

Building Beyond HLS: Graph Analysis and Others

Pedro Filipe Silva
Faculty of Engineering, University of
Porto
Porto, Portugal
pedro.filipe.silva@fe.up.pt

João Bispo
INESC-TEC and Faculty of
Engineering, University of Porto
Porto, Portugal
jbispo@fe.up.pt

Nuno Paulino
INESC-TEC and Faculty of
Engineering, University of Porto
Porto, Portugal
nuno.m.paulino@inesctec.pt

ABSTRACT

High-Level Synthesis has introduced reconfigurable logic to a new world – that of software development. The newest wave of HLS tools has been successful, and the future looks bright. But is HLS the end-all-be-all to FPGA acceleration? Is it enough to allow non-experts to program FPGAs successfully, even when dealing with troublesome data structures and complex control flows – such as those often encountered in graph algorithms? We take a look at the panorama of adoption of HLS by the software community, focusing on graph analysis in particular in order to generalise to *FPGA-unfriendly* problems. We argue for the existence of shortcomings in current HLS development flows which hinder adoption, and present our perspective on the path forward, including how these issues may be remedied via higher-level tooling.

1 INTRODUCTION

The High-Level Synthesis (HLS) approach to FPGA hardware synthesis has quite a storied past. From the 1970s-80s, generations of HLS (or behavioural synthesis, as it was once known) have come and gone [11] – but the current one has shown staying power. Current HLS tools have reached new heights in academic development, sophistication, and commercial success [13].

As a result, these tools are now marketed to pure software developers: by many accounts, HLS is ready to break into the software world [9, 10]. The opposing viewpoint persists, however: that implementing FPGA accelerators with current HLS tools still requires significant hardware development knowledge.

We present a somewhat more nuanced viewpoint: *Assuming the existence of higher-level tools such as libraries*, it is perfectly possible for a pure software engineer to implement an FPGA accelerator. Take a relatively mature product such as the the Vitis Vision Library [18] – browsing through examples shows the software abstraction mostly holds tight (if one does not descend to the internals). But this begs a question: what happens when there are no viable high-level tools?

2 WHAT’S MISSING? (A “CASE STUDY” ON GRAPH ANALYSIS)

Graphs are a common sight in a myriad of scientific and engineering problems and models (e.g., traffic navigation, electrical network

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

LATTE '21, April 15, 2021, Virtual, Earth

© 2021 Copyright held by the owner/author(s).

Table 1: FPGA Graph Analytic Work

Type	Number	NPDU	Problems
HLS	4	N/A	BFS*3, SpMV*2
HLS Framework	3	0	CCent, PR*2, BFS*2, SpMV*3, SSSP*2, WCC*2, MST, VC
RTL	15	N/A	PR, BFS*8, SSSP*3, APSP, MM, GC
RTL Framework	12	7	PR*7, BFS*7, SpMV*2, SSSP*4, WCC*5, CC, MST, TRW-S, CNN, VC

analysis) [8], making research on graph algorithms – and their acceleration – very active. Notably, graph analysis tends to be both slow and very common in performance-sensitive applications (as is traffic navigation), meaning acceleration is all the more relevant.

FPGA graph accelerators are notoriously difficult. Low global memory bandwidths and low amounts of fast local memory do not mesh with the high memory dependence (often including random access requirements) of graph analytic algorithms. As a result, the combination remains somewhat unexplored [2].

Table 1 shows work on FPGA graph algorithms from 2010 onward, grouped by type¹². Note that we define *framework* as a work aiming to provide a high-level generalised tool able to solve several distinct problems, rather than focusing on a specific problem (e.g., shortest path computation).

Register-Transfer Level (RTL) work overwhelms HLS work in number of occurrences. This may indicate that HLS is not yet considered fit for the task, or that RTL development methods, versus HLS, still carry significant momentum in the FPGA community. However, recent literature exists arguing that HLS is indeed ready (with some caveats) [9], and some novel HLS work claims better results when compared to RTL frameworks [4], so we tend towards the latter option.

An analysis of the works compiled in Table 1 brings us to the following observations:

Patterns. In general, most works on FPGA graph acceleration possess at least one of the following attributes:

- (1) Tackling a single problem.
- (2) Not having available/runnable code.
- (3) Being either RTL-based or unclear about implementation.

The Limits of RTL. When analysing non-HLS literature in particular, a key takeaway is that the majority of works *does* focus on a single problem, and this tends to be a variation of the shortest-path problem. Are pure RTL designs hitting an abstraction limit, where other kinds of computations are very difficult to model? Note

¹Results for the *RTL* line were sourced from [2], and only cover up to the year 2019.

²*NPDU* means “number of works which poorly define programmer usability” – we reference this below.

that several frameworks, especially RTL ones, fail to clearly define how a programmer might use them. This may indicate that usage of such frameworks still requires very specialised knowledge.

A Hard Problem. Much of the literature (including HLS work) is written mostly from a hardware engineer’s perspective. This is likely indicative of the special status of graph analysis in the FPGA world: it is an *FPGA-unfriendly*³ problem, still requiring a fair amount of hardware knowledge (and time investment) to tackle, even with HLS. Due to this, adoption of software paradigms, as has occurred in other areas, is slowed. We base our argument for higher-level tooling in this fact.

