
Bibliography

- ABRAHAMSEN, P. (1997) A review of Gaussian random fields and correlation functions. Technical Report 917, Norwegian Computing Center, Blindern, N-0314 Oslo, Norway. 2nd edition.
- ABRAMSON, N. (1963) *Information Theory and Coding*. McGraw-Hill.
- ADLER, S. L. (1981) Over-relaxation method for the Monte-Carlo evaluation of the partition function for multiquadratic actions. *Physical Review D – Particles and Fields* **23** (12): 2901–2904.
- AIYER, S. V. B. (1991) *Solving Combinatorial Optimization Problems Using Neural Networks*. Cambridge Univ. Engineering Dept. PhD dissertation. CUED/F-INFENG/TR 89.
- AJI, S., JIN, H., KHANDEKAR, A., MCELIECE, R. J., and MACKAY, D. J. C. (2000) BSC thresholds for code ensembles based on ‘typical pairs’ decoding. In *Codes, Systems and Graphical Models*, ed. by B. Marcus and J. Rosenthal, volume 123 of *IMA Volumes in Mathematics and its Applications*, pp. 195–210. Springer.
- AMARI, S., CICHOCKI, A., and YANG, H. H. (1996) A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, ed. by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, volume 8, pp. 757–763. MIT Press.
- AMIT, D. J., GUTFREUND, H., and SOMPOLINSKY, H. (1985) Storing infinite numbers of patterns in a spin glass model of neural networks. *Phys. Rev. Lett.* **55**: 1530–1533.
- ANGEL, J. R. P., WIZINOWICH, P., LLOYD-HART, M., and SANDLER, D. (1990) Adaptive optics for array telescopes using neural-network techniques. *Nature* **348**: 221–224.
- BAHL, L. R., COCKE, J., JELINEK, F., and RAVIV, J. (1974) Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Info. Theory* **IT-20**: 284–287.
- BALDWIN, J. (1896) A new factor in evolution. *American Naturalist* **30**: 441–451.
- BAR-SHALOM, Y., and FORTMANN, T. (1988) *Tracking and Data Association*. Academic Press.
- BARBER, D., and WILLIAMS, C. K. I. (1997) Gaussian processes for Bayesian classification via hybrid Monte Carlo. In *Neural Information Processing Systems 9*, ed. by M. C. Mozer, M. I. Jordan, and T. Petsche, pp. 340–346. MIT Press.
- BARNETT, S. (1979) *Matrix Methods for Engineers and Scientists*. McGraw-Hill.
- BATTAIL, G. (1993) We can think of good codes, and even decode them. In *Eurocode '92. Udine, Italy, 26-30 October*, ed. by P. Camion, P. Charpin, and S. Harari, number 339 in CISM Courses and Lectures, pp. 353–368. Springer.
- BAUM, E., BONEH, D., and GARRETT, C. (1995) On genetic algorithms. In *Proc. Eighth Annual Conf. on Computational Learning Theory*, pp. 230–239. ACM.
- BAUM, E. B., and SMITH, W. D. (1993) Best play for imperfect players and game tree search. Technical report, NEC, Princeton, NJ.
- BAUM, E. B., and SMITH, W. D. (1997) A Bayesian approach to relevance in game playing. *Artificial Intelligence* **97** (1-2): 195–242.
- BAUM, L. E., and PETRIE, T. (1966) Statistical inference for probabilistic functions of finite-state Markov chains. *Ann. Math. Stat.* **37**: 1559–1563.
- BEAL, M. J., GHAHRAMANI, Z., and RASMUSSEN, C. E. (2002) The infinite hidden Markov model. In *Advances in Neural Information Processing Systems 14*. MIT Press.
- BELL, A. J., and SEJNOWSKI, T. J. (1995) An information maximization approach to blind separation and blind deconvolution. *Neural Computation* **7** (6): 1129–1159.
- BENTLEY, J. (2000) *Programming Pearls*. Addison-Wesley, second edition.
- BERGER, J. (1985) *Statistical Decision theory and Bayesian Analysis*. Springer.
- BERLEKAMP, E. R. (1968) *Algebraic Coding Theory*. McGraw-Hill.
- BERLEKAMP, E. R. (1980) The technology of error-correcting codes. *IEEE Trans. Info. Theory* **68**: 564–593.
- BERLEKAMP, E. R., MCELIECE, R. J., and VAN TILBORG, H. C. A. (1978) On the intractability of certain coding problems. *IEEE Trans. Info. Theory* **24** (3): 384–386.
- BERROU, C., and GLAVIEUX, A. (1996) Near optimum error correcting coding and decoding: Turbo-codes. *IEEE Trans. on Communications* **44**: 1261–1271.
- BERROU, C., GLAVIEUX, A., and THITIMAJSHIMA, P. (1993) Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *Proc. 1993 IEEE International Conf. on Communications, Geneva, Switzerland*, pp. 1064–1070.
- BERZUINI, C., BEST, N. G., GILKS, W. R., and LARIZZA, C. (1997) Dynamic conditional independence models and Markov chain Monte Carlo methods. *J. American Statistical Assoc.* **92** (440): 1403–1412.
- BERZUINI, C., and GILKS, W. R. (2001) Following a moving target – Monte Carlo inference for dynamic Bayesian models. *J. Royal Statistical Society Series B – Statistical Methodology* **63** (1): 127–146.
- BHATTACHARYYA, A. (1943) On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **35**: 99–110.
- BISHOP, C. M. (1992) Exact calculation of the Hessian matrix for the multilayer perceptron. *Neural Computation* **4** (4): 494–501.
- BISHOP, C. M. (1995) *Neural Networks for Pattern Recognition*. Oxford Univ. Press.
- BISHOP, C. M., and WINN, J. M. (2000) Non-linear Bayesian image modelling. In *Proc. Sixth European Conf. on Computer Vision*, volume 1, pp. 3–17. Springer.
- BISHOP, C. M., and WINN, J. M. (2003) Structured variational distributions in VIBES. In *Advances in Neural Information Processing Systems*.
- BISHOP, C. M., WINN, J. M., and SPIEGELHALTER, D. (2002) VIBES: A variational inference engine for Bayesian networks. In *Advances in Neural Information Processing Systems XV*, ed. by S. Becker, S. Thrun, and K. Obermayer.
- BLAHUT, R. E. (1987) *Principles and Practice of Information Theory*. Addison-Wesley.

- BOTTOU, L., HOWARD, P. G., and BENGIO, Y. (1998) The Z-coder adaptive binary coder. In *Proc. Data Compression Conf., Snowbird, Utah, March 1998*, pp. 13–22.
