, A. D. and Hinton, G. E. (2001). Products of Hidden Markov Models. *Proceedings of Artificial Intelligence and Statistics 2001*

185. Mayraz, G. and Hinton, G. E. (2001) Recognizing Hand-Written Digits Using Hierarchical Products of Experts. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

186. Sallans, B. and Hinton, G. E. (2001) Using Free Energies to Represent Q-values in a Multiagent Reinforcement learning Task. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

187. Teh, Y. and Hinton, G. E. (2001) Rate-coded Restricted Boltzmann Machines for Face Recognition. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

188. Hinton, G. E. and Teh, Y. (2001) Discovering multiple constraints that are frequently approximately satisfied. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp 227-234.

189. Paccanaro, A. and Hinton, G. E. (2000) Extracting Distributed Representations of Concepts and Relations from Positive and Negative Propositions. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2000*.

190. Paccanaro, A. and Hinton, G. E. (2000) Learning Distributed Representations by Mapping Concepts and Relations into a Linear Space. In P. Langley (Ed.) *Proceedings of the Seventeenth International Conference on Machine Learning, ICML2000*, pp 711-718, Morgan Kaufmann Publishers, San Francisco.

191. Hinton, G.E., Ghahramani, Z. and Teh, Y.W. (2000) Learning to Parse Images. In S. A. Solla, T. K. Leen, K.-R. Muller, (Eds.) *Advances in Neural Information Processing Systems* **12**. Cambridge, MA: MIT Press.

192. Hinton, G. E. and Brown, A. (2000) Spiking Boltzmann Machines. In S. A. Solla, T. K. Leen, K.-R. Muller, (Eds.) *Advances in Neural Information Processing Systems* **12**. Cambridge, MA: MIT Press.

193. Ghahramani, Z., Korenberg, A., and Hinton, G.E. (1999) Scaling in a Hierarchical Unsupervised Network. *ICANN 99: Ninth international conference on Artificial Neural Networks*, Edinburgh.

194. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1999) SMEM Algorithm for Mixture Models. In M. S. Kearns, S. A. Solla, D. A. Cohn, (eds.) *Advances in Neural Information Processing Systems* **11**. Cambridge, MA: MIT Press.

195. Grzeszczuk, R., Terzopoulos, D., and Hinton, G. E. (1999) Fast Neural Network Emulation of Dynamical Systems for Computer Animation. In M. S. Kearns, S. A. Solla, D. A. Cohn, (eds.) *Advances in Neural Information Processing Systems* **11**. Cambridge, MA: MIT Press, pp 882-889.

196. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1999) Pattern classification using a mixture of factor analyzers. *IEEE Neural Networks for Signal Processing (NNSP99)*, pp. 525-533.

197. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1998) Split and Merge EM Algorithm for Improving Gaussian Mixture Density Estimates. *IEEE Neural Networks for Signal Processing (NNSP98)*, pp. 274-283.

198. Ghahramani, Z. and Hinton, G. E. (1998) Hierarchical Non-linear Factor Analysis and Topographic Maps. *Advances in Neural Information Processing Systems 10*. M. I. Jordan, M. J. Kearns, and S. A. Solla (Eds.) MIT Press: Cambridge, MA.

199. Grzeszczuk, R., Terzopoulos, D., and Hinton, G. E. (1998) NeuroAnimator: Fast Neural Network Emulation and Control of Physics-Based Models. *Proc. ACM SIGGRAPH-98, Computer Graphics Proceedings, Annual Conference Series*, pp 9-20.

200. Bishop, C. M., Hinton, G. E. and Strachan, I. D. G. (1997) GTM through time. *Proceedings IEE Fifth International Conference on Artificial Neural Networks*. pp 111–116. IEE, London.

201. Frey, B. J., and Hinton, G. E. (1996) Free energy coding. J. A. Storer and M. Cohn (Eds.), *Proceedings of the Data Compression Conference 1996*, IEEE Computer Society Press, Los Alamitos, CA.

202. Hinton, G. E. and Revow, M. (1996) Using Pairs of Data-Points to Define Splits for Decision Trees. *Advances in Neural Information Processing Systems 8*. D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds), pp 507-514. MIT Press, Cambridge MA.

203. Frey, B. J., Hinton, G. E. and Dayan, P. (1996) Does the wake-sleep algorithm learn good density estimators? *Advances in Neural Information Processing Systems 8*. D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds), pp 661-668. MIT Press, Cambridge MA.

204. Fels, S. S. and Hinton, G. E. (1995) GloveTalk: An adaptive interface that uses neural networks. *Proceedings of Computer Human Interface Conference*.

205. Xu, L., Jordan, M. I. and Hinton, G. E. (1995) An alternative model for mixtures of experts. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 633-640 MIT Press, Cambridge MA.

206. Hinton, G. E., Revow, M. and Dayan P. (1995) Recognizing handwritten digits using mixtures of local models. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 1015-1022 MIT Press, Cambridge MA.

207. Fels, S. S. and Hinton, G. E. (1995) GloveTalkII: Mapping hand gestures to speech using neural networks. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 843-850 MIT Press, Cambridge MA.

208. Williams, C. K. I., Hinton, G. E. and Revow, M. (1995) Using a neural net to instantiate a deformable model. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 965-972 MIT Press, Cambridge MA.

209. Zemel, R. S. and Hinton, G. E (1994) Developing Population Codes by Minimizing Description Length. *Advances in Neural Information Processing Systems 6.* J. D. Cowan, G. Tesauro and J. Alspector (Eds.), Morgan Kaufmann: San Mateo, CA.

210. Hinton, G. E. and Zemel, R. S. (1994) Autoencoders, Minimum Description Length, and Helmholtz Free Energy. *Advances in Neural Information Processing Systems 6.* J. D. Cowan, G. Tesauro and J. Alspector (Eds.), Morgan Kaufmann: San Mateo, CA.

211. Xu, L. Jordan, M. I. and Hinton, G. E. (1994) A modified gating network for the mixtures of experts architectures. *Proc. WCNN94*, San Diego, CA. vol. 2, pp. 405410.

212. Zemel, R. S. and Hinton, G. E. (1993) Developing Population Codes for Object Instantiation Parameters. *AAAI Fall Symposium Series: Machine Learning in Computer Vision* Raleigh, North Carolina USA.

213. Hinton, G. E. and van Camp, D. (1993) Keeping Neural Networks Simple by Minimizing the Description Length of the Weights. In: *Proceedings of COLT-93.*

214. Revow, M., Williams, C. K. I., and Hinton, G. E. (1993) Using mixtures of deformable models to capture variations in the shapes of hand-printed digits. *Third International Workshop on Frontiers of Handwriting Recognition.*

215. Dayan, P. and Hinton, G. E. (1993) Feudal reinforcement learning. *Advances in Neural Information Processing Systems 5.* S. J. Hanson, J. D. Cowan and C. L. Giles (Eds.), Morgan Kaufmann: San Mateo, CA.

216. Hinton, G. E., Williams, C. K. I., and Revow, M. (1992) Adaptive Elastic Models for Character Recognition. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

217. Nowlan, S. J. and Hinton, G. E. (1992) Adaptive Soft Weight Tying Using Gaussian Mixtures. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

218. Becker, S. and Hinton, G. E. (1992) Learning to make coherent predictions in domains with discontinuities. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

219. Nowlan, S. J. and Hinton, G. E. (1991) Evaluation of a system of competing experts on a vowel recognition task. *Advances in Neural Information Processing Systems 3.* R. P. Lippmann, J. E. Moody, and D. S. Touretzky (Eds.), Morgan Kaufmann: San Mateo, CA.

220. Zemel, R. and Hinton, G. E. (1991) Discovering viewpoint-invariant relationships that characterize objects. *Advances in Neural Information Processing Systems 3.* R. P. Lippmann, J. E. Moody, and D. S. Touretzky (Eds.), Morgan Kaufmann: San Mateo, CA.

221. Galland, C. G. and Hinton, G. E. (1990) Deterministic Boltzmann Learning in Networks with Asymmetric Connectivity. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J. and Hinton, G. E. (Eds.) Connectionist Models: Proceedings of the 1990 Connectionist Summer School. Morgan Kauffman: San Mateo, CA.

222. Williams, C. K. I. and Hinton, G. E. (1990) Mean field networks that learn to discriminate temporally distorted strings. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J. and Hinton, G. E. (Eds.) Connectionist Models: Proceedings of the 1990 Connectionist Summer School. Morgan Kauffman: San Mateo, CA.

223. Fels, S. S. and Hinton, G. E. (1990) Building adaptive interfaces with neural networks: The Glove-Talk pilot study. In D. Daiper, D. Gilmore, G. Cockton and B Shackel (Eds.) *Proceedings of the IFIP TC 13 Third International Conference on Human-Computer Interaction*, pages 683-688, North-Holland: Amsterdam.

224. Lang, K. J. and Hinton, G. E. (1990) Dimensionality reduction and prior knowledge in E-set recognition. In Touretzky, D. S., (Ed.) *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

225. Zemel, R., Mozer, M. and Hinton, G. E. (1990) TRAFFIC: Recognizing objects using local reference frame transformations. In Touretzky, D. S., (Ed.) *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

226. Galland, C. C. and Hinton, G. E. (1990) Discovering higher-order features with mean field networks. In Touretzky, D. S., (Ed.) *Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

227. Hinton, G. E. and Becker, S. (1990) An unsupervised learning procedure that discovers surfaces in random-dot stereograms. *Proceedings of the International Joint Conference on Neural Networks*, **Vol 1**, 218-222, Lawrence Erlbaum Associates, Hillsdale, NJ.