Symptoms. All factors appear to be symptoms of a problem: RTL cannot fully tackle graph analysis, and HLS isn’t quite there yet.

3 WHY ARE SOFTWARE DEVELOPERS INTERESTED IN FPGAS?

The answer is, of course, HLS [5, 10]. But why are they interested in using *FPGA boards and expansion cards*?

The Good. In the HPC community, factors of interest in FPGA technology include:

- (1) Potential time or energy efficiency increase [3, 10, 12, 16].
- (2) Interest in or perceived architectural advantages vs. traditional processing systems [3, 10].
- (3) Interest in reconfigurable heterogeneous computing systems [12, 16].

The Bad. Conversely, referenced issues include:

- (1) Difficulty in implementation [3] vs. GPU systems [12].
- (2) Lack of portability between FPGA vendors [16].
- (3) Optimisation challenges [16].
- (4) Reduced quality-of-results, in some instances, vs. RTL tools [9].

The Lovely. However, a fact stands out: *all issues could benefit from higher-level tooling*. Raising the level of abstraction is, as history shows, the most direct method of resolving issues 1 and 2. As for issues 3 and 4, modern optimising compilers give a hint: in many, or most, scenarios, programming in a high level (systems) language results in more efficient program code – an effect more visible the more complex the application or target hardware become, which is precisely where we are now heading with FPGA acceleration.

4 DISCUSSION AND CONCLUSION

The time is right for redoubling research into FPGA acceleration, especially in new libraries, abstractions, and toolsets. FPGA usage is increasing, and recent technologies, such as High-Bandwidth Memory (HBM), may come to increase popularity even further.

The state of the art in FPGA *graph* acceleration is also shifting: where single-issue, low-level, RTL implementations dominated, now appear high-level frameworks, and, most interestingly, *HLS-based* frameworks (we ourselves are evaluating such a framework [4] to implement graph centrality metrics for traffic navigation).

³Less amenable to implementation than, for instance, streaming-type algorithms. Note that we do not argue that efficient implementation is impossible or nonviable; only that it is *harder*.

We have used graph analysis to generalise to unfriendly applications, but some of the issues we’ve raised may very well be applicable to *all* domains of FPGA acceleration. The remainder this section may be read in either light.

4.1 Parallels

Abstraction is a key feature of hardware [6] and software [17] engineering alike. Increased abstraction levels are a necessity as system complexity increases. Not unlike the progression, in software engineering, from machine code to assembly to high-level languages⁴, hardware synthesis has evolved from manual design to logic synthesis, and, more recently, high-level synthesis. Could the two parallel paths actually converge at a distant point to a perfectly agnostic behavioural description language?

In 2009, Martin and Smith [11] divided the history of HLS into three distinct generations, with an upcoming fourth hinging on conquering control flow. This has been done (to what degree, however, is debatable). So will the *fifth* generation focus on raising the abstraction level further, perhaps via higher-level tooling, to conquer unfriendly applications?

4.2 So, What’s Missing?

The way forward for FPGAs goes through HLS – this is a reoccurring sentiment [9, 10, 14]. In fact, RTL may be hitting an abstraction wall. Perhaps it will come to be viewed as the equivalent of assembly for hardware design [9]. But research cannot stop at pure HLS.

While HLS has come to significantly raise the abstraction level⁵, this is not sufficient for unfriendly problems such as graph analysis: hardware knowledge remains a necessity in these instances [1, 4, 15]. Thus, we raise the following questions:

- Are FPGA HLS accelerators currently competitive for graph analysis, or will they become competitive in the near future? If so, how?
- Are HLS libraries and toolsets for FPGA acceleration, especially in unfriendly applications, mature enough for use in production? If not, when will they be?
- How will increasing support for HBM affect the adoption of FPGAs as graph accelerators for unfriendly applications such as graph analysis?
- What other promising technologies (such as multi-FPGA systems, overlays, dynamic partial reconfiguration, or run-time binary translation) have the potential to increase said adoption?
- What impact will these technologies have on the potential of FPGAs as general-purpose hardware accelerators?

We argue that the continued development and promotion of tools such as frameworks and libraries is *necessary* in order to move the burden of specialised knowledge away from the domain expert. As such, we hold that the future of FPGA acceleration, especially in remnant markets, will heavily depend not only on advances in compiler technology, but also on investment into high-level tooling.

⁴Although an argument could be made that C presents a faulty/leaky abstraction, due to its closeness to the underlying hardware [7, 17]. Perhaps perfect abstractions are unattainable, or even undesirable?

⁵Also not without fault – as a review of any highly optimised pure HLS quickly indicates (we reference this in Section 1).