- BOX, G. E. P., and TIAO, G. C. (1973) *Bayesian Inference in Statistical Analysis*. Addison–Wesley.
- BRAUNSTEIN, A., MÉZARD, M., and ZECCHINA, R., (2003) Survey propagation: an algorithm for satisfiability. cs.CC/0212002.
- BRETTTHORST, G. (1988) *Bayesian Spectrum Analysis and Parameter Estimation*. Springer. Also available at bayes.wustl.edu.
- BRIDLE, J. S. (1989) Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neuro-computing: Algorithms, Architectures and Applications*, ed. by F. Fogelman-Soulie and J. Héroult. Springer–Verlag.
- BULMER, M. (1985) *The Mathematical Theory of Quantitative Genetics*. Oxford Univ. Press.
- BURROWS, M., and WHEELER, D. J. (1994) A block-sorting lossless data compression algorithm. Technical Report 124, Digital SRC.
- BYERS, J., LUBY, M., MITZENMACHER, M., and REGE, A. (1998) A digital fountain approach to reliable distribution of bulk data. In *Proc. ACM SIGCOMM '98, September 2–4, 1998*.
- CAIRNS-SMITH, A. G. (1985) *Seven Clues to the Origin of Life*. Cambridge Univ. Press.
- CALDERBANK, A. R., and SHOR, P. W. (1996) Good quantum error-correcting codes exist. *Phys. Rev. A* **54**: 1098. quant-ph/9512032.
- CARROLL, L. (1998) *Alice's Adventures in Wonderland; and, Through the Looking-glass: and what Alice Found There*. Macmillan Children's Books.
- CHILDS, A. M., PATTERSON, R. B., and MACKAY, D. J. C. (2001) Exact sampling from non-attractive distributions using summary states. *Physical Review E* **63**: 036113.
- CHU, W., KEERTHI, S. S., and ONG, C. J. (2001) A unified loss function in Bayesian framework for support vector regression. In *Proc. 18th International Conf. on Machine Learning*, pp. 51–58.
- CHU, W., KEERTHI, S. S., and ONG, C. J. (2002) A new Bayesian design method for support vector classification. In *Special Section on Support Vector Machines of the 9th International Conf. on Neural Information Processing*.
- CHU, W., KEERTHI, S. S., and ONG, C. J. (2003a) Bayesian support vector regression using a unified loss function. *IEEE Trans. on Neural Networks*. Submitted.
- CHU, W., KEERTHI, S. S., and ONG, C. J. (2003b) Bayesian trigonometric support vector classifier. *Neural Computation*.
- CHUNG, S.-Y., RICHARDSON, T. J., and URBANKE, R. L. (2001) Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation. *IEEE Trans. Info. Theory* **47** (2): 657–670.
- CHUNG, S.-Y., URBANKE, R. L., and RICHARDSON, T. J., (1999) LDPC code design applet. lids.mit.edu/~sychung/gaopt.html.
- COMON, P., JUTTEN, C., and HERAULT, J. (1991) Blind separation of sources. 2. Problems statement. *Signal Processing* **24** (1): 11–20.
- COPAS, J. B. (1983) Regression, prediction and shrinkage (with discussion). *J. R. Statist. Soc. B* **45** (3): 311–354.
- COVER, T. M. (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. on Electronic Computers* **14**: 326–334.
- COVER, T. M., and THOMAS, J. A. (1991) *Elements of Information Theory*. Wiley.
- COWLES, M. K., and CARLIN, B. P. (1996) Markov-chain Monte-Carlo convergence diagnostics – a comparative review. *J. American Statistical Assoc.* **91** (434): 883–904.
- COX, R. (1946) Probability, frequency, and reasonable expectation. *Am. J. Physics* **14**: 1–13.
- CRESSIE, N. (1993) *Statistics for Spatial Data*. Wiley.
- DAVEY, M. C. (1999) *Error-correction using Low-Density Parity-Check Codes*. Univ. of Cambridge PhD dissertation.
- DAVEY, M. C., and MACKAY, D. J. C. (1998) Low density parity check codes over GF(q). *IEEE Communications Letters* **2** (6): 165–167.
- DAVEY, M. C., and MACKAY, D. J. C. (2000) Watermark codes: Reliable communication over insertion/deletion channels. In *Proc. 2000 IEEE International Symposium on Info. Theory*, p. 477.
- DAVEY, M. C., and MACKAY, D. J. C. (2001) Reliable communication over channels with insertions, deletions and substitutions. *IEEE Trans. Info. Theory* **47** (2): 687–698.
- DAWID, A., STONE, M., and ZIDEK, J. (1996) Critique of E.T Jaynes's 'paradoxes of probability theory'. Technical Report 172, Dept. of Statistical Science, Univ. College London.
- DAYAN, P., HINTON, G. E., NEAL, R. M., and ZEMEL, R. S. (1995) The Helmholtz machine. *Neural Computation* **7** (5): 889–904.
- DIVSALAR, D., JIN, H., and McELIECE, R. J. (1998) Coding theorems for 'turbo-like' codes. In *Proc. 36th Allerton Conf. on Communication, Control, and Computing, Sept. 1998*, pp. 201–210. Allerton House.
- DOUCET, A., DE FREITAS, J., and GORDON, N. eds. (2001) *Sequential Monte Carlo Methods in Practice*. Springer.
- DURBIN, R., EDDY, S. R., KROGH, A., and MITCHISON, G. (1998) *Biological Sequence Analysis. Probabilistic Models of Proteins and Nucleic Acids*. Cambridge Univ. Press.
- DYSON, F. J. (1985) *Origins of Life*. Cambridge Univ. Press.
- ELIAS, P. (1975) Universal codeword sets and representations of the integers. *IEEE Trans. Info. Theory* **21** (2): 194–203.
- EYRE-WALKER, A., and KEIGHTLEY, P. (1999) High genomic deleterious mutation rates in hominids. *Nature* **397**: 344–347.
- FELSENSTEIN, J. (1985) Recombination and sex: is Maynard Smith necessary? In *Evolution. Essays in Honour of John Maynard Smith*, ed. by P. J. Greenwood, P. H. Harvey, and M. Slatkin, pp. 209–220. Cambridge Univ. Press.
- FERREIRA, H., CLARKE, W., HELBERG, A., ABDEL-GHAFFAR, K. S., and VINCK, A. H. (1997) Insertion/deletion correction with spectral nulls. *IEEE Trans. Info. Theory* **43** (2): 722–732.
- FEYNMAN, R. P. (1972) *Statistical Mechanics*. Addison–Wesley.
