Parallel Programming Applied Research Projects for Teaching Parallel Programming to Beginner Students

Youry Khmelevsky  
Computer Science, Okanagan College  
Kelowna, BC, Canada  
Email: ykhmelevsky@okanagan.bc.ca

Gaëtan J.D.R. Hains  
LACL, Univ. Paris-Est Créteil  
Paris, France  
Email: gaetan.hains@gmail.com

*Also Affiliated with Mathematics, Statistics, Physics, and Computer Science  
Irving K. Barber School of Arts and Sciences, UBC Okanagan, BC Canada

Abstract—In this paper, we discuss the educational value of a few mid-size and one large applied research projects at the Computer Science Department of Okanagan College (OC) and at the Universities of Paris East Créteil (LACL) and Orléans (LIFO) in France.

We found, that some freshmen students are very active and eager to be involved in applied research projects starting from the second semester. They are actively participating in programming competitions and want to be involved in applied research projects to compete with sophomore and older students. Our observation is based on five NSERC Engage College and Applied Research and Development (ARD) grants, and several small applied projects.

Student involvement in applied research is a key motivation and success factor in our activities, but we are also involved in transferring some results of applied research, namely programming techniques, into the parallel programming courses for beginners at the senior- and first-year MSc levels. We illustrate this feedback process with programming notions for beginners, practical tools to acquire them and the overall success/failure of students as experienced for more than 10 years in our French University courses.

I. INTRODUCTION

This paper is an attempt to analyze our results of 12 years of small and mid-size applied research projects introduction to students initially started at COSC 224 Capstone Project course (4th semester) and at COSC 470/471 Software Engineering Capstone Projects at the Computer Science department at Okanagan College and then to freshmen students from 2nd semester in sponsored by NSERC and by industry applied research projects. We also analyze the results of almost two decades of teaching beginner’s parallel programming at the Universities of Orléans and Paris-Est Créteil, in this case to final-year undergraduates and first-year MSc Computer Science students [1], [2], [3], [4]. Every year, starting from 2005 we searched for new solutions and new pedagogical approaches within COSC 224 — Projects in Computer Science, COSC 470 — Software Engineering and COSC 471 — Software Engineering Project Project courses [5], [6], [7], [8], [9]. Starting from 2007 we tried to introduce research projects to all our capstone projects and some local industrial clients supported our idea [10], but it took almost 7 extra years to get support for our student research or research related projects from NSERC. Several industrial projects were so successful that we were able to obtain an NSERC Community College Innovation (CCI) Applied Research grants (Engage College grants), two Mitacs grants with SFU and UBCO, an Amazon's AWS in Education Research grant, and several small grants funded by the College for the student research projects with industrial sponsors [11], [12], [13], [14].

As other educators and computer science (CS) researchers discovered, CS “research has an opportunity to change all other disciplines. Much of the current CS research is of an incremental stovepipe or inward-looking nature, even though CS has a unique ability to fundamentally transform virtually every other discipline” [15]. Moreover, the “undergraduate research experiences are widely promoted in science, technology, engineering, and mathematics (STEM) for their potential to bring several benefits to students, like: (1) improved retention in both the major and discipline-related careers; (2) ability to work independently and to communicate well with a team; (3) increased confidence in academic knowledge and technical skill; (3) broader awareness of the discipline and and awareness of career opportunities and support for making career choices.”

Another researcher in [16] concluded: “these positive outcomes are not guaranteed simply because a student participates; however, the experience must truly be an experience of research and with enough structure that the student is able to be successful.”

A few years ago several freshmen students contacted us with requests to organize students programming competitions and to be accepted into our undergraduate research projects. They successfully competed in ICPC and IEEExtreme programming competitions and achieved good results. As an experiment, we accepted a few freshmen students to our research projects and we found, that some talented freshman students can contribute to the research projects on the same level, if not better as 3rd
and 4th years undergraduate students.
In the next section we will discuss our ongoing projects related to automatic code generation and parallel programming programming initial project results. This is concrete evidence of the upward transfer of skills from undergraduate education to applied research in parallel programming. We also summarize results of a long-term experiment in teaching parallel programming to beginners with simplified and formally verified programming interfaces, that are the result of applied research projects by academics and PhD students.

II. Teaching Parallel Programming for Freshmen Student in Applied Research Projects

About 7 years ago we started parallel programming and automatic code generation as small applied research projects within our capstone framework at the Computer Science Department of Okanagan College, and also at the UBC Okanagan Campus, [17][18], [19], [20]. Students enjoyed projects within capstone project-courses [17], [21], [22], [23] and several projects resulted in student research papers [24], [25], [26], [27]. Starting from 2016 we also accepted freshmen students to the parallel programming and automatic code generation small applied research projects [28], [29], [30], [31], [32], [33]. We found that such projects, additionally to a small stipend for the student research work creates a strong motivation in courses starting from the first year. Students see immediate application of their knowledge in the small applied research projects at the College. They understand better the need for first year courses such as COSC 126—System Analysis and Design, COSC 109 — Technical Aspects of Operating Systems, COSC 118 — Networks and Telecommunications I, COSC 131 — Visual Programming, and two communication courses in addition to traditional programming courses.

