

**Fioretto, Ferdinando; Pontelli, Enrico; Yeoh, William; Dechter, Rina**  
**Accelerating exact and approximate inference for (distributed) discrete optimization with GPUs.** (English) [Zbl 1395.90187](#)  
[Constraints](#) 23, No. 1, 1-43 (2018).

Summary: Discrete optimization is a central problem in artificial intelligence. The optimization of the aggregated cost of a network of cost functions arises in a variety of problems including Weighted Constraint Programs (WCSPs), Distributed Constraint Optimization (DCOP), as well as optimization in stochastic variants such as the tasks of finding the most probable explanation (MPE) in belief networks. Inference-based algorithms are powerful techniques for solving discrete optimization problems, which can be used independently or in combination with other techniques. However, their applicability is often limited by their compute intensive nature and their space requirements. This paper proposes the design and implementation of a novel inference-based technique, which exploits modern massively parallel architectures, such as those found in Graphical Processing Units (GPUs), to speed up the resolution of exact and approximated inference-based algorithms for discrete optimization. The paper studies the proposed algorithm in both centralized and distributed optimization contexts. The paper demonstrates that the use of GPUs provides significant advantages in terms of runtime and scalability, achieving up to two orders of magnitude in speedups and showing a considerable reduction in execution time (up to 345 times faster) with respect to a sequential version.

**MSC:**

[90C10](#) Integer programming

**Keywords:**

[GPU](#); [WCSP](#); [MPE](#); [DCOP](#); [\(mini-\)bucket elimination](#); [\(A\)DPOP](#)

**Software:**

[CUDA](#); [ToulBar2](#)

**Full Text:** [DOI](#)

**References:**

- [1] Abdennadher, S; Schlenker, H, Nurse scheduling using constraint logic programming, 838-843, (1999)
- [2] Allouche, D; André, I; Barbe, S; Davies, J; Givry, S; Katsirelos, G; O'Sullivan, B; Prestwich, SD; Schiex, T; Traoré, S, Computational protein design as an optimization problem, *Artificial Intelligence*, 212, 59-79, (2014) · [Zbl 1407.92099](#)
- [3] Allouche, D., de Givry, S., Nguyen, H., & Schiex, T. (2013). Toulbar2 to solve Weighted Partial max-SAT. Tech. rep. INRA.
- [4] Apt, K. (2003). [\textit{Principles of constraint programming}](#). Cambridge University Press. · [Zbl 1187.68132](#)
- [5] Arbelaez, A; Codognet, P, A GPU implementation of parallel constraint-based local search, 648-655, (2014)
- [6] Barabási, AL; Albert, R, Emergence of scaling in random networks, *Science*, 286, 509-512, (1999) · [Zbl 1226.05223](#)
- [7] Bistaffa, F; Bomberì, N; Farinelli, A, CUBE: a CUDA approach for bucket elimination on gpus, (2016)
- [8] Bistarelli, S; Montanari, U; Rossi, F, Semiring-based constraint satisfaction and optimization, *Journal of the ACM*, 44, 201-236, (1997) · [Zbl 0890.68032](#)
- [9] Boyer, V; El Baz, D; Elkihel, M, Solving knapsack problems on GPU, *Computers & Operations Research*, 39, 42-47, (2012) · [Zbl 1251.90014](#)
- [10] Brito, I; Meseguer, P, Improving DPOP with function filtering, 141-158, (2010)
- [11] Burke, EK; De Causmaecker, P; Berghe, GV; Landeghem, H, The state of the art of nurse rostering, *Journal of scheduling*, 7, 441-499, (2004) · [Zbl 1154.90422](#)
- [12] Campeotto, F; Dovier, A; Fioretto, F; Pontelli, E, A GPU implementation of large neighborhood search for solving constraint optimization problems, 189-194, (2014)
- [13] Campeotto, F; Palù, AD; Dovier, A; Fioretto, F; Pontelli, E, A constraint solver for flexible protein model, *Journal of Artificial Intelligence Research*, 48, 953-1000, (2013)