- FORNEY, JR., G. D. (1966) *Concatenated Codes*. MIT Press.
- FORNEY, JR., G. D. (2001) Codes on graphs: Normal realizations. *IEEE Trans. Info. Theory* **47** (2): 520–548.
- FREY, B. J. (1998) *Graphical Models for Machine Learning and Digital Communication*. MIT Press.
- GALLAGER, R. G. (1962) Low density parity check codes. *IRE Trans. Info. Theory* **IT-8**: 21–28.
- GALLAGER, R. G. (1963) *Low Density Parity Check Codes*. Number 21 in MIT Research monograph series. MIT Press. Available from www.inference.phy.cam.ac.uk/mackay/gallager/papers/.
- GALLAGER, R. G. (1968) *Information Theory and Reliable Communication*. Wiley.
- GALLAGER, R. G. (1978) Variations on a theme by Huffman. *IEEE Trans. Info. Theory* **IT-24** (6): 668–674.
- GIBBS, M. N. (1997) *Bayesian Gaussian Processes for Regression and Classification*. Cambridge Univ. PhD dissertation. www.inference.phy.cam.ac.uk/mng10/.
- GIBBS, M. N., and MACKAY, D. J. C., (1996) Efficient implementation of Gaussian processes for interpolation. www.inference.phy.cam.ac.uk/mackay/abstracts/gpros.html.

- GIBBS, M. N., and MACKEY, D. J. C. (2000) Variational Gaussian process classifiers. *IEEE Trans. on Neural Networks* **11** (6): 1458–1464.
- GILKS, W., ROBERTS, G., and GEORGE, E. (1994) Adaptive direction sampling. *Statistician* **43**: 179–189.
- GILKS, W., and WILD, P. (1992) Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**: 337–348.
- GILKS, W. R., RICHARDSON, S., and SPIEGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- GOLDIE, C. M., and PINCH, R. G. E. (1991) *Communication theory*. Cambridge Univ. Press.
- GOLOMB, S. W., PEILE, R. E., and SCHOLTZ, R. A. (1994) *Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111*. Plenum Press.
- GOOD, I. J. (1979) Studies in the history of probability and statistics. XXXVII. A.M. Turing's statistical work in World War II. *Biometrika* **66** (2): 393–396.
- GRAHAM, R. L. (1966) On partitions of a finite set. *Journal of Combinatorial Theory* **1**: 215–223.
- GRAHAM, R. L., and KNOWLTON, K. C., (1968) Method of identifying conductors in a cable by establishing conductor connection groupings at both ends of the cable. U.S. Patent 3,369,177.
- GREEN, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.
- GREGORY, P. C., and LOREDO, T. J. (1992) A new method for the detection of a periodic signal of unknown shape and period. In *Maximum Entropy and Bayesian Methods*, ed. by G. Erickson and C. Smith. Kluwer. Also in *Astrophysical Journal*, **398**, pp. 146–168, Oct 10, 1992.
- GULL, S. F. (1988) Bayesian inductive inference and maximum entropy. In *Maximum Entropy and Bayesian Methods in Science and Engineering, vol. 1: Foundations*, ed. by G. Erickson and C. Smith, pp. 53–74. Kluwer.
- GULL, S. F. (1989) Developments in maximum entropy data analysis. In *Maximum Entropy and Bayesian Methods, Cambridge 1988*, ed. by J. Skilling, pp. 53–71. Kluwer.
- GULL, S. F., and DANIELL, G. (1978) Image reconstruction from incomplete and noisy data. *Nature* **272**: 686–690.
- HANSON, R., STUTZ, J., and CHEESEMAN, P. (1991a) Bayesian classification theory. Technical Report FIA-90-12-7-01, NASA Ames.
- HANSON, R., STUTZ, J., and CHEESEMAN, P. (1991b) Bayesian classification with correlation and inheritance. In *Proc. 12th Intern. Joint Conf. on Artificial Intelligence, Sydney, Australia*, volume 2, pp. 692–698. Morgan Kaufmann.
- HARTMANN, C. R. P., and RUDOLPH, L. D. (1976) An optimum symbol by symbol decoding rule for linear codes. *IEEE Trans. Info. Theory* **IT-22**: 514–517.
- HARVEY, M., and NEAL, R. M. (2000) Inference for belief networks using coupling from the past. In *Uncertainty in Artificial Intelligence: Proc. Sixteenth Conf.*, pp. 256–263.
- HEBB, D. O. (1949) *The Organization of Behavior*. Wiley.
- HENDIN, O., HORN, D., and HOPFIELD, J. J. (1994) Decomposition of a mixture of signals in a model of the olfactory bulb. *Proc. Natl. Acad. Sci. USA* **91** (13): 5942–5946.
- HERTZ, J., KROGH, A., and PALMER, R. G. (1991) *Introduction to the Theory of Neural Computation*. Addison-Wesley.
- HINTON, G. (2001) Training products of experts by minimizing contrastive divergence. Technical Report 2000-004, Gatsby Computational Neuroscience Unit, Univ. College London.
- HINTON, G., and NOWLAN, S. (1987) How learning can guide evolution. *Complex Systems* **1**: 495–502.
- HINTON, G. E., DAYAN, P., FREY, B. J., and NEAL, R. M. (1995) The wake-sleep algorithm for unsupervised neural networks. *Science* **268** (5214): 1158–1161.
- HINTON, G. E., and GHAHRAMANI, Z. (1997) Generative models for discovering sparse distributed representations. *Philosophical Trans. Royal Society B*.
- HINTON, G. E., and SEJNOWSKI, T. J. (1986) Learning and relearning in Boltzmann machines. In *Parallel Distributed Processing*, ed. by D. E. Rumelhart and J. E. McClelland, pp. 282–317. MIT Press.
- HINTON, G. E., and TEH, Y. W. (2001) Discovering multiple constraints that are frequently approximately satisfied. In *Uncertainty in Artificial Intelligence: Proc. Seventeenth Conf. (UAI-2001)*, pp. 227–234. Morgan Kaufmann.
- HINTON, G. E., and VAN CAMP, D. (1993) Keeping neural networks simple by minimizing the description length of the weights. In *Proc. 6th Annual Workshop on Comput. Learning Theory*, pp. 5–13. ACM Press, New York, NY.
- HINTON, G. E., WELLING, M., TEH, Y. W., and OSINDERO, S. (2001) A new view of ICA. In *Proc. International Conf. on Independent Component Analysis and Blind Signal Separation*, volume 3.