The second author (Hains) assisted with project goals development and with the students supervision. He taught for about 20 years bulk-synchronous parallel (BSP) programming, which is based on a paradigm of global synchronous communications allowing a realistic performance (“cost”) model—that leads to portable and scalable algorithm design, or in the case of beginner students, the realistic prototyping of parallel algorithms without parallel machine. We already introduced BSP parallel programming paradigm into small applied research projects successfully [34] and continue our educational development.

In 2017 we started a new freshman parallel programming short applied research project on the performance evaluation of the Parallel MATLAB library. This is both an exercise in declarative numerical programming, and a systematic introduction to performance benchmarking and modelling. In previous research projects sophomore and upper level students successfully investigated servers workload and many different applications/game servers performance issues [34], [25], [27], [35], but those projects were too complex for the freshman students. After observing this problem we adapted our tasks and research projects for very young, but skillful students. We assigned the following tasks for the freshmen students realizing the parallel MATLAB project:

- Install the Parallel MATLAB library.
- Try examples with it to write pure-declarative code (pure operations on arrays) and obtain multicore/multinode parallel execution.
- Design a simple performance model for it (how fast for how many cores, how much communication etc) and test that model against measurements. The “model” can be as simple as a low-degree polynomial giving execution time against data dimensions and the number of processing cores.
- Document all the test data with the experiment setup: software (SW) versions, hardware, number of cores used, (number of threads if its visible), measured quantity (time, space etc).
- For a simple operation like Matrix x vector that would mean the size of matrix etc show the Matlab expressions what are compiling.
- Then try to cover the space of tests to obtain ideally 3D graphs with (time or memory) vs cores x (data size). From that we can interpolate surfaces (or just curves if the graphs were 2D) that with the help of your supervisor(s) guess from parallel performance models. That helps understand the general compiler or algorithm as applied to a specific parallel machine (in this case a one-node or two-node multicore PC). Each graph is about running the same Parallel MATLAB expression for various cores x (data size).
- Finally if time allows, a more creative project phase can lead to the initial phase of a serious project: design a language-based general cost model that can unify all the previous performance measurements in a syntax-based model of performance. In other words, design a parallel cost model for a whole sub-language of Parallel MATLAB.

The list of tasks above is just an initial investigation, but the freshman students were able to install MATLAB, add necessary libraries and run almost all tests in a few weeks. The project is still on-going, but students are excited and like to do experiments with the parallel code execution and coding experiments. Such a project is a good example of a simplified research project (namely, benchmarking and modelization of the a parallel library’s performance) with quantitative objectives that can be converted into a precisely defined and progressively difficult student project. Given enough time and effort it could lead young undergraduates to Master’s level research topics and an introduction to the general research problems of parallel programming and parallel software engineering.

III. Research Results in Parallel Programming and MSC-Level Teaching

The BSP (bulk-synchronous parallel) model of parallel computing [36], [37], [38] has been studied, applied and implemented since the late 1990s to improve the reliability, predictability and portable scalability of parallel programs.
It is based on the notion that a parallel algorithm is a sequence of supersteps where a vector of \( p \) computation units (cores, nodes, computers) execute as many asynchronous and independent phases of computation (threads, processes) before all exchanging necessary data in a globally synchronous communication phase. The advantages of the BSP model over ad-hoc parallel methods are many: determinism, a performance model that relates the communication phase to its underlying architecture in a linear way, possible checkpoints or debugging steps at the end of supersteps, “immortal” algorithms that can adapt to any future architecture etc.

We have taught theoretical BSP algorithms to doctoral students since the invention of the model by Valiant ca. 1988 and its in-depth investigation by McColl in the 1990s. We then developed a declarative API for programming BSP algorithms that can serve both as a parallel functional language and simplified tool for learning parallel programming. This API is called BSMLlib or “BSP-ML library” [39] because it is based on the OCaml (Objective-Ca-ML) functional language, a dialect of the ML language (a strongly-typed form of Lisp). Whatever the host language and functional programming details, BSMLlib summarized BSP programming to the following primitives:

- A global, load-time constant \( nprocs \) for the number \( p \) of processing units.
- A constructor and a transformer for building asynchronous computation steps of width \( p \). Those operators bind the local variable \( pid \) that identifies a local thread by its rank \( 0,1,\ldots,p−1 \).
- A destructor for folding parallel data into a local array or list that must then be processed locally and sequentially.
- A transformer based on a \( p \times p \) relation of message sending-reception for creating global communication-synchronization phases.

We have used BSMLlib to teach parallel programming to final-year undergraduates and first-year MSc students so that they learn the expected performance (BSP model) of any simple parallel program written with BSMLlib. The most demanding exercises involved advanced parallel data structures like balanced parallel hashing, or simulations like the n-body problem, or BSP sample-sort programming and benchmarking. One great advantage of BSMLlib is that a purely-sequential simulator of the library (1000 lines of OCaml) is able to produce exact counts of the communication and synchronization costs. If complemented with a bytecode-instruction counter for the asynchronous parallel phases, it is able to produce realistic and scalable estimates of runtime without any more complex technology than the bytecode-interpreted version of OCaml (which runs on Unix, Windows, a web-server service at OCamlPro SARL and even an Android port).