228. LeCun, Y., Galland, C. C., and Hinton, G. E. (1989) GEMINI: Gradient Estimation by Matrix Inversion after Noise Injection. In Touretzky, D. S., (Ed.) *Neural Information Processing Systems 1*, Morgan Kaufmann: San Mateo, CA.

229. Zemel, R. S., Mozer, M. C. and Hinton, G. E. (1988) TRAFFIC: A model of object recognition based on transformations of feature instances. In Touretzky, D. S., Hinton, G. E. and Sejnowski, T. J., editors, Proceedings of the 1988 Connectionist Summer School, Morgan Kauffman: Los Altos, CA.

230. Hinton, G. E. (1988) Representing part-whole hierarchies in connectionist networks. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, Montreal, Canada.

231. Hinton, G. E. and McClelland, J. L. (1988) Learning representations by recirculation. In D. Z. Anderson, editor, *Neural Information Processing Systems*, pages 358–366, American Institute of Physics: New York.

232. Hinton, G. E. and Plaut, D. C. (1987) Using fast weights to deblur old memories. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Seattle, WA.

233. Hinton, G. E. (1987) Learning translation invariant recognition in a massively parallel network. In Goos, G. and Hartmanis, J., editors, *PARLE: Parallel Architectures and Languages Europe*, pages 1–13, Lecture Notes in Computer Science, Springer-Verlag, Berlin.

234. Hinton, G. E. (1986) Learning distributed representations of concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, Amherst, Mass.
Reprinted in Morris, R. G. M. editor, Parallel Distributed Processing: Implications for Psychology and Neurobiology, Oxford University Press, Oxford, UK.

235. Pearlmutter, B. A. and Hinton, G. E. (1986) G-maximization: An unsupervised learning procedure for discovering regularities. In Denker, J., editor, *Neural Networks for Computing: American Institute of Physics Conference Proceedings* **151**,pp. 333-338

236. Touretzky, D. S. and Hinton, G. E. (1985) Symbols among the neurons: Details of a connectionist inference architecture. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles.

237. Hinton, G. E. and Lang, K. J. (1985) Shape recognition and illusory conjunctions. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, pp 252-259.

238. Szeliski, R. and Hinton, G. E. (1985) Solving random-dot stereograms using the heat equation. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, San Francisco.

239. Hammond, N., Hinton, G., Barnard, P., Long, J., and Whitefield, A. (1984) Evaluating the interface of a document processor: A comparison of expert judgement and user observation. *Proceedings of the First IFIP Conference on Human-Computer Interaction*, North-Holland.

240. Hinton, G. E. and Sejnowski, T. J. (1983) Analyzing cooperative computation. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*, Rochester NY.

241. Hinton, G. E. and Sejnowski, T. J. (1983) Optimal perceptual inference. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Washington DC.

242. Fahlman, S. E., Hinton, G. E., and Sejnowski, T. J. (1983) Massively parallel architectures for A.I.: Netl, Thistle, and Boltzmann machines. *Proceedings of the National Conference on Artificial Intelligence*, Washington DC.

243. Hinton, G. E. (1981) Shape representation in parallel systems. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vol 2*, Vancouver BC, Canada.

244. Hinton, G. E. (1981) A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vol 2*, Vancouver BC, Canada.

245. Hinton, G. E. (1981) The role of spatial working memory in shape perception. *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley CA.

246. Sloman, A., Owen, D., Hinton, G., Birch, F., and O'Gorman, F. (1978) Representation and control in vision. *Proceedings of the A.I.S.B. Summer Conference*, Hamburg.

247. Hinton, G. E. (1976) Using relaxation to find a puppet. *Proceedings of the A.I.S.B. Summer Conference*, University of Edinburgh.

Invited Papers

248. Hinton, G. E. (2018) Deep learning: a technology with the potential to transform health care. *Journal of the American Medical Association* [published online August 30, 2018].JAMA. doi:10.1001/jama.2018.11100

249. Vinyals, O., Kaiser, L., Koo, T., Petrov, S., Sutskever, I. and Hinton, G.E. (2015) Grammar as a foreign language. http://arxiv.org/abs/1412.7449v3

250. Lee, Q. L., Jaitly, N. and Hinton, G. E. (2015) A simple way to initialize recurrent networks of rectified linear units. http://arxiv.org/abs/1504.00941

251. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. (2012) Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580

252. Hinton, G. E. (2005) What kind of a graphical model is the brain? *International Joint Conference on Artificial Intelligence 2005.*

253. Hinton, G. E. (2003) The ups and downs of Hebb synapses. *Canadian Psychology*, **44**, pp 10-13.

254. Hinton, G. E., Welling, M., Teh, Y. W., and Osindero, S. (2001) A new view of ICA. In *ICA-2001*, San Diego, CA.

255. Hinton, G. E. and Brown, A. D. (2001) Training many small hidden markov models. WISP-2001 Workshop on Innovation in Speech Processing. *Proceeding of the Institute of Acoustics*, **23**, Part 3.

256. Hinton, G. E. (2000) Modelling High-Dimensional Data by Combining Simple Experts. *AAAI-2000: Seventeenth National Conference on Artificial Intelligence*, Austin, Texas.

257. Hinton, G. E. (1999) Products of Experts. *ICANN 99: Ninth International Conference on Artificial Neural Networks*, Edinburgh. 1-6. Institution of Electrical Engineers, London, UK.

258. R. Grzeszczuk, R. D. Terzopoulos, D., and Hinton, G. (1999) Fast Neural Network Emulation and Control of Dynamical Systems. *Proc. AAAI 1999 Spring Symposium Series: Hybrid Systems and AI: Modeling, Analysis and Control of Discrete + Continuous Systems*, Stanford, CA, March, 1999, 83–88.

259. Hinton, G. E. and Ghahramani, Z. (1997) Towards Neurally Plausible Bayesian Networks. *Proceedings of the 1997 International Conference on Neural Networks* Houston, Texas. *The paper was accidentally ommitted from the proceedings but the conference organizers distributed copies to the attendees.*

260. Hinton, G. E. and Frey, B. J. (1995) Using neural networks to monitor for rare failures. *Proceedings of the* 37^{th} *Mechanical Working and Steel Processing Conference*, Hamilton, Ontario.

261. Hinton, G. E., Dayan, P., To, A. and Neal R. M. (1995) The Helmholtz Machine Through Time. *Artificial Neural Networks V: Proceedings of ICANN-95*, pp 483-490. Elsevier North-Holland.

262. Hinton, G. E., Dayan, P., Neal, R. M., and Zemel, R. S. (1994) Using Neural Networks to Learn Intractable Generative Models. *American Statistical Association, 1994 Proceedings of the Statistical Computing Section.*, American Statistical Association, Alexandria, VA.

263. Hinton, G. E., Plaut, D. C. and Shallice, T. (1993) Simulating Brain Damage *Scientific American, October Issue*

264. Hinton, G. E. and van Camp, D. (1993) Keeping Neural Networks Simple. In: *Artificial Neural Networks III: Proceedings of ICANN-93.* Elsevier North-Holland.

265. Hinton, G. E., Williams, C. K. I., and Revow, M. (1992) Combining Two Methods of Recognizing Hand-Printed Digits. *Artificial Neural Networks II: Proceedings of ICANN-92.* I. Aleksander and J. Taylor (Eds.), Elsevier North-Holland.

266. Hinton, G. E. (1992) How Neural Networks Learn from Experience. *Scientific American, September Issue*

267. McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1987) Une nouvelle approche de la cognition: Le connexionnisme. *Débat*, **47**, Novembre - Decembre.

268. Hinton, G. E. (1987) Learning procedures that construct representations in neural networks. *Kagaku*, **57**, 228–237.

269. Hinton, G. E. and Sejnowski, T. J. (1985) Learning in Boltzmann Machines. *Cognitiva - Colloque Scientifique* Paris.

270. Hinton, G. E. (1985) Learning in parallel networks. *Byte*, April issue.

271. Hinton, G. E. and Sejnowski, T. J. (1984) Learning semantic features. In *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, Boulder, CO.

Book Chapters

272. Hinton, G. E. (2010) Deep Belief Networks In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.

273. Hinton, G. E. (2010) Boltzmann Machines In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.

274. Susskind, J.M., Hinton, G. E., Movellan, J.R., and Anderson, A.K.(2008) Generating Facial Expressions with Deep Belief Nets. In V. Kordic (ed.) *Affective Computing, Emotion Modelling, Synthesis and Recognition.* ARS Publishers.

275. Hinton, G. E. (2007) To recognize shapes, first learn to generate images. In P. Cisek, T. Drew and J. Kalaska (Eds.) *Computational Neuroscience: Theoretical insights into brain function.* Elsevier.

276. Hinton, G. E. and Brown, A. D. (2002) Learning to Use Spike Timing in a Restricted Boltzmann Machine. In R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki (Eds.) *Probabilistic Models of the Brain.* MIT Press.

277. Hinton, G. E. Sallans, B. and Ghahramani, Z. (1998) Hierarchical Communities of Experts. In M. I. Jordan (Ed.) *Learning in Graphical Models.* Kluwer Academic Press.