- HINTON, G. E., and ZEMEL, R. S. (1994) Autoencoders, minimum description length and Helmholtz free energy. In *Advances in Neural Information Processing Systems 6*, ed. by J. D. Cowan, G. Tesauro, and J. Alspector. Morgan Kaufmann.
- HODGES, A. (1983) *Alan Turing: The Enigma*. Simon and Schuster.
- HOJEN-SORENSEN, P. A., WINTHER, O., and HANSEN, L. K. (2002) Mean field approaches to independent component analysis. *Neural Computation* **14**: 889–918.
- HOLMES, C., and MALLICK, B. (1998) Perfect simulation for orthogonal model mixing. Technical report, Imperial College, London.
- HOPFIELD, J. J. (1974) Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA* **71** (10): 4135–4139.
- HOPFIELD, J. J. (1978) Origin of the genetic code: A testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proc. Natl. Acad. Sci. USA* **75** (9): 4334–4338.
- HOPFIELD, J. J. (1980) The energy relay: A proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein synthesis. *Proc. Natl. Acad. Sci. USA* **77** (9): 5248–5252.
- HOPFIELD, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79**: 2554–8.
- HOPFIELD, J. J. (1984) Neurons with graded response properties have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **81**: 3088–92.
- HOPFIELD, J. J. (1987) Learning algorithms and probability distributions in feed-forward and feed-back networks. *Proc. Natl. Acad. Sci. USA* **84**: 8429–33.
- HOPFIELD, J. J., and BRODY, C. D. (2000) What is a moment? “Cortical” sensory integration over a brief interval. *Proc. Natl. Acad. Sci.* **97**: 13919–13924.
- HOPFIELD, J. J., and BRODY, C. D. (2001) What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl. Acad. Sci.* **98**: 1282–1287.
- HOPFIELD, J. J., and TANK, D. W. (1985) Neural computation of decisions in optimization problems. *Biol. Cybernetics* **52**: 1–25.
- HOWARTH, P., and BRADLEY, A. (1986) The longitudinal aberration of the human eye and its correction. *Vision Res.* **26**: 361–366.
- HUBER, M. (1998) Exact sampling and approximate counting techniques. In *Proc. 30th ACM Symposium on the Theory of Computing*, pp. 31–40.

- ICHIKAWA, K., BHADESHIA, H. K. D. H., and MACKAY, D. J. C. (1996) Model for hot cracking in low-alloy steel weld metals. *Science and Technology of Welding and Joining* **1**: 43–50.
- ISARD, M., and BLAKE, A. (1996) Visual tracking by stochastic propagation of conditional density. In *Proc. Fourth European Conf. Computer Vision*, pp. 343–356.
- ISARD, M., and BLAKE, A. (1998) Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision* **29** (1): 5–28.
- JAAKKOLA, T. S., and JORDAN, M. I. (1996) Computing upper and lower bounds on likelihoods in intractable networks. In *Proc. Twelfth Conf. on Uncertainty in AI*. Morgan Kaufman.
- JAAKKOLA, T. S., and JORDAN, M. I. (2000a) Bayesian logistic regression: a variational approach. *Statistics and Computing* **10**: 25–37.
- JAAKKOLA, T. S., and JORDAN, M. I. (2000b) Bayesian parameter estimation via variational methods. *Statistics and Computing* **10** (1): 25–37.
- JAYNES, E. T. (1983) Bayesian intervals versus confidence intervals. In *E.T. Jaynes. Papers on Probability, Statistics and Statistical Physics*, ed. by R. D. Rosenkrantz, p. 151. Kluwer.
- JAYNES, E. T. (2003) *Probability Theory: The Logic of Science*. Cambridge Univ. Press. Edited by G. Larry Bretthorst.
- JENSEN, F. V. (1996) *An Introduction to Bayesian Networks*. UCL press.
- JOHANNESSON, R., and ZIGANGIROV, K. S. (1999) *Fundamentals of Convolutional Coding*. IEEE Press.
- JORDAN, M. I. ed. (1998) *Learning in Graphical Models*. NATO Science Series. Kluwer Academic Publishers.
- JPL, (1996) Turbo codes performance. Available from www331.jpl.nasa.gov/public/TurboPerf.html.
- JUTTEN, C., and HERAULT, J. (1991) Blind separation of sources. 1. An adaptive algorithm based on neuromimetic architecture. *Signal Processing* **24** (1): 1–10.
- KARPLUS, K., and KRIT, H. (1991) A semi-systolic decoder for the PDSC–73 error-correcting code. *Discrete Applied Mathematics* **33**: 109–128.
- KEPLER, T., and OPREA, M. (2001) Improved inference of mutation rates: I. An integral representation of the Luria-Delbrück distribution. *Theoretical Population Biology* **59**: 41–48.
- KIMELDORF, G. S., and WAHBA, G. (1970) A correspondence between Bayesian estimation of stochastic processes and smoothing by splines. *Annals of Math. Statistics* **41** (2): 495–502.
- KITANIDIS, P. K. (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research* **22**: 499–507.
- KNUTH, D. E. (1968) *The Art of Computer Programming. Volume 1: Fundamental Algorithms*. Addison Wesley.
- KONDRASHOV, A. S. (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature* **336** (6198): 435–440.
- KSCHISCHANG, F. R., FREY, B. J., and LOELIGER, H.-A. (2001) Factor graphs and the sum-product algorithm. *IEEE Trans. Info. Theory* **47** (2): 498–519.
- KSCHISCHANG, F. R., and SOROKINE, V. (1995) On the trellis structure of block codes. *IEEE Trans. Info. Theory* **41** (6): 1924–1937.
- LAURITZEN, S. L. (1981) Time series analysis in 1880, a discussion of contributions made by T. N. Thiele. *ISI Review* **49**: 319–333.
- LAURITZEN, S. L. (1996) *Graphical Models*. Number 17 in Oxford Statistical Science Series. Clarendon Press.
- LAURITZEN, S. L., and SPIEGELHALTER, D. J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society B* **50**: 157–224.
- LEVENSHTEIN, V. I. (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics – Doklady* **10** (8): 707–710.
- LIN, S., and COSTELLO, JR., D. J. (1983) *Error Control Coding: Fundamentals and Applications*. Prentice-Hall.
- LITSYN, S., and SHEVELEV, V. (2002) On ensembles of low-density parity-check codes: asymptotic distance distributions. *IEEE Trans. Info. Theory* **48** (4): 887–908.
- LOREDO, T. J. (1990) From Laplace to supernova SN 1987A: Bayesian inference in astrophysics. In *Maximum Entropy and Bayesian Methods, Dartmouth, U.S.A., 1989*, ed. by P. Fougere, pp. 81–142. Kluwer.