- [14] Chakroun, I; Mezmaiz, MS; Melab, N; Bendjoudi, A, Reducing thread divergence in a GPU-accelerated branch-and-bound algorithm, *Concurrency and Computation: Practice and Experience*, 25, 1121-1136, (2013)
- [15] Dechter, R, Bucket elimination: a unifying framework for reasoning, *Artificial Intelligence*, 113, 41-85, (1999) · [Zbl 0939.68847](#)
- [16] Dechter, R. (2003). *Constraint processing*. San Francisco: Morgan Kaufmann Publishers Inc. · [Zbl 1057.68114](#)
- [17] Dechter, R, Reasoning with probabilistic and deterministic graphical models: exact algorithms, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 7, 1-191, (2013) · [Zbl 1297.68006](#)
- [18] Dechter, R., & Pearl, J. (1988). *Network-based heuristics for constraint-satisfaction problems*. Springer. · [Zbl 0643.68156](#)
- [19] Dechter, R; Rish, I, Mini-buckets: a general scheme for bounded inference, *Journal of the ACM*, 50, 107-153, (2003) · [Zbl 1326.68335](#)
- [20] Diamos, GF; Ashbaugh, B; Maiyuran, S; Kerr, A; Wu, H; Yalamanchili, S, SIMD re-convergence at thread frontiers, 477-488, (2011)
- [21] Dovier, A; Formisano, A; Pontelli, E, Autonomous agents coordination: action languages meet CLP() and linda, *Theory and Practice of Logic Programming*, 13, 149-173, (2013) · [Zbl 1267.68079](#)
- [22] Edelkamp, S; Jabbar, S; Schrödl, S, External A\*, 226-240, (2004) · [Zbl 1132.68698](#)
- [23] Farinelli, A; Rogers, A; Petcu, A; Jennings, N, Decentralised coordination of low-power embedded devices using the MAX-sum algorithm, 639-646, (2008)
- [24] Fioretto, F; Dovier, A; Pontelli, E, Constrained community-based gene regulatory network inference, *ACM Trans. Model. Comput. Simul.*, 25, 11, (2015) · [Zbl 1369.92039](#)
- [25] Fioretto, F; Le, T; Yeoh, W; Pontelli, E; Son, TC, Improving DPOP with branch consistency for solving distributed constraint optimization problems, 307-323, (2014)
- [26] Fioretto, F; Le, T; Yeoh, W; Pontelli, E; Son, TC, Exploiting GPUs in solving (distributed) constraint optimization problems with dynamic programming, 121-139, (2015)
- [27] Fioretto, F; Yeoh, W; Pontelli, E, A dynamic programming-based MCMC framework for solving DCOPs with gpus, 813-831, (2016)
- [28] Fioretto, F; Yeoh, W; Pontelli, E, Multi-variable agent decomposition for dcops, 2480-2486, (2016)
- [29] Fioretto, F; Yeoh, W; Pontelli, E, A multiagent system approach to scheduling devices in smart homes, 981-989, (2017)
- [30] Fioretto, F; Yeoh, W; Pontelli, E; Ma, Y; Ranade, S, A DCOP approach to the economic dispatch with demand response, 981-989, (2017)
- [31] Fishelson, M; Geiger, D, Exact genetic linkage computations for general pedigrees, *Bioinformatics*, 18, s189-s198, (2002)
- [32] Friedman, N; Linal, M; Nachman, I; Pe'er, D, Using Bayesian networks to analyze expression data, *Journal of Computational Biology*, 7, 601-620, (2000)
- [33] Gaudreault, J; Frayret, JM; Pesant, G, Distributed search for supply chain coordination, *Computers in Industry*, 60, 441-451, (2009)
- [34] Gupta, S., Yeoh, W., Pontelli, E., Jain, P., & Ranade, S.J. (2013). Modeling microgrid islanding problems as DCOPs. In *North American power symposium (NAPS)* (pp. 1-6): IEEE.
- [35] Hamadi, Y; Bessière, C; Quinqueton, J, Distributed intelligent backtracking, 219-223, (1998)
- [36] Han, TD; Abdelrahman, TS, Reducing branch divergence in GPU programs, 3:1-3:8, (2011), New York
- [37] Kask, K., Dechter, R., & Gelfand, A.E. (2012). Beem: bucket elimination with external memory. [arXiv:1203.3487](#).
- [38] Kumar, A; Faltings, B; Petcu, A, Distributed constraint optimization with structured resource constraints, 923-930, (2009)
- [39] Lalami, ME; El Baz, D; Boyer, V, Multi GPU implementation of the simplex algorithm, 179-186, (2011)
- [40] Larrosa, J, Node and arc consistency in weighted csp, 48-53, (2002)
- [41] Lars, O., & Rina, D. (2017). And/or branch-and-bound on a computational grid. *Journal of Artificial Intelligence Research* (to appear). · [Zbl 1418.68189](#)
- [42] Le, T; Fioretto, F; Yeoh, W; Son, TC; Pontelli, E, ER-DCOPS: a framework for distributed constraint optimization with uncertainty in constraint utilities, 605-614, (2016)
- [43] Lerner, U; Parr, R; Koller, D; Biswas, G; et al., Bayesian fault detection and diagnosis in dynamic systems, 531-537, (2000)
- [44] Lim, H., Yuan, C., & Hansen, E.A. (2010). Scaling up map search in bayesian networks using external memory. *On Probabilistic Graphical Models*, 177. · [Zbl 1226.05223](#)
- [45] Maheswaran, R; Tambe, M; Bowring, E; Pearce, J; Varakantham, P, Taking DCOP to the real world: efficient complete solutions for distributed event scheduling, 310-317, (2004)
- [46] Marinescu, R; Dechter, R, Memory intensive and/or search for combinatorial optimization in graphical models, *Artificial Intelligence*, 173, 1492-1524, (2009) · [Zbl 1185.68649](#)
- [47] Modi, P; Shen, WM; Tambe, M; Yokoo, M, ADOPT: asynchronous distributed constraint optimization with quality guarantees, *Artificial Intelligence*, 161, 149-180, (2005) · [Zbl 1132.68706](#)
- [48] Montanari, U, Networks of constraints: fundamental properties and applications to picture processing, *Information Sciences*, 7, 95-132, (1974) · [Zbl 0284.68074](#)
- [49] Pawłowski, K., Kurach, K., Michalak, T., & Rahwan, T. (2104). Coalition structure generation with the graphic processor