278. Neal, R., and Hinton, G. E. (1998) A new view of the EM algorithm that justifies incremental and other variants. In M. I. Jordan (Ed.) *Learning in Graphical Models.* Kluwer Academic Press.

279. Frey, B. J. and Hinton, G. E. (1996) A simple algorithm that discovers efficient perceptual codes. In L. Harris and M. Jenkin (Eds) *Computational and Biological Mechanisms of Visual Coding*, Cambridge University press, New York.

280. Becker, S. and Hinton, G. E. (1995) Using Spatial Coherence as an Internal Teacher for a Neural Network. In Y. Chauvin and D. E. Rumelhart (Eds) Advances in back-propagation. Erlbaum, Hillsdale, NJ.

281. Williams, C. K. I., Revow, M. and Hinton, G. E. (1993) Hand-printed digit recognition using deformable models. In L. Harris and M. Jenkin (Eds) *Spatial Vision in Humans and Robots*, Cambridge University press, New York.

282. Hinton, G. E. and Becker, S. (1992) Using coherence assumptions to discover the underlying causes of the sensory input. In S. Davis (Ed.) *Connectionism: Theory and practice*, Oxford University Press, New York.

283. Hinton, G. E. (1991) The unity of consciousness: A connectionist account. In Kessen, Ortony & Craik (Eds.) Festschrift in honor of George Mandler. Erlbaum, Hillsdale, NJ.

284. Hinton, G. E. and Anderson, J. A. (1989) Introduction to the second edition. In Hinton, G. E. and Anderson, J. A, editors, *Parallel Models of Associative Memory* (second edition), Erlbaum, Hillsdale, NJ.

285. Touretzky, D. S. and Hinton, G. E. (1987) Pattern matching and variable binding in a stochastic neural network. In Davis, L., editor, *Genetic Algorithms and Simulated Annealing*, Pitman, London.

286. Hinton, G. E. Learning to recognize shapes in a parallel network. In Imbert, M., editor, *Proceedings of the 1986 Fyssen Conference*, (Since I sent the finished manuscript in 1986, I have been unable to discover what has happened to this proceedings).

287. Sejnowski, T. J. and Hinton, G. E. (1987) Separating figure from ground using a Boltzmann machine. In Arbib, M. and Hanson, A. R., editors, *Vision, Brain and Cooperative Computation*, MIT Press, Cambridge, MA.

288. Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E. (1986) Parallel distributed models of schemata and sequential thought processes. In McClelland, J. L. and Rumelhart, D. E., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Applications*, MIT Press, Cambridge, MA.

289. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

290. Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986) A general framework. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

291. McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1986) The appeal of parallel distributed processing. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

292. Hinton, G. E. and Sejnowski, T. J. (1986) Learning and relearning in Boltzmann machines. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

293. Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986) Distributed representations. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

294. Hinton, G. E. (1984) Some computational solutions to Bernstein's problems. In Whiting, H., editor, *Human Motor Actions: Bernstein Reassessed*, North-Holland, New York.

295. Hinton, G. E. and Parsons, L. A. (1981) Frames of reference and mental imagery. In Long, J. and Baddeley, A., editors, *Attention and Performance IX*, Erlbaum, Hillsdale, NJ.

296. Hinton, G. E. (1981) Implementing semantic networks in parallel hardware. In Hinton, G. E. and Anderson, J. A., editors, *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ.

297. Anderson, J. A. and Hinton, G. E. (1981) Models of information processing in the brain. In Hinton, G. E. and Anderson, J. A, editors, *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ.

Books

298. Hinton, G. E. and Sejnowski, T. J. (Editors) *Unsupervised Learning: Foundations of Neural Computation*. MIT Press, Cambridge, Massachusetts, 1999.

299. Hinton G. E. (Ed.) Connectionist Symbol Processing. 1990 Special issue of the journal *Artificial Intelligence* issued as a book by MIT press in 1991.

300. Touretzky, D. S., Elman, J., Sejnowski, T. J. and Hinton, G. E. (Eds.) Proceedings of the 1990 Connectionist Models Summer School, Morgan Kauffman: Los Altos, CA, 1990.

301. Touretzky, D. S., Hinton, G. E. and Sejnowski, T. J. (Eds.) Proceedings of the 1988 Connectionist Models Summer School, Morgan Kauffman: Los Altos, CA, 1988.

302. Hinton, G. E. and Anderson, J. A. (Eds.) *Parallel Models of Associative Memory* Hillsdale, NJ: Erlbaum, 1981.
(Updated second edition, 1989).

Technical Reports

*Technical reports that were subsequently published as papers or chapters are marked with a * and are not numbered.*

303. Susskind, J., Anderson, A. and Hinton, G. E. (2010) The Toronto Face Database. Technical Report UTML TR 2010-001, University of Toronto.

304. Sminchisescu, C., Welling, M., and Hinton, G.E. (2003) A Mode-Hopping MCMC Sampler Technical Report CSRG-478, University of Toronto.

* Hinton, G. E. (2000) Training Products of Experts by Minimizing Contrastive Divergence. Technical Report GCNU 2000-004, Gatsby Computational Neuroscience Unit, University College London.

* Paccanaro, A and Hinton, G. E. (2000) Learning Distributed Representation of Concepts using Linear Relational Embedding. Technical Report GCNU 2000-002, Gatsby Computational Neuroscience Unit, University College London.

305. Hinton, G. E. and Revow, M. (1997) Using Mixtures of Factor Analyzers for Segmentation and Pose Estimation. (available at http://www.cs.toronto.edu/ hinton/papers.html)

306. Rasmussen, C. E., Neal, R. M., Hinton, G. E., van Camp, D, Revow, M.. Ghahramani, Z., Kustra, R, and Tibshirani, R. (1996) The DELVE Manual. Department of Computer Science, University of Toronto

* Ghahramani, Z. and Hinton, G. E. (1996) Switching Mixtures of State space Models. Technical Report CRG-TR-96-3, University of Toronto.

307. Ghahramani, Z. and Hinton, G. E. (1996) Parameter Estimation for Linear Dynamical Systems. Technical Report CRG-TR-96-2, University of Toronto.

308. Ghahramani, Z. and Hinton, G. E. (1996) The EM algorithm for Mixtures of Factor Analyzers. Technical Report CRG-TR-96-1, University of Toronto.

* Galland, C. C. and Hinton, G. E. (1990) Experiments on discovering higher-order features with mean field networks. Technical Report CRG-TR-90-3, Department of Computer Science, University of Toronto, Toronto, Canada.

* Nowlan, S. J. and Hinton, G. E. (1989) Maximum Likelihood Decision-Directed Adaptive Equalization. Technical Report CRG-TR-89-8, Department of Computer Science, University of Toronto, Toronto, Canada.

* Becker, S. and Hinton, G. E. (1990) Using Spatial Coherence as an Internal Teacher for a Neural Network. Technical Report CRG-TR-89-7, Department of Computer Science, University of Toronto, Toronto, Canada.

* Galland, C. G. and Hinton, G. E. (1989) Deterministic Boltzmann Learning in Networks with Asymmetric Connectivity. Technical Report CRG-TR-89-6, Department of Computer Science, University of Toronto, Toronto, Canada.

* Lang, K. and Hinton, G. E. (1988) A time-delay neural network architecture for speech recognition. Technical Report CMU-CS-88-152. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

* Hinton, G. E. (1988) Representing part-whole hierarchies in connectionist networks. Technical Report CRG-TR-88-2, Department of Computer Science, University of Toronto, Toronto, Canada.

* Waibel, A. Hanazawa, T. Hinton, G. Shikano, K. and Lang, K. (1987) Phoneme Recognition Using Time-Delay Neural Networks. Technical Report TR-1-0006. ATR Interpreting Telephony Research Laboratories, Japan.

* Hinton, G. E. and Nowlan, S. J. (1986) How learning can guide evolution. Technical Report CMU-CS-86-128. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

309. Plaut, D., Nowlan, S. and Hinton, G. E. (1986) Experiments on learning by back-propagation. Technical Report CMU-CS-86-126. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

* Hinton, G. E. (1984) Distributed Representations. Technical Report CMU-CS-84-157. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

310. Hinton, G. E., and Smolensky, P. (1984) Parallel computation and the mass-spring model of motor control. Technical Report. Center for Human Information Processing, University of California, San Diego.

* Hinton, G. E., Sejnowski, T. J., and Ackley, D. H. (1984) Boltzmann Machines: Constraint satisfaction networks that learn. Technical Report CMU-CS-84-119, Carnegie-Mellon University.

311. Hammond, N., MacLean, A., Hinton, G., Long, J., Barnard, P., and Clark, I. (1983) Novice use of an interactive graph-plotting system. *Human factors report HF083* IBM (UK) Laboratories, Hursley Park.

312. Hinton, G. E. (1978) Relaxation and its role in vision. PhD Thesis, University of Edinburgh.

Working Papers

313. Dudek, G. and Hinton, G. E. (1993) Navigating without a map by directly transforming sensory inputs into location.

314. Hinton, G. E. (1982) Displays for network management. Applied Psychology Unit report for British Telecom.

315. Hinton, G. E. (1981) Some examples of novices problems with the CHART utility. MRC Applied Psychology Unit internal working paper.

316. Hinton, G. E. (1980) Larger receptive fields give more accurate representations. Program in Cognitive Science internal paper, University of California, San Diego.

317. Hinton, G. E. (1980) Self-tuning feature detectors. Program in Cognitive Science internal paper, University of California, San Diego.