- LOWE, D. G. (1995) Similarity metric learning for a variable kernel classifier. *Neural Computation* **7**: 72–85.
- LUBY, M. (2002) LT codes. In *Proc. The 43rd Annual IEEE Symposium on Foundations of Computer Science, November 16–19 2002*, pp. 271–282.
- LUBY, M. G., MITZENMACHER, M., SHOKROLLAHI, M. A., and SPIELMAN, D. A. (1998) Improved low-density parity-check codes using irregular graphs and belief propagation. In *Proc. IEEE International Symposium on Info. Theory*, p. 117.
- LUBY, M. G., MITZENMACHER, M., SHOKROLLAHI, M. A., and SPIELMAN, D. A. (2001a) Efficient erasure correcting codes. *IEEE Trans. Info. Theory* **47** (2): 569–584.
- LUBY, M. G., MITZENMACHER, M., SHOKROLLAHI, M. A., and SPIELMAN, D. A. (2001b) Improved low-density parity-check codes using irregular graphs and belief propagation. *IEEE Trans. Info. Theory* **47** (2): 585–584.
- LUBY, M. G., MITZENMACHER, M., SHOKROLLAHI, M. A., SPIELMAN, D. A., and STEMANN, V. (1997) Practical loss-resilient codes. In *Proc. Twenty-Ninth Annual ACM Symposium on Theory of Computing (STOC)*.
- LUO, Z., and WAHBA, G. (1997) Hybrid adaptive splines. *J. Amer. Statist. Assoc.* **92**: 107–116.
- LURIA, S. E., and DELBRÜCK, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511. Reprinted in *Microbiology: A Centenary Perspective*, Wolfgang K. Joklik, ed., 1999, ASM Press, and available from www.esp.org/.
- LUTTRELL, S. P. (1989) Hierarchical vector quantisation. *Proc. IEE Part I* **136**: 405–413.
- LUTTRELL, S. P. (1990) Derivation of a class of training algorithms. *IEEE Trans. on Neural Networks* **1** (2): 229–232.
- MACKAY, D. J. C. (1991) *Bayesian Methods for Adaptive Models*. California Institute of Technology PhD dissertation.
- MACKAY, D. J. C. (1992a) Bayesian interpolation. *Neural Computation* **4** (3): 415–447.
- MACKAY, D. J. C. (1992b) The evidence framework applied to classification networks. *Neural Computation* **4** (5): 698–714.
- MACKAY, D. J. C. (1992c) A practical Bayesian framework for backpropagation networks. *Neural Computation* **4** (3): 448–472.
- MACKAY, D. J. C. (1994a) Bayesian methods for backpropagation networks. In *Models of Neural Networks III*, ed. by E. Domany, J. L. van Hemmen, and K. Schulten, chapter 6, pp. 211–254. Springer.
- MACKAY, D. J. C. (1994b) Bayesian non-linear modelling for the prediction competition. In *ASHRAE Trans., V.100, Pt.2*, pp. 1053–1062. American Society of Heating, Refrigeration, and Air-conditioning Engineers.
- MACKAY, D. J. C. (1995a) Free energy minimization algorithm for decoding and cryptanalysis. *Electronics Letters* **31** (6): 446–447.
- MACKAY, D. J. C. (1995b) Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems* **6**: 469–505.
- MACKAY, D. J. C., (1997a) Ensemble learning for hidden Markov models. www.inference.phy.cam.ac.uk/mackay/abstracts/ensemblePaper.html.

- MACKEY, D. J. C., (1997b) Iterative probabilistic decoding of low density parity check codes. Animations available on world wide web. www.inference.phy.cam.ac.uk/mackay/codes/gifs/.
- MACKEY, D. J. C. (1998a) Choice of basis for Laplace approximation. *Machine Learning* **33** (1): 77–86.
- MACKEY, D. J. C. (1998b) Introduction to Gaussian processes. In *Neural Networks and Machine Learning*, ed. by C. M. Bishop, NATO ASI Series, pp. 133–166. Kluwer.
- MACKEY, D. J. C. (1999a) Comparison of approximate methods for handling hyperparameters. *Neural Computation* **11** (5): 1035–1068.
- MACKEY, D. J. C. (1999b) Good error correcting codes based on very sparse matrices. *IEEE Trans. Info. Theory* **45** (2): 399–431.
- MACKEY, D. J. C., (2000) An alternative to runlength-limiting codes: Turn timing errors into substitution errors. Available from www.inference.phy.cam.ac.uk/mackay/.
- MACKEY, D. J. C., (2001) A problem with variational free energy minimization. www.inference.phy.cam.ac.uk/mackay/abstracts/minima.html.
- MACKEY, D. J. C., and DAVEY, M. C. (2000) Evaluation of Gallager codes for short block length and high rate applications. In *Codes, Systems and Graphical Models*, ed. by B. Marcus and J. Rosenthal, volume 123 of *IMA Volumes in Mathematics and its Applications*, pp. 113–130. Springer.
- MACKEY, D. J. C., MITCHISON, G., and MCFADDEN, P. L., (2003) Sparse-graph codes for quantum error-correction. [quant-ph/0304161](http://arxiv.org/abs/quant-ph/0304161). Submitted to *IEEE Trans. Info. Theory* May 8, 2003.
- MACKEY, D. J. C., and NEAL, R. M. (1995) Good codes based on very sparse matrices. In *Cryptography and Coding. 5th IMA Conf., LNCS 1025*, ed. by C. Boyd, pp. 100–111. Springer.
- MACKEY, D. J. C., and NEAL, R. M. (1996) Near Shannon limit performance of low density parity check codes. *Electronics Letters* **32** (18): 1645–1646. Reprinted *Electronics Letters*, **33**(6):457–458, March 1997.
- MACKEY, D. J. C., and PETO, L. (1995) A hierarchical Dirichlet language model. *Natural Language Engineering* **1** (3): 1–19.
- MACKEY, D. J. C., WILSON, S. T., and DAVEY, M. C. (1998) Comparison of constructions of irregular Gallager codes. In *Proc. 36th Allerton Conf. on Communication, Control, and Computing, Sept. 1998*, pp. 220–229. Allerton House.
- MACKEY, D. J. C., WILSON, S. T., and DAVEY, M. C. (1999) Comparison of constructions of irregular Gallager codes. *IEEE Trans. on Communications* **47** (10): 1449–1454.
- MACKEY, D. M., and MACKAY, V. (1974) The time course of the McCollough effect and its physiological implications. *J. Physiol.* **237**: 38–39.