REFERENCES

- [1] Dario Baptista, Leonel Sousa, and Fernando Morgado-Dias. 2020. Raising the Abstraction Level of a Deep Learning Design on FPGAs. *IEEE Access* 8 (2020), 205148–205161. <https://doi.org/10.1109/ACCESS.2020.3036975>
- [2] Maciej Besta, Dimitri Stanojevic, Johannes De Fine Licht, Tal Ben-Nun, and Torsten Hoefler. 2019. Graph Processing on FPGAs: Taxonomy, Survey, Challenges. *arXiv:1903.06697 [cs]* (April 2019). arXiv:1903.06697 [cs]
- [3] Nick Brown. 2020. Weighing Up the New Kid on the Block: Impressions of Using Vitis for HPC Software Development. In *2020 30th International Conference on Field-Programmable Logic and Applications (FPL)*. IEEE, Gothenburg, Sweden, 335–340. <https://doi.org/10.1109/FPL50879.2020.00062>
- [4] Xinyu Chen, Hongshi Tan, Yao Chen, Bingsheng He, Weng-Fai Wong, and Deming Chen. 2021. ThunderGP: HLS-Based Graph Processing Framework on FPGAs. In *The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. ACM, Virtual Event USA, 69–80. <https://doi.org/10.1145/3431920.3439290>
- [5] Jason Cong, Bin Liu, Stephen Neuendorffer, Juanjo Noguera, Kees Vissers, and Zhiru Zhang. 2011. High-Level Synthesis for FPGAs: From Prototyping to Deployment. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 30, 4 (April 2011), 473–491. <https://doi.org/10.1109/TCAD.2011.2110592>
- [6] P. Coussy, D.D. Gajski, M. Meredith, and A. Takach. 2009. An Introduction to High-Level Synthesis. *IEEE Design & Test of Computers* 26, 4 (July 2009), 8–17. <https://doi.org/10.1109/MDT.2009.69>
- [7] S.A. Edwards. 2006. The Challenges of Synthesizing Hardware from C-Like Languages. *IEEE Design & Test of Computers* 23, 5 (May 2006), 375–386. <https://doi.org/10.1109/MDT.2006.134>
- [8] L. R Foulds. 1992. *Graph Theory Applications*. Springer New York, New York.
- [9] Sakari Lahti, Panu Sjoval, Jarno Vanne, and Timo D. Hamalainen. 2019. Are We There Yet? A Study on the State of High-Level Synthesis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 38, 5 (May 2019), 898–911. <https://doi.org/10.1109/TCAD.2018.2834439>
- [10] Joshua Lant, Javier Navaridas, Mikel Lujan, and John Goodacre. 2020. Toward FPGA-Based HPC: Advancing Interconnect Technologies. *IEEE Micro* 40, 1 (Jan 2020), 25–34. <https://doi.org/10.1109/MM.2019.2950655>
- [11] G. Martin and G. Smith. 2009. High-Level Synthesis: Past, Present, and Future. *IEEE Design & Test of Computers* 26, 4 (July 2009), 18–25. <https://doi.org/10.1109/MDT.2009.83>
- [12] Fahad Bin Muslim, Liang Ma, Mehdi Roozmeh, and Luciano Lavagno. 2017. Efficient FPGA Implementation of OpenCL High-Performance Computing Applications via High-Level Synthesis. *IEEE Access* 5 (2017), 2747–2762. <https://doi.org/10.1109/ACCESS.2017.2671881>
- [13] Razvan Nane, Vlad-Mihai Sima, Christian Pilato, Jongsok Choi, Blair Fort, Andrew Canis, Yu Ting Chen, Hsuan Hsiao, Stephen Brown, Fabrizio Ferrandi, Jason Anderson, and Koen Bertels. 2016. A Survey and Evaluation of FPGA High-Level Synthesis Tools. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 35, 10 (Oct. 2016), 1591–1604. <https://doi.org/10.1109/TCAD.2015.2513673>
- [14] Maxime Pelcat, Cedric Bourrasset, Luca Maggiani, and Francois Berry. 2016. Design Productivity of a High Level Synthesis Compiler versus HDL. In *2016 International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation (SAMOS)*. IEEE, Agios Konstantinos, Samos Island, Greece, 140–147. <https://doi.org/10.1109/SAMOS.2016.7818341>
- [15] Nik Sultana, Salvator Galea, David Greaves, Marcin Wojcik, Jonny Shipton, Richard Clegg, Luo Mai, Pietro Bressana, Robert Soulé, Richard Mortier, Paolo Costa, Peter Pietzuch, Jon Crowcroft, Andrew W Moore, and Noa Zilberman. 2017. Emu: Rapid Prototyping of Networking Services. In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*. USENIX Association, Santa Clara, CA, 459–471.
- [16] Dennis Weller, Fabian Oboril, Dimitar Lukarski, Juergen Becker, and Mehdi Tahoori. 2017. Energy Efficient Scientific Computing on FPGAs Using OpenCL. In *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. ACM, Monterey California USA, 247–256. <https://doi.org/10.1145/3020078.3021730>
- [17] Niklaus Wirth. 2008. A Brief History of Software Engineering. *IEEE Annals of the History of Computing* 30, 3 (July 2008), 32–39. <https://doi.org/10.1109/MAHC.2008.33>
- [18] Xilinx. 2021. Vitis Vision Library.