318. Hinton, G. E. (1979) Are mental images like 2-D arrays? Program in Cognitive Science internal paper, University of California, San Diego.

Commentaries, Book Reviews and other Minor Publications

319. Hinton, G. E. (2011) Machine learning for neuroscience. *Neural Systems and Circuits*, **1**, Aug 2011.

320. Hinton, G. E. (2011) A better way to learn features: technical perspective. *Communications of the ACM*, **54**, No. 10, p 94.

321. Hinton, G. E. (2010) Deep Belief Networks In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.(3 pages) *An almost identical entry appears in Scholarpedia*

322. Hinton, G. E. (2010) Boltzmann Machines In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer. (4 pages) *An almost identical entry appears in Scholarpedia*

323. Taylor, G.W., Hinton, G. E. and Roweis, S. (2008) Deep Generative Models for Modeling Animate Motion. Proc. 4th Int. Symp. Adaptive Motion of Animals and Machines.

324. Hinton, G. E. (2003) Neural Networks *Van Nostrand's Scientific Encyclopedia*

325. Hinton, G. E. (2000) Computation by Neural Networks. *Nature Neuroscience Supplement*, **3**, p1170

326. Hinton, G. E. (1999) Supervised Learning in Multilayer Neural Networks In The MIT Encyclopedia of the Cognitive Sciences Edited by Robert Wilson and Frank Keil The MIT Press, Cambridge, Mass.

327. R. Grzeszczuk, D. Terzopoulos, G. Hinton (1997) Learning fast neural network emulators for physics-based models (technical sketch) *Proc. ACM SIGGRAPH 97 Conference*, Los Angeles, CA, August, 1997, in *Computer Graphics Visual Proceedings, Annual Conference Series, 1997*, 167.

328. Hinton, G. E. (1995) Foreword to the book "Neural Networks for Pattern Recognition" by Chris Bishop. Oxford University Press, Oxford.

329. Hinton, G. E. and Nowlan, S. J. (1994) Preface to "Simplifying neural networks by soft weight-sharing". In D. H. Wolpert (Ed.) *The Mathematics of Generalization.* Santa Fe Institute Studies in the Sciences of Complexity.

330. Hinton, G. E. (1990) Review of Aleksander and Morton *Introduction to Neuro-Computing*, In *Nature*, **347**, 627-628.

331. Hinton, G. E., and LeCun, Y. (1988) Review of: R. K. Miller *Neural Networks: Implementing associative memory models in neurocomputers*, In Canadian Artificial Intelligence, **41**.

332. Hinton, G. E. (1987) Models of human inference. Invited commentary on a paper by D. McDermott. *Computational Intelligence*, **3**, 189-190.

333. Hinton, G. E. (1987) Boltzmann Machines. In S. Shapiro (Ed.) *The Encyclopedia of Artificial Intelligence* , New York: Wiley and Sons.

334. Hinton, G. E. (1982) Review of: S. E. Fahlman *NETL: A system for representing and using real-world knowledge.* In *A.I.S.B. Quarterly*, **42/43**.

335. Hinton, G. E. (1985) Three frames suffice. Invited commentary on a paper by J. Feldman. *The Behavioral and Brain Sciences*,

336. Hinton, G. E. (1980) Inferring the meaning of direct perception. Invited commentary on a paper by Ullman, S. *The Behavioral and Brain Sciences*, **3**, 387-388.

337. Hinton, G. E. (1979) Imagery without arrays. Invited commentary on a paper by S. M. Kosslyn, S. Pinker, G.E. Smith, and S. P. Shwartz. *The Behavioral and Brain Sciences*, **2**, 555-556

338. Hinton, G. E. (1979) Report on The La Jolla Conference on Cognitive Science, In *A.I.S.B. Quarterly*, **35**.

339. Hinton, G. E. (1979) Review of: D. C. Dennett *Brainstorms.* In *Contemporary Psychology*, **24**, 746-748.

340. Hinton, G. E. (1979) Review of: E. L. J. Leeuwenberg and H. F. J. M. Buffart (Eds.) *Formal theories of visual perception.* In *Journal of the Optical Society of America*, **69**, p.1492.

341. Hinton, G. E. (1978) Review of: J. Metzler (Ed.) *Systems Neuroscience.* In *Perception*, **7**, 364-365.

Invited Conference Addresses

this does not include talks accompanying refereed conference papers.

1981
Seventh International Joint Conference on A. I., Vancouver BC, Canada.
Third Annual Conference of the Cognitive Science Society, Berkeley CA.
Annual Meeting of the British Experimental Psychology Society, Oxford.

1983
Fifth Annual Conference of the Cognitive Science Society, Rochester NY.

1984
Conference on "Statistical Physics in Computer Engineering and Models of Perception."
IBM Watson Research Center, Yorktown Heights NY.
Conference on "Visual Perception." Lake Ontario Visual Establishment, Canada.
Sixth Annual Conference of the Cognitive Science Society, Boulder CO.
Workshop on "Cognitive Neuroscience", Moorea.

1985
British Pattern Recognition Society (Keynote lecture plus additional lecture)
"From Neurons to Thought", Florence Center for the History and Philosophy of Science.
Society for Artificial Intelligence and the Simulation of Behaviour.
Cognitiva 85. Paris

1986
American Association for Artificial Intelligence, Philadelphia PA.
Eighth Annual Conference of the Cognitive Science Society, Amherst Mass.
Conference on "Organizing Principles of Sensory Processing." Pittsburgh.

1987
Neural Network Conference, Snowbird, Utah.
Philosophy of Mind Conference, Pittsburgh PA.
Annual Meeting of the British Experimental Psychology Society, Oxford.
European Psychology Society Summer School, Bernried, West Germany

1988
Canadian Society for Artificial Intelligence, Edmonton.
McDonnell Foundation workshop on vision, Monterey, CA.
Machine Learning Conference, Ann Arbor.
Workshop on Parallel Distributed Processing, Emory University (4 lectures).
Connectionist Summer School, Carnegie-Mellon.
Conference on "How the Brain Works", Boston (Organized by *Nature*)
Canadian Institute for Artificial Intelligence meeting, Whistler, BC
International Neural Network Society Conference, Boston
Canadian Institute for Artificial Intelligence meeting, Mt. St. Marie, Quebec
Fourth biennial conference on cellular mechanisms of conditioning and behavioral plasticity
Toronto, Canada.

1989
First International Conference on Knowledge Representation, Toronto, Ontario
International Joint Conference on A. I., Detroit, Michigan.
Annual Conference of the Cognitive Science Society, Ann Arbor, Michigan.

1990
International Joint Conference on Neural Networks, Washington DC.
Conference on "Connectionism: Theory and Practice." Vancouver, BC.
Workshop on Neural Networks, Niagara, Ontario

1991
Snowbird Neural Networks Conference, Utah
Neural Network Summer School, Cambridge, UK
International Joint Conference on Neural Networks, Seattle, WA.

Cognitive Science Society Conference, Chicago, MI.
Conference on Computational Learning Theory and Natural Learning Systems, Berkeley

1992
Neural Networks Applications Forum, Birmingham UK
Canadian Institute for Artificial Intelligence meeting, Montreal
Applications of Artificial Intelligence in Engineering, Waterloo, Ontario
American Statistical Association, Boston, Mass.
Santa Fe Institute meeting on supervised learning, Santa Fe, NM.
International Conference on Neural Networks, Brighton, UK
Neural Network Summer School, Cambridge, UK

1993
Australian Neural Networks Society Conference, Melbourne
Australian Cognitive Science Society Conference, Melbourne
Third International Workshop on Frontiers of Handwriting Recognition, Buffalo, NY
Cognitive Science Society Conference, Boulder, Colorado
Connectionist Summer School, Boulder, Colorado
Computational Learning Theory Conference (COLT-93), Santa Cruz, CA
International Conference on Neural Networks, Amsterdam
Neural Network Summer School, Cambridge, UK

1994
The Royal Society of Canada, Symposium, Calgary
Cognitive Science Summer School, Buffalo
American Statistical Association, Toronto
American Association for Artificial Intelligence, Seattle
Neural Information Processing Conference, Denver

1995
Machines that Learn, Snowbird, Utah.
Conference on Visual Coding, York University, Ontario
Neural Network Summer School, Cambridge, UK
37^{th} Metal Working and Steel Processing Conference, Hamilton, Ontario
Workshop on Belief nets, Vail Colorado

1996
Workshop on Neural Networks, Montreal.
Connectionism for Cognitivists Conference, Carleton University, Ottawa
Workshop on Neuromorphic Engineering, Telluride, Colorado
Cognitive Science Society Conference, San Diego, CA.