- MACKEY, D. M., and MCCULLOCH, W. S. (1952) The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* **14**: 127–135.
- MACWILLIAMS, F. J., and SLOANE, N. J. A. (1977) *The Theory of Error-correcting Codes*. North-Holland.
- MANDELBROT, B. (1982) *The Fractal Geometry of Nature*. W.H. Freeman.
- MAO, Y., and BANIHASHEMI, A. (2000) Design of good LDPC codes using girth distribution. In *IEEE International Symposium on Info. Theory, Italy, June, 2000*.
- MAO, Y., and BANIHASHEMI, A. (2001) A heuristic search for good LDPC codes at short block lengths. In *IEEE International Conf. on Communications*.
- MARINARI, E., and PARISI, G. (1992) Simulated tempering – a new Monte-Carlo scheme. *Europhysics Letters* **19** (6): 451–458.
- MATHERON, G. (1963) Principles of geostatistics. *Economic Geology* **58**: 1246–1266.
- MAYNARD SMITH, J. (1968) ‘Haldane’s dilemma’ and the rate of evolution. *Nature* **219** (5159): 1114–1116.
- MAYNARD SMITH, J. (1978) *The Evolution of Sex*. Cambridge Univ. Press.
- MAYNARD SMITH, J. (1988) *Games, Sex and Evolution*. Harvester–Wheatsheaf.
- MAYNARD SMITH, J., and SZÁTHMARY, E. (1995) *The Major Transitions in Evolution*. Freeman.
- MAYNARD SMITH, J., and SZÁTHMARY, E. (1999) *The Origins of Life*. Oxford Univ. Press.
- MCCOLLOUGH, C. (1965) Color adaptation of edge-detectors in the human visual system. *Science* **149**: 1115–1116.
- MCELIECE, R. J. (2002) *The Theory of Information and Coding*. Cambridge Univ. Press, second edition.
- MCELIECE, R. J., MACKAY, D. J. C., and CHENG, J.-F. (1998) Turbo decoding as an instance of Pearl’s ‘belief propagation’ algorithm. *IEEE Journal on Selected Areas in Communications* **16** (2): 140–152.
- MCMILLAN, B. (1956) Two inequalities implied by unique decipherability. *IRE Trans. Inform. Theory* **2**: 115–116.
- MINKA, T. (2001) *A family of algorithms for approximate Bayesian inference*. MIT PhD dissertation.
- MISKIN, J. W. (2001) *Ensemble Learning for Independent Component Analysis*. Dept. of Physics, Univ. of Cambridge PhD dissertation.
- MISKIN, J. W., and MACKAY, D. J. C. (2000) Ensemble learning for blind image separation and deconvolution. In *Advances in Independent Component Analysis*, ed. by M. Girolami. Springer.
- MISKIN, J. W., and MACKAY, D. J. C. (2001) Ensemble learning for blind source separation. In *ICA: Principles and Practice*, ed. by S. Roberts and R. Everson. Cambridge Univ. Press.
- MOSTELLER, F., and WALLACE, D. L. (1984) *Applied Bayesian and Classical Inference. The case of The Federalist papers*. Springer.
- NEAL, R. M. (1991) Bayesian mixture modelling by Monte Carlo simulation. Technical Report CRG–TR–91–2, Computer Science, Univ. of Toronto.
- NEAL, R. M. (1993a) Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems 5*, ed. by C. L. Giles, S. J. Hanson, and J. D. Cowan, pp. 475–482. Morgan Kaufmann.
- NEAL, R. M. (1993b) Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG–TR–93–1, Dept. of Computer Science, Univ. of Toronto.
- NEAL, R. M. (1995) Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation. Technical Report 9508, Dept. of Statistics, Univ. of Toronto.
- NEAL, R. M. (1996) *Bayesian Learning for Neural Networks*. Springer.
- NEAL, R. M. (1997a) Markov chain Monte Carlo methods based on ‘slicing’ the density function. Technical Report 9722, Dept. of Statistics, Univ. of Toronto.
- NEAL, R. M. (1997b) Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report CRG–TR–97–2, Dept. of Computer Science, Univ. of Toronto.
- NEAL, R. M. (1998) Annealed importance sampling. Technical Report 9805, Dept. of Statistics, Univ. of Toronto.
- NEAL, R. M. (2001) Defining priors for distributions using Dirichlet diffusion trees. Technical Report 0104, Dept. of Statistics, Univ. of Toronto.
- NEAL, R. M. (2003) Slice sampling. *Annals of Statistics* **31** (3): 705–767.
- NEAL, R. M., and HINTON, G. E. (1998) A new view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, ed. by M. I. Jordan, NATO Science Series, pp. 355–368. Kluwer.

- NIELSEN, M., and CHUANG, I. (2000) *Quantum Computation and Quantum Information*. Cambridge Univ. Press.
- OFFER, E., and SOLJANIN, E. (2000) An algebraic description of iterative decoding schemes. In *Codes, Systems and Graphical Models*, ed. by B. Marcus and J. Rosenthal, volume 123 of *IMA Volumes in Mathematics and its Applications*, pp. 283–298. Springer.
- OFFER, E., and SOLJANIN, E. (2001) LDPC codes: a group algebra formulation. In *Proc. Internat. Workshop on Coding and Cryptography WCC 2001, 8-12 Jan. 2001, Paris*.
- O'HAGAN, A. (1978) On curve fitting and optimal design for regression. *J. Royal Statistical Society, B* **40**: 1–42.
- O'HAGAN, A. (1987) Monte Carlo is fundamentally unsound. *The Statistician* **36**: 247–249.
- O'HAGAN, A. (1994) *Bayesian Inference*, volume 2B of *Kendall's Advanced Theory of Statistics*. Edward Arnold.
- OMRE, H. (1987) Bayesian kriging – merging observations and qualified guesses in kriging. *Mathematical Geology* **19**: 25–39.
- OPPER, M., and WINTHER, O. (2000) Gaussian processes for classification: Mean-field algorithms. *Neural Computation* **12** (11): 2655–2684.
- PATRICK, J. D., and WALLACE, C. S. (1982) Stone circle geometries: an information theory approach. In *Archaeoastronomy in the Old World*, ed. by D. C. Heggie, pp. 231–264. Cambridge Univ. Press.
- PEARL, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- PEARL, J. (2000) *Causality*. Cambridge Univ. Press.
- PEARLMUTTER, B. A., and PARRA, L. C. (1996) A context-sensitive generalization of ICA. In *International Conf. on Neural Information Processing, Hong Kong*, pp. 151–157.