unit. Tech. Rep. CS-RR-13-07, Department of Computer Science, University of Oxford.

- [50] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufmann Publishers Inc. · [Zbl 0746.68089](#)
- [51] Pesant, G, A regular language membership constraint for finite sequences of variables, 482-495, (2004) · [Zbl 1152.68573](#)
- [52] Petcu, A; Faltings, B, Approximations in distributed optimization, 802-806, (2005) · [Zbl 1153.90583](#)
- [53] Petcu, A; Faltings, B, A scalable method for multiagent constraint optimization, 1413-1420, (2005)
- [54] Quimper, C.G., & Walsh, T. (2006). Global grammar constraints. In *Proceedings of the international conference on principles and practice of constraint programming (CP)* (pp. 751-755): Springer. · [Zbl 1160.68560](#)
- [55] Rodrigues, L., & Magatao, L. (2007). Enhancing supply chain decisions using constraint programming: a case study. In *MICAI 2007: advances in artificial intelligence*, (Vol. LNCS 4827 pp. 1110-1121): Springer.
- [56] Rossi, F., van Beek, P., & Walsh, T. (eds.) (2006). *Handbook of constraint programming*. Elsevier. · [Zbl 1175.90011](#)
- [57] Rust, P; Picard, G; Ramparany, F, Using message-passing DCOP algorithms to solve energy-efficient smart environment configuration problems, 468-474, (2016)
- [58] Sanders, J., & Kandrot, E. (2010). *CUDA By example. An introduction to general-purpose GPU programming*. Addison Wesley. · [Zbl 1132.68698](#)
- [59] Sandholm, T, Algorithm for optimal winner determination in combinatorial auctions, *Artificial Intelligence*, 135, 1-54, (2002) · [Zbl 0984.68039](#)
- [60] Schiex, T; Fargier, H; Verfaillie, G; etal., Valued constraint satisfaction problems: hard and easy problems, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 95, 631-639, (1995)
- [61] Shapiro, LG; Haralick, RM, Structural descriptions and inexact matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3, 504-519, (1981)
- [62] Silberstein, M., Schuster, A., Geiger, D., Patney, A., & Owens, J.D. (2008). Efficient computation of sum-products on gpus through software-managed cache. In *Proceedings of the 22nd annual international conference on supercomputing* (pp. 309-318): ACM. · [Zbl 0939.68847](#)
- [63] Sturtevant, NR; Rutherford, MJ, Minimizing writes in parallel external memory search, (2013)
- [64] Sultanik, E; Modi, PJ; Regli, WC, On modeling multiagent task scheduling as a distributed constraint optimization problem, 1531-1536, (2007)
- [65] Trick, MA, A dynamic programming approach for consistency and propagation for knapsack constraints, *Annals of Operations Research*, 118, 73-84, (2003) · [Zbl 1027.90075](#)
- [66] Yeoh, W; Felner, A; Koenig, S, Bnb-ADOPT: an asynchronous branch-and-bound DCOP algorithm, *Journal of Artificial Intelligence Research*, 38, 85-133, (2010) · [Zbl 1191.68733](#)
- [67] Yeoh, W; Yokoo, M, Distributed problem solving, *AI Magazine*, 33, 53-65, (2012)
- [68] Zivan, R; Yedidsion, H; Okamoto, S; Glinton, R; Sycara, K, Distributed constraint optimization for teams of mobile sensing agents, *Journal of Autonomous Agents and Multi-Agent Systems*, 29, 495-536, (2015)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.