1997
Royal Society Meeting on the role of knowledge in perception. London, UK.
Canadian Applied Mathematics Society Conference. Toronto
IEEE International Conference on Neural Networks. Houston, Texas
Canadian Workshop on Information Theory. Toronto
Nato workshop on Neural Networks and Generalization. Cambridge, UK

1998
NEC/NYU Workshop on Learning in Vision, New York

1999
Connectionist Models Summer School, Oxford.

Cognitive Neuroscience Autumn School, Oxford.
International Conference on Artificial Neural Networks, Edinburgh

2000
NICE 2000, Grindelwald, Switzerland
Neural networks for Computing, Snowbird, Utah
Model selection and learning in Computer vision, Grasmere, UK
Artificial Intelligence and the Simulation of Behaviour, Birmingham
Neural Applications Computing Forum, London
Tenth anniversary neural networks workshop, Kings College London
XIII International Congress on Mathematical Physics, London
American Association for Artificial Intelligence, Austin, Texas
Cambridge Philosophical Society, Cambridge

2001
Workshop on Innovation in Speech Processing, Stratford-on-Avon
International Conference on Artificial Neural Networks and Genetic Algorithms, Prague
European Congress of Psychology, London
Conference on Uncertainty in Artificial Intelligence, Seattle
Third International Workshop on Energy Minimization Methods
in Computer Vision and Pattern Recognition, Sophia-Antipolis
Autumn School in Cognitive Neuroscience, Oxford
ICA-2001, San Diego

2002
CIAR Workshop on Machine Learning, Toronto.
KDD-02 (Data-Mining Conference), Edmonton
CASCON workshop on machine learning, Toronto

2003
Uncertainty in Artificial Intelligence, Key West, Florida
Workshop on "The Probabilistic Brain", Cambridge, UK
CIAR Workshop on Computational Neuroscience, Vancouver

2004
Symposium on the Mind, University College, University of Toronto
Southern California Neural Computation Workshop (keynote address), UC Irvine
Workshop on Object Recognition, Sicily
Workshop on learning for object recognition, Vancouver BC

2005
Robotics: Science and Systems, Cambridge Mass. June 2005.
CIAR workshop on the perception of human motion, Toronto, July 2005
International Joint Conference on AI, Edinburgh, Aug 2005.
Society for Computers in Psychology (keynote address), Toronto, Nov 2005.
Sunnybrook Artificial Intelligence in Medicine workshop, Toronto, Oct 2005.
CIAR workshop on sequential vision, Vancouver, Dec 2005
CIAR workshop on learning energy-based models, Vancouver, Dec 2005

2006
CIAR workshop on computational neuroscience, Alton, Ontario, April 2006
XXVIIIth International Symposium on computational neuroscience, Montreal, May 2006
Canadian AI conference (keynote talk), Quebec City, June 2006
CIAR workshop on "Learning and Vision", Toronto, Oct 2006

CIAR workshop on Cortical Learning Mechanisms, Vancouver, Dec 2006

2007
Poster (with S. Becker) at Computational and Systems Neuroscience, Utah, Feb 2007
Frontiers of Theoretical Neuroscience, Waterloo, April 13 2007
"Brain Day", Waterloo, April 27 2007
Microsoft Research Laboratory Cambridge 10 year celebration, invited plenary talk, June 2007.
Talk at CIAR workshop on "Achieving Perceptual Invariance", Vancouver BC, Dec 2007.
Talk at NIPS workshop on "Deep Belief Nets", Vancouver, BC, Dec 2007

2008
Talk at CIFAR workshop on "Inhibitory Interneurons", Toronto, ON, Apr 2008
Talk at NIPS workshop on "New Multi-level Models for High-dimensional Sequential Data", Whistler, BC, Dec 2008
Talk at CIFAR workshop on "Recent Developments in Deep Learning", Vancouver, BC, Dec 2008

2009
The Ed Posner lecture (plenary invited talk) at NIPS, Vancouver, BC, Dec 2009
Talk at CIFAR workshop on "Image retrieval using deep auto-encoders", Vancouver, BC, Dec 2009
talk at NIPS workshop on "Deep learning for speech recognition", Whistler, BC, Dec 2009.

2010
Talk at Computational and Systems Neuroscience 2010, Salt Lake City, Utah, Feb 2010.
Talk at First Annual Mini-Symposium, Center for Mind, Brain, and Computation. Stanford university, March 2010.
Talk at a conference on "Taming Complexity", CUNY, New York
The Sam Roweis Memorial Symposium (plenary talk) at NIPS, Vancouver, BC, Dec 2010.
Talk at CIFAR workshop on "Transforming Autoencoders", Vancouver, BC, Dec 2010
talk at Conference on "Hierarchical Distributed Representations", Windsor, UK. July 2011
Talk NIPS workshop talk on "Transforming Autoencoders", Grenada, Spain (by phone) Dec 2011
Talk at CIFAR joint workshop on "Neural Development", Alton, Ontario, April 2012.

2012
Plenary invited talk at NIPS, Lake Tahoe, CA
Talk at Workshop on Deep Learning, NIPS, Lake Tahoe, CA
Talk at Gairdner Foundation Symposium, Toronto.
Talk at 50th anniversary symposium of IBBME, Toronto.
Talk at CIFAR joint workshop on "Neural Development", Alton, ON.
Talk at CIFAR workshop, Tahoe, CA.

2013
Plenary invited talk at ICASSP-13, Vancouver, Canada.

2014
Invited talk at "Baylearn", Berkeley.
Invited talk at Berkeley summer course in mining and modeling of neuroscience data. MSRI, Berkeley.

2015
Keynote Talk, Royal Society Meeting on Machine Learning, London, UK.

2016
Invited talk, Rotman Business School Conference on AI, Toronto
Keynote talk, University Health Network annual retreat, Toronto

2017
Invited talk, Go North meeting, Toronto
Keynote talk: Deep learning Summit, Montreal
Invited talk, Economics Workshop, Toronto

2018
keynote talk, Thompson-Reuters Conference, Toronto
Invited talk, Rotman Business School Conference on AI, Toronto
Keynote talk: Deep learning Summit, Toronto

I currently get several invitations per week to talk at conferences or universities, but I have not been free to travel so I decline them.

Invited Lectures

1978
University of Maryland, Computer Vision Group

1979
University of California at Berkeley, Computer Science Department
Yale University, Computer Science Department
Brown University, Cognitive Science Group
University of Rochester, Cognitive Science Group
Bolt, Beranek and Newman, Cambridge Mass
Applied Psychology Unit, Cambridge England
Sussex University, Cognitive Studies Group
Warwick University, Psychology Department
Bristol University, Psychology Department

1980
Stanford University, A.I. group
University of California at Berkeley, Psychology Department
University of California at Berkeley, Computer Science Department
IBM, San Jose
Stanford University, Psychology Department
Lawrence Berkeley Laboratory
Brown University, Cognitive Science Group
University of Rochester, Computer Science Department
M.I.T, Artificial Intelligence Laboratory
Essex University, Cognitive Science Group
Sussex University, Cognitive Studies Group
Cambridge University, Psychology Department

1981
University of California at Berkeley, Computer Science Department
University of California at San Diego, Computer Science Department
Hewlett Packard, Palo Alto
Stirling University, Psychology Department
Carnegie-Mellon University, Computer Science Department
Carnegie-Mellon University, Psychology Department
Sussex University, Cognitive Studies Group

System Development Foundation, Palo Alto
Manchester University, Psychology Department
London University, Birkbeck College
Hatfield Polytechnic, Psychology Department
Warwick University, Psychology Department

1983
University of Pennsylvania, Computer Science Department
M.I.T, Artificial Intelligence Laboratory
University of Rochester, Computer Science Department
University of California at San Diego, Center for Human Information Processing
Fairchild AI Lab, Palo Alto
University of Pennsylvania, Psychology Department
University of California at Berkeley, Cognitive Science Group
California Institute of Technology, Parallel Computation Group
Bell Labs, Murray Hill New Jersey

1984
Applied Psychology Unit, Cambridge England
University of Pennsylvania, Psychology Department
University of Pittsburgh, Neuroscience Program
Sussex University, Cognitive Studies Group
Brown University, Psychology Department
Harvard University, Computer Science Department
Bolt, Beranek and Newman, Cambridge Mass
IBM, Yorktown Heights, Physics Group
Princeton University, Cognitive Science Group
M.I.T., Cognitive Science Seminar
University of Toronto, Computer Science Department
University of Toronto, Psychology Department

1985
Sussex University, Institute for Cognitive and Information Sciences
Applied Psychology Unit, Cambridge, England
University of Rhode Island, Computer Science Department
The Open University, England, Psychology Department
Imperial College London, Computer Science Department
University of Michigan, Psychology Department
Georgia Institute of Technology, Computer Science Department
Johns Hopkins University, Psychology Department
Yale University, Computer Science Department
M.I.T, Artificial Intelligence Laboratory

1986
Imperial College London, Computer Science Department (4 lectures)
Edinburgh University, Artificial Intelligence Department
Oxford University, Psychology Department
Applied Psychology Unit, Cambridge, England
Center for the Study of Systems and Advanced Technology, Paris
Sussex University, Institute for Cognitive and Information Sciences
M.I.T, Artificial Intelligence Laboratory
Memphis State University, Psychology Department (2 lectures)
Bolt, Beranek and Newman, AI group and Speech Recognition group

1987
Stanford University, Psychology Department
University of California at Berkeley, Psychology Department
University of California at Berkeley, Cognitive Science Group (2 talks)
University of California at Santa Barbara, Cognitive Science Group
University of California at Los Angeles, Cognitive Science Group
University of California at Los Angeles, Computer Science Department
University of California at San Diego, Center for Human Information Processing
Columbia University, Psychology Department
University of Toronto, Psychology Department
University of Toronto, Computer Science Department
University of Maryland, Computer Science Department
York University, Psychology Department
York University, Vision Group
University of Rochester, Computer Science Department
Cornell University, Computer Science Department