- PEARLMUTTER, B. A., and PARRA, L. C. (1997) Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In *Advances in Neural Information Processing Systems*, ed. by M. C. Mozer, M. I. Jordan, and T. Petsche, volume 9, p. 613. MIT Press.
- PINTO, R. L., and NEAL, R. M. (2001) Improving Markov chain Monte Carlo estimators by coupling to an approximating chain. Technical Report 0101, Dept. of Statistics, Univ. of Toronto.
- POGGIO, T., and GIROSI, F. (1989) A theory of networks for approximation and learning. Technical Report A.I. 1140, MIT.
- POGGIO, T., and GIROSI, F. (1990) Networks for approximation and learning. *Proc. IEEE* **78**: 1481–1497.
- POLYA, G. (1954) *Induction and Analogy in Mathematics*. Princeton Univ. Press.
- PROPP, J. G., and WILSON, D. B. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** (1-2): 223–252.
- RABINER, L. R., and JUANG, B. H. (1986) An introduction to hidden Markov models. *IEEE ASSP Magazine* pp. 4–16.
- RASMUSSEN, C. E. (1996) *Evaluation of Gaussian Processes and Other Methods for Non-Linear Regression*. Univ. of Toronto PhD dissertation.
- RASMUSSEN, C. E. (2000) The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems 12*, ed. by S.olla, T. Leen, and K.-R. Müller, pp. 554–560. MIT Press.
- RASMUSSEN, C. E., (2002) Reduced rank Gaussian process learning. Unpublished manuscript.
- RASMUSSEN, C. E., and GHAHRAMANI, Z. (2002) Infinite mixtures of Gaussian process experts. In *Advances in Neural Information Processing Systems 14*, ed. by T. G. Diettrich, S. Becker, and Z. Ghahramani. MIT Press.
- RASMUSSEN, C. E., and GHAHRAMANI, Z. (2003) Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems XV*, ed. by S. Becker, S. Thrun, and K. Obermayer.
- RATLIFF, F., and RIGGS, L. A. (1950) Involuntary motions of the eye during monocular fixation. *J. Exptl. Psychol.* **40**: 687–701.
- RATZER, E. A., and MACKAY, D. J. C. (2003) Sparse low-density parity-check codes for channels with cross-talk. In *Proc. 2003 IEEE Info. Theory Workshop, Paris*.
- REIF, F. (1965) *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill.
- RICHARDSON, T., SHOKROLLAHI, M. A., and URBANKE, R. (2001) Design of capacity-approaching irregular low-density parity check codes. *IEEE Trans. Info. Theory* **47** (2): 619–637.
- RICHARDSON, T., and URBANKE, R. (2001a) The capacity of low-density parity check codes under message-passing decoding. *IEEE Trans. Info. Theory* **47** (2): 599–618.
- RICHARDSON, T., and URBANKE, R. (2001b) Efficient encoding of low-density parity-check codes. *IEEE Trans. Info. Theory* **47** (2): 638–656.
- RIDLEY, M. (2000) *Mendel's Demon: gene justice and the complexity of life*. Phoenix.
- RIPLEY, B. D. (1991) *Statistical Inference for Spatial Processes*. Cambridge Univ. Press.
- RIPLEY, B. D. (1996) *Pattern Recognition and Neural Networks*. Cambridge Univ. Press.
- RUMELHART, D. E., HINTON, G. E., and WILLIAMS, R. J. (1986) Learning representations by back-propagating errors. *Nature* **323**: 533–536.
- RUSSELL, S., and WEFALD, E. (1991) *Do the Right Thing: Studies in Limited Rationality*. MIT Press.
- SCHNEIER, B. (1996) *Applied Cryptography*. Wiley.
- SCHOLKOPF, B., BURGES, C., and VAPNIK, V. (1995) Extracting support data for a given task. In *Proc. First International Conf. on Knowledge Discovery and Data Mining*, ed. by U. M. Fayyad and R. Uthurusamy. AAAI Press.
- SCHOLTZ, R. A. (1982) The origins of spread-spectrum communications. *IEEE Trans. on Communications* **30** (5): 822–854.
- SEEGER, M., WILLIAMS, C. K. I., and LAWRENCE, N. (2003) Fast forward selection to speed up sparse Gaussian process regression. In *Proc. Ninth International Workshop on Artificial Intelligence and Statistics*, ed. by C. Bishop and B. J. Frey. Society for Artificial Intelligence and Statistics.
- SEJNOWSKI, T. J. (1986) Higher order Boltzmann machines. In *Neural networks for computing*, ed. by J. Denker, pp. 398–403. American Institute of Physics.
- SEJNOWSKI, T. J., and ROSENBERG, C. R. (1987) Parallel networks that learn to pronounce English text. *Journal of Complex Systems* **1** (1): 145–168.
- SHANNON, C. E. (1948) A mathematical theory of communication. *Bell Sys. Tech. J.* **27**: 379–423, 623–656.
- SHANNON, C. E. (1993) The best detection of pulses. In *Collected Papers of Claude Shannon*, ed. by N. J. A. Sloane and A. D. Wyner, pp. 148–150. IEEE Press.
- SHANNON, C. E., and WEAVER, W. (1949) *The Mathematical Theory of Communication*. Univ. of Illinois Press.
- SHOKROLLAHI, A. (2003) Raptor codes. Technical report, Laboratoire d'algorithmique, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. Available from algo.epfl.ch/.
- SIPSER, M., and SPIELMAN, D. A. (1996) Expander codes. *IEEE Trans. Info. Theory* **42** (6.1): 1710–1722.
- SKILLING, J. (1989) Classic maximum entropy. In *Maximum Entropy and Bayesian Methods, Cambridge 1988*, ed. by J. Skilling. Kluwer.
- SKILLING, J. (1993) Bayesian numerical analysis. In *Physics and Probability*, ed. by W. T. Grandy, Jr. and P. Milonni. Cambridge Univ. Press.
- SKILLING, J., and MACKAY, D. J. C. (2003) Slice sampling – a binary implementation. *Annals of Statistics* **31** (3): 753–755. Discussion of *Slice Sampling* by Radford M. Neal.

- SMOLA, A. J., and BARTLETT, P. (2001) Sparse Greedy Gaussian Process Regression. In *Advances in Neural Information Processing Systems 13*, ed. by T. K. Leen, T. G. Diettrich, and V. Tresp, pp. 619–625. MIT Press.
- SPIEGEL, M. R. (1988) *Statistics*. Schaum's outline series. McGraw-Hill, 2nd edition.
