



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Graph Neural Network Acceleration with SODA Framework

MASTER OF SCIENCE THESIS IN
COMPUTER SCIENCE AND ENGINEERING

Author: **Giovanni Demasi**

Student ID: 987062

Advisor: Prof. Fabrizio Ferrandi

Co-advisors: Serena Curzel

Academic Year: 2022-23

Abstract

Here goes the Abstract in English of your thesis followed by a list of keywords. The Abstract is a concise summary of the content of the thesis (single page of text) and a guide to the most important contributions included in your thesis. The Abstract is the very last thing you write. It should be a self-contained text and should be clear to someone who hasn't (yet) read the whole manuscript. The Abstract should contain the answers to the main scientific questions that have been addressed in your thesis. It needs to summarize the adopted motivations and the adopted methodological approach as well as the findings of your work and their relevance and impact. The Abstract is the part appearing in the record of your thesis inside POLITesi, the Digital Archive of PhD and Master Theses (Laurea Magistrale) of Politecnico di Milano. The Abstract will be followed by a list of four to six keywords. Keywords are a tool to help indexers and search engines to find relevant documents. To be relevant and effective, keywords must be chosen carefully. They should represent the content of your work and be specific to your field or sub-field. Keywords may be a single word or two to four words.

Keywords: here, the keywords, of your thesis

Abstract in lingua italiana

Qui va l'Abstract in lingua italiana della tesi seguito dalla lista di parole chiave.

Parole chiave: qui, vanno, le parole chiave, della tesi

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
Introduction	1
1 Background	3
2 Conclusions and future developments	5
Bibliography	7
A Appendix A	9
B Appendix B	11
List of Figures	13
List of Tables	15
List of Symbols	17
Acknowledgements	19

Introduction

Introduction here

1 Background

Background here...

2 Conclusions and future developments

Final chapter containing the main conclusions of my research and possible future developments.

Bibliography

- [1] S. Abi-Karam, Y. He, R. Sarkar, L. Sathidevi, Z. Qiao, and C. Hao. Gengnn: A generic FPGA framework for graph neural network acceleration. *CoRR*, abs/2201.08475, 2022. URL <https://arxiv.org/abs/2201.08475>.
- [2] N. B. Agostini, S. Curzel, J. J. Zhang, A. Limaye, C. Tan, V. Amatya, M. Minutoli, V. G. Castellana, J. Manzano, D. Brooks, G.-Y. Wei, and A. Tumeo. Bridging python to silicon: The soda toolchain. *IEEE Micro*, 42(5):78–88, 2022. doi: 10.1109/MM.2022.3178580.
- [3] A. Auten, M. Tomei, and R. Kumar. Hardware acceleration of graph neural networks. In *2020 57th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6, 2020. doi: 10.1109/DAC18072.2020.9218751.
- [4] A. Bik, P. Koanantakool, T. Shpeisman, N. Vasilache, B. Zheng, and F. Kjolstad. Compiler support for sparse tensor computations in MLIR. *ACM Transactions on Architecture and Code Optimization*, 19(4):1–25, sep 2022. doi: 10.1145/3544559. URL <https://doi.org/10.1145%2F3544559>.
- [5] U. Bondhugula. High performance code generation in MLIR: an early case study with GEMM. *CoRR*, abs/2003.00532, 2020. URL <https://arxiv.org/abs/2003.00532>.
- [6] S. Böhm. How to optimize a cuda matmul kernel for cublas-like performance: a worklog, 2022. URL <https://siboehm.com/articles/22/CUDA-MMM>.
- [7] L. He. Engn: A high-throughput and energy-efficient accelerator for large graph neural networks. *CoRR*, abs/1909.00155, 2019. URL <http://arxiv.org/abs/1909.00155>.
- [8] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 22118–22133. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf.

- [9] Y. Hu, Y. Du, E. Ustun, and Z. Zhang. Graphlily: Accelerating graph linear algebra on hbm-equipped fpgas. In *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pages 1–9, 2021. doi: 10.1109/ICCAD51958.2021.9643582.
- [10] K. Kinningham, C. Ré, and P. A. Levis. GRIP: A graph neural network accelerator architecture. *CoRR*, abs/2007.13828, 2020. URL <https://arxiv.org/abs/2007.13828>.
- [11] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, abs/1609.02907, 2016. URL <http://arxiv.org/abs/1609.02907>.
- [12] S. Liang, C. Liu, Y. Wang, H. Li, and X. Li. Deepburning-gl: an automated framework for generating graph neural network accelerators. In *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pages 1–9, 2020.
- [13] B. Sanchez-Lengeling, E. Reif, A. Pearce, and A. B. Wiltschko. A gentle introduction to graph neural networks. *Distill*, 2021. doi: 10.23915/distill.00033. <https://distill.pub/2021/gnn-intro>.
- [14] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks?, 2019.
- [15] M. Yan, L. Deng, X. Hu, L. Liang, Y. Feng, X. Ye, Z. Zhang, D. Fan, and Y. Xie. Hygcn: A GCN accelerator with hybrid architecture. *CoRR*, abs/2001.02514, 2020. URL <http://arxiv.org/abs/2001.02514>.

A Appendix A

If you need to include an appendix to support the research in your thesis, you can place it at the end of the manuscript. An appendix contains supplementary material (figures, tables, data, codes, mathematical proofs, surveys, . . .) which supplement the main results contained in the previous chapters.

B Appendix B

It may be necessary to include another appendix to better organize the presentation of supplementary material.

List of Figures

List of Tables

List of Symbols

Variable	Description	SI unit
\boldsymbol{u}	solid displacement	m
\boldsymbol{u}_f	fluid displacement	m

Acknowledgements

Acknowledgements here...