1988
University of Sussex, Psychology Department (2 talks)
University of Western Ontario, Cognitive Science Group
Oregon Graduate Center, Beaverton
University of Oregon, Psychology Department
University of Toronto, Anatomy Department
Brown University, Computer Science Department
Brown University, Philosophy Department
McMaster University, Psychology Department
University of Waterloo, Psychology Department
Ontario Institute for Studies in Education
International Computer Science Institute, Berkeley

1989
IBM Canada, Toronto
Carnegie-Mellon University, Computer Science Department
Queens University, Psychology Department
University of Waterloo, Computer Science Department
Stanford University, Psychology Department
McGill University, Computer Science Department
McGill University, Psychology Department
University of Manitoba, Computer Science Department
University of Manitoba, Psychology Department
University of Toronto, Non-Linear Systems Group

1990
University of California, Berkeley, Cognitive Science Group
Synaptics Inc., San Jose
International Computer Science Institute, Berkeley
University of Toronto, Physics Department Colloquium
University of California, San Diego, Institute for Neural Computation
California Institute of Technology, Division of Biology
University of Pennsylvannia, Cognitive Science Group
MIT, Department of Cognitive and Brain Sciences
Brandeis University, Computer Science Department
Purdue University, Computer Science Department
University of Indiana, Psychology Department

University of Illinois, Computer Science Department
Northwestern University, Psychology Department
University of Toronto, Biomedical Engineering Department
NCR Waterloo, Recognition Modules Group
SUNY Buffalo, Computer Science Department

1991
Rotman Institute, Toronto, Neuropsychology Group
University of North Carolina, Cognitive Science Group
North Carolina Triangle Park Neural Networks Group
University of California, San Diego, Computer Science Department
Synaptics Inc., San Jose
Bell Northern Research, Ottawa
The Salk Institute, California
University of Wisconsin, Computer Science Department
University of Waterloo, Electrical Engineering Department
Synaptics Inc., San Jose
Apple Computer, San Jose
University of Toronto, Semiotics Group
Yale University, Computer Science Department
University of Massachusetts, Computer Science Department
University of Toronto, Statistics Department Colloquium

1992
University of Ottawa, Physics Department
NCR Waterloo, Recognition Modules Group
McGill University, Electrical Engineering Department
Montreal Neurological Institute
Oxford University, Psychology Department
University of Cardiff, Computer Science Department
Oxford University, Department of Theoretical Physics
Synaptics Inc., San Jose
University of British Columbia, Computer Science Department
University of British Columbia, Physics Department
SUNY Buffalo, Center of Excellence for Document Analysis and Recognition
University of Rochester, Computer Science Department

1993
University of Sydney, Psychology Department
MIT, Department of Brain and Cognitive Sciences
MIT, Center for Biological and Computational Learning
Queens University, Computer Science Department
Queens University, Cognitive Science
Carleton University, Computer Science Department
IEEE Chapter, Waterloo, Ontario
Synaptics Inc., San Jose
International Computer Science Institute, Berkeley
Carnegie-Mellon University, Psychology Department (2 lectures)

1994
University of Sussex, Cognitive Studies Group
University of Waterloo, Cognitive Science
Sunnybrook Hospital, Toronto, Imaging Research Group
Northwestern University, Institute for Learning Studies

International Center for Advanced Studies, Trieste, Italy

1995
Princeton University, Psychology Department
Clark Institute of Psychiatry, Toronto
University of California, Berkeley, Cognitive Science Group
University of California at Berkeley, Statistics Department
Synaptics Inc., San Jose
MIT, Department of Brain and Cognitive Sciences
Stanford University, Electrical Engineering Department
MIT, Artificial Intelligence Laboratory
Aston University, Neural Networks Group (3 talks)
Fields Institute, Toronto, Financial Mathematics lecture series

1996
Carnegie-Mellon University, Computer Science Department
University of Toronto, Physiology Department
AT&T Research, Machine Learning Group
California Institute of Technology, Neural Computation Program
CIPS/CIAR meeting, Toronto
Brandeis University, Computer Science Department

1997
Imperial College London, Neural networks group
University of Colorado at Boulder, Psychology Department
University of Colorado at Boulder, Computer Science Department
Cornell University, Psychology Department
Cornell University, Cognitive Science Group

1998
University of Pennsylvania, Computer Science Department
Brown University, Computer Science Department
Caltech, Neuromorphic Engineering Group
Salk Institute, Computational Neurobiology Laboratory
University College London, Department of Statistical Science
Oxford University, Cognitive Neuroscience group
Applied Psychology Unit, Cambridge

1999
University College London, Computer Science Department
University of Edinburgh, Cognitive Science Group
University of Edinburgh, Computer Science
University of Bristol, Psychology Department
University of Glasgow, Statistics Department
Imperial College London, Neural networks group
University College London, Institute of Cognitive Neuroscience
University of Toronto, Department of Computer Science
University of Warwick, Psychology Dept
University of Edinburgh, Electrical Engineering
Microsoft Research Laboratories, Cambridge

2000
Cambridge University, Astrophysics Department
Rutgers University, Cognitive Science Department

University of Exeter, Computer Science.
University of Colorado at Boulder, Computer Science
University of Waterloo, Computer Science
University of Toronto, Computer Science
University of California at Berkeley, Computer Science
Salk Institute, Computational Neurobiology Laboratory
University of Montreal
University of Southampton, Electronics and Computer Science

2001
Sussex University, COGS department (2 talks)
Stanford University, Computer Science and Statistics
University of Oxford, Engineering Science

2002
Yale University, Computer Science Department
University of Montreal, Dept. IRO
McGill University, Psychology Department
University of Chicago, Departments of Statistics and Computer Science
York University, Vision Group
University of Toronto, Computational Neuroscience Group
Georgetown University, Medical School
Johns Hopkins, Speech and Language group
Ohio State University, Computer Science
MIT, AI laboratory
Carnegie-Mellon University, Computer Science Department
UC Irvine, Computer Science Department
Microsoft Research Laboratories, Redmond
University of Alberta, Computer Science Department
University of British Columbia, Computer Science Department

2003
University of California San Diego, Institute for Neural Computation
Redwood Neuroscience Institute, CA
Stanford University, Computational Neuroscience Lecture Series
University of Rochester, Computer Science

2004
Salk Institute, Computational Neurobiology Lab.
UC San Diego, Computer Science Department
Caltech, Neural Computation program
Redwood Neuroscience Institute, CA (2 talks)
Microsoft Research Laboratory, Cambridge
MRC Cognition and Brain Sciences Unit, Cambridge, UK
Gatsby Computational Neuroscience Unit, University College London
Cambridge University, Physics Department
Simon Fraser University, Computer Science Department

2005
University of Montreal, , Dept. IRO
MIT, AI Laboratory
Microsoft Research Laboratory, Cambridge
Gatsby Computational Neuroscience Unit, University College London
Cambridge University, Inference and Learning Group, Physics Department

2006
McGill University, Psychology Department
University of Toronto, Scarborough, Mathematical Sciences Dept
University of Montreal, Dept. IRO
McGill University, Physics Department

2007
New York University, Computer Science Department
Columbia University, Theoretical Neuroscience Group
Carnegie-Mellon University, Computer Science Department
Microsoft Research Laboratory, Cambridge U.K (via video link)
MacMaster University, Electrical Engineering Group
Salk Institute, Computational Neurobiology Lab.
UC San Diego, Computer Science Department
UC Irvine, Computer Science Department
Caltech, Neural Computation program
UC Berkeley, Redwood Neuroscience Institute
Stanford University, Computer Science Department
Google, Mountain View, CA
UC Berkeley, Electrical Engineering and Computer Science

2008
University of Waterloo, Statistics Department
Brown University, Compter Science Department
MIT, Brain and Behavioural Sciences (90 mins)
MIT, Electrical and Computer Engineering Department
IBM, Watson Research Center, Yorktown Heights
New York University, Computer Science Department
McGill University, CRM Applied Mathematics Seminar
University of Montreal, Dept. Computer Science and Operations Research
University of British Columbia, Dept. Computer Science

2009
Toyota Technology Institute, Chicago
University of Montreal, Machine Learning Group.
Cambridge University, Inference and Learning Group, Physics Department
Microsoft Research Laboratory, Cambridge
Sussex University, Informatics Department
University of Oxford, Engineering Science
Edinburgh University, Informatics Department
Gatsby Computational Neuroscience Unit, University College London
University of Hertfordshire, Computer Science Department
Manchester University, Computer Science Department (two talks)
University of Quebec at Montreal, Cognitive Science Group.
Dalhousie University, School of Computer Science
"C2C" video lecture broadcast to universities across Canada
Microsoft Research Redmond (talk on object recognition)
Microsoft Research Redmond (talk on image retrieval)
Microsoft Research Redmond (talk on language modeling)

2010
Queens University, Neuroscience Program
University of Waterloo, Computer Science

Stanford for University, Statistics Department
University of California, Los Angeles, Computer Science
University of California, Irvine, Computer Science
Caltech, Computer Vision Group
University of California, Berkeley, Redwood Neuroscience Institute
Google, Mountain View, Machine Learning Group
Johns Hopkins University, Computer Science
MIT, Behavioural and Brain Sciences
IBM Thomas J. Watson Research Labs, Speech group
Microsoft Research, Redmond
MacMaster University, Origins Institute

2011
IBM Thomas J. Watson Research Labs, Speech group
Columbia University, Theoretical Neuroscience Group
Princeton University, Computer Science Department
Microsoft Research Laboratory, Cambridge, UK.
University of Sheffield, Computer Science Department
Edinbugh University, Informatics Department
University College London, Computer Science Department
Cambridge University, Engineering Department
University of Birmingham, Computer Science Department
University of Swansea, Computer Science Department
Sussex University, Informatics Department

2012
Rockefeller University, Center for Studies in Physics and Biology
New York University, Computer Science Department
University of Toronto, Department of Computer Science faculty lecture series
Stanford University, Computer Science Dept.
McGill University, Computer Science Dept.