- SPIELMAN, D. A. (1996) Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Info. Theory* **42** (6.1): 1723–1731.
- SUTTON, R. S., and BARTO, A. G. (1998) *Reinforcement Learning: An Introduction*. MIT Press.
- SWANSON, L. (1988) A new code for Galileo. In *Proc. 1988 IEEE International Symposium Info. Theory*, pp. 94–95.
- TANNER, M. A. (1996) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. Springer, 3rd edition.
- TANNER, R. M. (1981) A recursive approach to low complexity codes. *IEEE Trans. Info. Theory* **27** (5): 533–547.
- TEAHAN, W. J. (1995) Probability estimation for PPM. In *Proc. NZCSRSC'95*. Available from citeseer.nj.nec.com/teahan95probability.html.
- TEN BRINK, S. (1999) Convergence of iterative decoding. *Electronics Letters* **35** (10): 806–808.
- TEN BRINK, S., KRAMER, G., and ASHIKHMIN, A., (2002) Design of low-density parity-check codes for multi-antenna modulation and detection. Submitted to *IEEE Trans. on Communications*.
- TERRAS, A. (1999) *Fourier Analysis on Finite Groups and Applications*. Cambridge Univ. Press.
- THOMAS, A., SPIEGELHALTER, D. J., and GILKS, W. R. (1992) BUGS: A program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics 4*, ed. by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 837–842. Clarendon Press.
- TRESP, V. (2000) A Bayesian committee machine. *Neural Computation* **12** (11): 2719–2741.
- URBANKE, R., (2001) LdpcOpt – a fast and accurate degree distribution optimizer for LDPC code ensembles. 1thcwww.epfl.ch/research/ldpcopt/.
- VAPNIK, V. (1995) *The Nature of Statistical Learning Theory*. Springer.
- VITERBI, A. J. (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Info. Theory* **IT-13**: 260–269.
- WAHBA, G. (1990) *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics. CBMS-NSF Regional Conf. series in applied mathematics.
- WAINWRIGHT, M. J., JAAKKOLA, T., and WILLSKY, A. S. (2003) Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Trans. Info. Theory* **45** (9): 1120–1146.
- WALD, G., and GRIFFIN, D. (1947) The change in refractive power of the eye in bright and dim light. *J. Opt. Soc. Am.* **37**: 321–336.
- WALLACE, C., and BOULTON, D. (1968) An information measure for classification. *Comput. J.* **11** (2): 185–194.
- WALLACE, C. S., and FREEMAN, P. R. (1987) Estimation and inference by compact coding. *J. R. Statist. Soc. B* **49** (3): 240–265.
- WARD, D. J., BLACKWELL, A. F., and MACKAY, D. J. C. (2000) Dasher – A data entry interface using continuous gestures and language models. In *Proc. User Interface Software and Technology 2000*, pp. 129–137.
- WARD, D. J., and MACKAY, D. J. C. (2002) Fast hands-free writing by gaze direction. *Nature* **418** (6900): 838.
- WELLING, M., and TEH, Y. W. (2001) Belief optimization for binary networks: A stable alternative to loopy belief propagation. In *Uncertainty in Artificial Intelligence: Proc. Seventeenth Conf. (UAI-2001)*, pp. 554–561. Morgan Kaufmann.
- WIBERG, N. (1996) *Codes and Decoding on General Graphs*. Dept. of Electrical Engineering, Linköping, Sweden PhD dissertation. Linköping Studies in Science and Technology No. 440.
- WIBERG, N., LOELIGER, H.-A., and KÖTTER, R. (1995) Codes and iterative decoding on general graphs. *European Trans. on Telecommunications* **6**: 513–525.
- WIENER, N. (1948) *Cybernetics*. Wiley.
- WILLIAMS, C. K. I., and RASMUSSEN, C. E. (1996) Gaussian processes for regression. In *Advances in Neural Information Processing Systems 8*, ed. by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo. MIT Press.
- WILLIAMS, C. K. I., and SEEGER, M. (2001) Using the Nyström Method to Speed Up Kernel Machines. In *Advances in Neural Information Processing Systems 13*, ed. by T. K. Leen, T. G. Diettrich, and V. Tresp, pp. 682–688. MIT Press.
- WITTEN, I. H., NEAL, R. M., and CLEARY, J. G. (1987) Arithmetic coding for data compression. *Communications of the ACM* **30** (6): 520–540.
- WOLF, J. K., and SIEGEL, P. (1998) On two-dimensional arrays and crossword puzzles. In *Proc. 36th Allerton Conf. on Communication, Control, and Computing, Sept. 1998*, pp. 366–371. Allerton House.
- WORTHEN, A. P., and STARK, W. E. (1998) Low-density parity check codes for fading channels with memory. In *Proc. 36th Allerton Conf. on Communication, Control, and Computing, Sept. 1998*, pp. 117–125.
- YEDIDIA, J. S. (2000) An idiosyncratic journey beyond mean field theory. Technical report, Mitsubishi Electric Research Laboratories. TR-2000-27.
- YEDIDIA, J. S., FREEMAN, W. T., and WEISS, Y. (2000a) Bethe free energy, Kikuchi approximations and belief propagation algorithms. Technical report, Mitsubishi Electric Research Laboratories. TR-2001-16.
- YEDIDIA, J. S., FREEMAN, W. T., and WEISS, Y. (2000b) Characterization of belief propagation and its generalizations. Technical report, Mitsubishi Electric Research Laboratories. TR-2001-15.
- YEDIDIA, J. S., FREEMAN, W. T., and WEISS, Y. (2000c) Generalized belief propagation. Technical report, Mitsubishi Electric Research Laboratories. TR-2000-26.
- YEDIDIA, J. S., FREEMAN, W. T., and WEISS, Y. (2002) Constructing free energy approximations and generalized belief propagation algorithms. Technical report, Mitsubishi Electric Research Laboratories. TR-2002-35.
- YEUNG, R. W. (1991) A new outlook on Shannon-information measures. *IEEE Trans. Info. Theory* **37** (3.1): 466–474.
- YUILLE, A. L. (2001) A double-loop algorithm to minimize the Bethe and Kikuchi free energies. In *Energy Minimization Methods in Computer Vision and Pattern Recognition. Proc. Third International Workshop, Sophia Antipolis France, September 3-5, 2001*, ed. by M. Figueiredo, J. Zerubia, and A. Jain, number 2134 in LNCS, pp. 3–18. Springer.
- ZIPF, G. K. (1949) *Human Behavior and the Principle of Least Effort*. Addison-Wesley.