2013
University of British Columbia, Computer Science Dept.
Stanford University, Statistics Dept.
University of Toronto, Institute of Biomedical Engineering

2014
Toyota Technology Institute, Chicago
University of California, San Diego, Computer Science Department
University of California, Berkeley, Computer Science
University of California, Irvine, Computer Science
Caltech, Computer Vision Group
Netflix, Los Gatos
University of Toronto, Physics Department
MIT, Behavioural and Brain Sciences
MIT, Electrical and Computer Engineering Department

2015
Oxford University, Statistics Department
Kings College London, Institute of Psychiatry
Cambridge University, Department of Engineering
Google DeepMind, London (3 talks)
Georgia Tech, Computer Science

Rutgers, Computer Science

2016
Toronto General Hospital
Princess Margaret Hospital
Fields Institute

2017
IIT Madras
Fields Institute
Engineering Science, Toronto

2018
Invited talk on AI, University of Toronto Governing Council

Grants Awarded

1982:
$462,011 from the System Development Foundation for the period Oct 1 1982 to Sept 30 1985. The grant was for studying ways of organizing computation in massively parallel networks of neuron-like processors.

1983:
$62,000 from the Alfred P. Sloan Foundation for the period Jan 1 1984 to Dec 31 1985 (Jointly with T. J. Sejnowski). The grant was for running a series of workshops.

1984:
$61,709 from the System Development Foundation for the purchase of a Symbolics 3600 Lisp Machine and for organizing a small workshop.

1985:
$47,503 from the System Development Foundation to continue research on Boltzmann Machines for the period Oct 1 1985 to Dec 31 1985.

1986:
$394,843 from the Office of Naval Research for the period Jan 1 1986 to Dec 31 1987 (jointly with J.L. McClelland). The grant was for studying "Learning in massively parallel networks".

$70,735 from the National Science Foundation for the period May 1 1986 to April 30 1987. The grant was for studying "Search methods for massively parallel networks".

1987:
$70,735 from the National Science Foundation for the period May 1 1987 to April 30 1988. Renewal of previous NSF grant.

$200,000 (US) from the Alfred P. Sloan Foundation for the period Jan 1 1987 to Dec 31 1989 (Jointly with M. Farah). The grant was for studying "The Neural Basis of Spatial Cognition".

$108,000 ($36,000 per year for three years) from the Natural Science and Engineering Research Council of Canada for the period May 1 1987 to April 30 1990. Operating grant.

$36,157 from the Natural Science and Engineering Research Council of Canada. Equipment grant.

1988:
$148,527 from the Information Technology Research Center.

$65,858 from the Natural Science and Engineering Research Council of Canada. Equipment grant.

$50,000 (US) from DuPont for investigating applications of artificial neural networks.

$50,000 (US) from Apple for investigating applications of artificial neural networks to speech recognition.

$62,500 (US) from DuPont. Contract for "Neural nets, parallel computation, and the prediction of structures for biological miolecules

1989:
$212,000 from the Information Technology Research Center.

$50,000 (US) from Apple for continued investigation of applications of artificial neural networks to speech recognition.

1990:
$279,517 from the Institute for Robotics and Intelligent Systems (A federal network of centres of excellence) for applying connectionist learning techniques to the development of adaptive interfaces.

$166,284 ($55,428 per year for three years) from the Natural Science and Engineering Research Council of Canada for the period May 1 1990 to April 30 1993. Operating grant.

$51,000 from the Natural Science and Engineering Research Council of Canada. Equipment grant.

$175,800 ($58,600 per year for three years) from the Natural Science and Engineering Research Council of Canada. Strategic Grant for developing the Xerion neural network simulator.

1991:
$218,000 from the Information Technology Research Center (jointly with D. Terzopoulos).

$50,000 (US) from Apple for research on handwritten character recognition using neural networks.

1993:
$185,250 ($61,750 per year for three years) from the Natural Science and Engineering Research Council of Canada for the period May 1 1993 to April 30 1996. Operating grant.

$55,467 from the Natural Science and Engineering Research Council of Canada. Equipment grant.

$250,000 from the Information Technology Research Center. Two year grant.

$9,800 from Burns Fry Analytics for the period June 1 to Dec 31 1993 for applying neural networks to financial forecasting.

$214,000 (over 3 years) from the Natural Science and Engineering Research Council of Canada. Strategic grant.

1994:
$669,376 (over 4 years) from the Institute for Robotics and Intelligent Systems, a federal network of centres of excellence (with R. Tibshirani, D. Lowe and M. Leblanc).

1995:

$123,593 from the Natural Science and Engineering Research Council of Canada. Equipment grant (I was the principal investigator and there were 17 co-investigators).

$186,000 (over 2 years) from the Information Technology Research Center

1996:
$412,500 ($82,500 per year for five years) from the Natural Science and Engineering Research Council of Canada for the period Apr 1 1996 to Mar 31 2001. Research grant.

$10,000 from Ontario Hydro Technologies.

1997:
$71,803 from the Natural Science and Engineering Research Council of Canada for equipment.

1998:
10,000,000 pounds for the period 1998-2008 from the Gatsby Foundation

2002:
$484,000 ($121,000 per year for five years) from the Natural Science and Engineering Research Council of Canada for the period Apr 1 2002 to Mar 31 2006. Research grant.

$487,714 from the Canada Foundation for Innovation and the Ontario Innovation Trust for the period Aug 2002 to July 2006 for the purchase of computing equipment and software.

$1,400,000 ($200,000 per year for seven years) for a Tier 1 Canada Research Chair in Machine Learning.

2003:
$1,500,000 ($300,000 per year for 5 years from CIHR.
"Modern Statistical Approaches in High-Throughput Genomic Research"
with M. Escobar (principal applicant), C. Boone, T. Hughes, T. Westwood and R. Kustra

$10,000 from the Canadian Institute for Advanced Research to develop a proposal for a program in "Neural Computation and Adaptive Perception".

2004:
$2,500,000 (for a 5 year program) from the Canadian Institute for Advanced Research. The funding is for a program on "Neural Computation and Adaptive Perception" that I will direct. The program will involve researchers across Canada and international researchers.

2007:
$610,000 ($122,000 per year for five years) from the Natural Science and Engineering Research Council of Canada for the period Apr 1 2007 to Mar 31 2012. Research grant.

$10,000 unsolicited gift from Microsoft Research for "the general support of Hinton's research".

$50,000 from Merfin LLC (a US company) to support a postdoc for part of 2008.

2008:
$40,000 gift from Google. For research on Semantic Hashing.

$194,000 ($97,000 per year for two years) from the Natural Science and Engineering Research Council of Canada for the period Apr 1 2008 to Mar 31 2010. PI of Strategic Research Grant (jointly with Richard Zemel)

2009:
$2,500,000 (for a 5 year renewal) from the Canadian Institute for Advanced Research. The funding is for a program on "Neural Computation and Adaptive Perception" that I direct. The program involves researchers across Canada and international researchers.

$94,028 from the Canada Foundation for Innovation and the Ontario Innovation Trust for the purchase of computing equipment and software.

2010:
$20,000 another unsolicited gift from Microsoft Research.

2011:
$390,000 ($78,000 per year for five years) from the Natural Science and Engineering Research Council of Canada to increase my NSERC discovery grant to $200,000 for the period Apr 1 2011 to Mar 31 2016 (Gerhard Herzberg Canada Gold Medal Award.)

$ 59,951 from the Natural Science and Engineering Research Council of Canada. Equipment grant. (I was the PI and Richard Zemel was the co-PI). April 2011 - March 2012.

$20,000 yet another unsolicited gift from Microsoft Research.

$22,250 gift from Google. For research on speech recognition.

2012:
$610,000 ($122,000 per year for five years) from the Natural Science and Engineering Research Council of Canada for the period Apr 1 2012 to Mar 31 2017. Research grant.

$22,250 another gift from Google for research on speech recognition.

2013:
$1000,000 focussed research grant from Google for 2013 and 2014. I am the PI and the co-PIs are Richard Zemel and Ruslan Salakhutdinov at Toronto, Yann LeCun at NYU, and Yoshua Bengio at U. Montreal.

Service to the Research Community

Editorial Boards

Currently: Neural Computation, The Journal of Machine Learning Research
Formerly: Artificial Intelligence, Cognitive Science, Machine Learning

Governing boards & committees outside my university

McDonnell Foundation committee on Attention and Perception (1987-88)
Cognitive Science Society Governing Board (1986-1991)
President, Cognitive Science Society (1992-1993)
American Association for Artificial Intelligence Governing Council (1988-1991)
American Association for Artificial Intelligence, Fellows Selection Committee (1993-1996)
Information Technology Research Center of Ontario: Area coordinator for AI, (1990-1995)

Conferences Organized or Co-Organized

"Dense Representations", La Jolla, Jan 1979
"Associative Memory and Parallel Computation", La Jolla, June 1979
"Parallel Models of Biological Computation", Carnegie-Mellon, June 1983.
"Stochastic Parallel Computation", Boston, May 1984.
"Connectionist Models of Symbol Processing", Maryland, Oct 1985
"The Connectionist Summer School", Pittsburgh, June 1986
"Connectionist Models of Attention and Perception", Pajaro Dunes, CA, Jan 1988.
"The Connectionist Summer School", Pittsburgh, June 1988
"The Connectionist Summer School", San Diego, June 1990
"Mixture Day", Gatsby Computational Neuroscience Unit, May 1999
"Inference and Learning", Gatsby Computational Neuroscience Unit, May 2001
CIFAR workshop on "The Future of Machine Learning", Toronto, July 2002
CIFAR workshop on "Neural Computation", Vancouver BC, Dec 2003
CIFAR workshop on "Learning for Object Recognition", Vancouver BC, Dec 2004
CIFAR workshop on "The Perception of Human Motion", Toronto, July 2004
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, July 2005
CIFAR workshop on "The Sequential Organization of Vision", Vancouver BC, Dec 2005
CIFAR workshop on "Energy-Based Learning Algorithms", Vancouver BC, Dec 2005
CIFAR workshop on "Energy-Based Learning Algorithms", Alton, Ontario, April 2006
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2006
CIFAR workshop on "Cortical learning mechanisms", Vancouver BC, Dec 2006
CIFAR workshop on "The perception of wholes and parts", Vancouver BC, Dec 2006
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2007
CIFAR workshop on "Modeling time-series", Alton, Ontario, June 2007
CIFAR workshop on "Achieving Perceptual Invariance", Vancouver BC, Dec 2007
CIAR workshop on "Inhibitory Interneurons", Toronto, ON, Apr 2008
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2008
CIFAR workshop on "Learning vision", Vancouver BC, Dec 2008
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2009
CIFAR workshop on "Visual attention and top-down connections", Vancouver BC, Dec 2009
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2010
CIFAR workshop on "Perceiving images motions and sounds", Vancouver BC, Dec 2010
CIFAR workshop on "Using GPUs for Machine Learning", Toronto, May 2011
CIFAR-Gatsby workshop on "Hierarchical Distributed Representations", Windsor, UK, July 2011
Five-day Summer School on "Neural Computation and Adaptive Perception" Toronto, August 2011
CIFAR workshop on "Advances in Deep Learning", Grenada, Spain, Dec 2011
CIFAR joint workshop on "Neural Development", Alton, Ontario, April 2012

Tutorials

I have taught the following tutorials on neural networks:

1987: American Association for Artificial Intelligence (4 hours)
1988: AT&T Bell Labs (2 days)
Technology Transfer Institute: Los Angeles (3 days)
Technology Transfer Institute: Washington (3 days)
Technology Transfer Institute: San Francisco (3 days)
American Association for Artificial Intelligence (4 hours)
Information Technology Research Center of Ontario (2 days)

1989: Apple, Cupertino CA (1 day)
Information Technology Research Center of Ontario (2 days)
1990: International Joint Conference on Neural Networks (1 hour)
Digital Equipment Corporation (2 days)
American Association for Artificial Intelligence (4 hours)
Information Technology Research Center of Ontario (2 days)
1991: American Association for Artificial Intelligence (4 hours)
Information Technology Research Center of Ontario (2 days)
1993: Sydney University (2 days)
Information Technology Research Center of Ontario (2 days)
1994: Canadian Government, Ottawa (2 days)
Information Technology Research Center of Ontario (2 days)
1995: Neural Information Processing Systems Conference (2 hours)
Information Technology Research Center of Ontario (2 days)
1996: Information Technology Research Center of Ontario (2 days)
Tutorial on Neural Networks and Machine Learning (with Mike Jordan), Boston (2 days)
1997: Information Technology Research Center of Ontario (2 days)
Tutorial on Neural Networks and Machine Learning (with Mike Jordan), Washington D.C. (2 days)
Tutorial on Neural Networks and Machine Learning (with Mike Jordan), Los Angeles (2 days)
1998: Canadian Government, Ottawa (1 day)
Information Technology Research Center of Ontario (1 day)
1999: Gatsby Computational Neuroscience Unit Tutorial (4 hours).
2007: Neural Information Processing Systems Conference (2 hours)
2009: Tutorial on "Deep Learning" at University College London (3 hours)
Tutorial on "Deep Learning" at Cambridge University Summer School (3 hours)
Tutorial on "Deep Learning" at NIPS speech workshop, Whistler BC (1.5 hours)

Contributions to Industry

I have consulted for Dragon Systems, DuPont, Apple, Synaptics, Nesbitt-Burns, Ontario Hydro Technologies, Morphometrics, Athene, Animotion, and Microsoft Redmond. I was on the Technical Advisory Board of the Microsoft Research Laboratory in Cambridge, UK from 2000 to 2006 and on the advisory board of Deepmind from 2012 to 2014.

I founded a start-up called DNNresearch in November 2012 with two of my graduate students. We sold its IP to Google in March 2013. Google "DNNresearch" for a large number of media stories.

The deep learning methods for recognizing phonemes in speech (2009) and objects in images (2012) that were developed in my group at the University of Toronto revolutionized speech recognition and object recognition.

Public Dissemination of my Research

In 2012 I taught an eight week Coursera Massive Open Online Course called *Neural Networks for Machine Learning.* It has had more than 100,000 registered students.

I have given three popular Google techtalks:

Brains, Sex and Machine Learning, 2012, (132,000 views)

Recent Developments in Deep Learning, 2009, (80,000 views)
The Next Generation of Neural Networks, 2007, (486,000 views)

Media coverage

My research on deep learning has been described in the New York Times (2012), the Globe and Mail (2012), Wired Magazine (2014), the Toronto Star (2015), CBC radio (2014 & 2015), the Guardian, UK (2015), and in many other media outlets. MIT Technology Review named Deep Learning as one of the top 10 innovations of 2013.

Graduate students

I have been the adviser for 22 completed MSc's and the following 37 completed PhD's:

Peter Brown (1987)
The Acoustic-Modeling Problem in Automatic Speech Recognition.

David Ackley (1987)
Stochastic Iterated Genetic Hillclimbing.

Mark Derthick (1988)
Mundane Reasoning by Parallel Constraint Satisfaction.

Richard Szeliski (1988)
Bayesian Modeling of Uncertainty in Low-Level Vision.

Kevin Lang (1989)
Phoneme Recognition Using Time-Delay Neural Nets.

Steven Nowlan (1991)
Soft Competitive Adaptation.

David Plaut (1991)
Connectionist Neuropsychology.

Conrad Galland (1991)
Learning in Deterministic Boltzmann Machine Networks.

S. Becker (1992)
An Information Theoretic Unsupervised Learning Algorithm for Neural Networks.

Richard Zemel (1994)
A Minimum Description Length Framework for Unsupervised Learning.

Tony Plate (1994)
Distributed Representations and Nested Compositional Structure.

Sidney Fels (1994)
Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks.

Christopher Williams (1994)
Combining Deformable Models and Neural Networks for Handprinted Digit Recognition.

Radford Neal (1994)
Bayesian Learning in Neural Networks

Carl Rasmussen (1996)
Evaluation of Gaussian Processes and Other Methods for Non-linear Regression.

Brendan Frey (1997)
Bayesian Networks for Pattern Classification, Data Compression and Channel Coding

Evan Steeg (1997)
Automated Motif Discovery in Protein Structure Prediction.

Radek Grzeszczuk (1998) (co-advised by Demitri Terzopoulos)
NeuroAnimator: Fast neural network emulation and control of physics-based models.

Brian Sallans (2002)
Reinforcement Learning for Factored Markov Decision Processes.

Sageev Oore (2002)
Digital Marionette: Augmenting Kinematics with Physics for Multi-Track Desktop Performance Animation.

Andrew Brown (2002)
Product Models for Sequences.

Alberto Paccanaro (2002)
Learning Distributed Representations of Relational Data using Linear Relational Embedding.

Yee-Whye Teh (2003)
Bethe Free Energy and Contrastive Divergence Approximations for Undirected Graphical Models.

Simon Osindero (2004)
Contrastive Topographic Models: Energy-based density models applied to the understanding of sensory coding and cortical topography.

Roland Memisevic (2007)
Non-linear Latent Factor Models for Revealing Structure in High-dimensional Data.

Ruslan Salakhutdinov (2009)
Learning deep generative models.

Graham Taylor (2009)
Composable, distributed-state models for high-dimensional time-series.

Andriy Mnih (2009)
Learning distributed representations for language modeling and collaborative filtering.

Vinod Nair (2010)
Visual object recognition using generative models of images.

Josh Susskind (2011)
Interpreting faces with neurally inspired generative models.

Ilya Sutskever (2012)
Training Recurrent Neural Networks.

Abdel-rahman Mohamed (2013)
Deep Neural Network Acoustic Models for ASR.

Vlad Mnih (2013)
Machine learning for aerial image labeling.

Navdeep Jaitly (2014)
Exploring Deep Learning Methods for Discovering Features in Speech Signals.

Tijmen Tieleman (2014)
Optimizing Neural Networks that Generate Images.

George Dahl (2015)
Deep Learning Approaches to Problems in Speech Recognition, Computational Chemistry and Natural Language Processing.

Charlie) Yichuan Tang (2015)
Learning Generative Models using Structured Latent Variables.