In all classes we observed that a large portion of the students became familiar and proficient with the constructor-destructor sub-language which allowed them to write a large subset of our elementary algorithms (sorting being a notable exception). The communication-synchronization transformer (called \( put \)) requires a two-dimensional coding of messages which is objectively difficult to acquire. In a more recent research project, [40] we designed a simplified form of BSMLlib called SGL (scatter-gather language) where the transformer operator \( put \) is absent so that all global operations appear like one-to-many constructors or many-to-one destructors. This language has the same advantages as BSMLlib plus:

- Is easier to learn as experienced in class and for a very precise reason learned from BSMLlib teaching.
- A centralized structure than can be nested to simulate heterogeneous architectures that are not “flat” e.g. multi-node systems of multicore nodes.
- Can express more than 80% of a set of basic application libraries (like list- and array APIs) without the use of a \( p \times p \) operator like \( put \).

Again SGL, the result of parallel programming research, has been applied to improved teaching for the benefit of all beginners, at well as inspiring new concrete ideas in parallel software engineering.

### IV. CONCLUSION

Software development skills are a key asset of modern economies. Training undergraduates in parallel programming is an important objective that has not yet been well covered by the CS curriculum. The following quotations from educators illustrate the general importance of teaching methods. “The software development and software development project courses with external and internal sponsors drastically increased learning activity, creativity, productivity and success of student work. Students like to see computer science as social, relevant, important, and caring endeavour” [41].

“The research workforce of the future is important for many STEM disciplines, but the computer science research workforce has become critical to the US and global economies” [16] and to Canada. “All of the engineering and science discoveries and innovations are now dependent on computational science” [8]. “Students recognize that research is a collaborative experience requiring communication skills to complete and disseminate the work” [42].

As a faculty at Okanagan College mentioned, we have a new generation of students, are looking for challenges during their study at post-secondary institutions. They like to work hard, but they want to be rewarded for their hard work. Many students apply to the College because they like small classes and the lower tuition to compare with the universities. On the other hand they are demanding a high quality education, involvement in the worldwide programming competitions (IEEE Xtreme 24-Hour Programming Competition\(^1\), the ACM International Collegiate Programming Contest (ICPC)\(^2\) and others), and involvement in industrial applied research projects, especially supported by Government and/or by large corporations. Many freshmen students want to continue their education at top universities within MSc and even in

http://www.ieee.org/membership_services/membership/students/competitions/xtreme/index.html
https://icpc.baylor.edu/
PhD programs. We believe that the results of MSc-level teaching and research work can be applied in very concrete ways to help more undergraduates acquire top skills in the development and performance-modelling of parallel programs. In return, undergraduate teaching can help improve parallel programming tools themselves, as well as motivating some of the students to pursue their technical training in the form of applied research projects.

The experiences described here somewhat contradict the traditional dichotomy between on the one hand, community-college teaching to undergraduates who have no exposure to research, and on the other hand University students who are exposed to the results of ongoing research projects.

We are confident that they are applicable to many other areas of software engineering, if not most of the computing disciplines.

V. ACKNOWLEDGMENTS

We thank the Natural Sciences and Engineering Research Council of Canada for supporting five of our applied research projects in 2014–2017. Our thanks go also to the AWS Programs for Research and Education program for supporting our research and educational projects, as well as to Atlassian. We thank the IBM Academic Initiative program for supporting us with Rational Rose engineering software, licenses and with training materials.

Research on BSMLib and SGL has benefited from many French-government grants including: the CARAML national project sponsored by the Ministry of Research, the HPIAF project sponsored by EXQIM S.A.S, Chong Li’s CIFRE doctoral project sponsored by ANRT and EXQIM. The second author thanks his many partners in BSMLib research: Frédéric Louergue, Frédéric Gava and Chong Li, to name just a few. Huawei Technologies’ Central Software Institute supports many efforts in parallel programming with the BSP model.

REFERENCES

[1] Y. Khmelevsky and S. Dhanjal, “Information security and data protection in computer science education,” in 12th Western Canadian Conference Education on Computing Education (WCCCE-2007), Thompson Rivers University, Kamloops, Canada, May, 2007, pp. 3–5.

[2] M. Govorov, Y. Khmelevsky, V. Ustimenko, and A. Khorev, “Security for gis n-tier architecture,” in Developments in Spatial Data Handling, Springer, 2005, pp. 71–83.

[3] Y. Khmelevsky, “SW Development Projects in Academia,” in Proceedings of the 14th Western Canadian Conference on Computing Education, ser. WCCCE ’09. New York, NY, USA: ACM, 2009, pp. 60–64. [Online]. Available: http://doi.acm.org/10.1145/1536274.1536292

[4] Y. Khmelevsky, M. Govorov, and L. Burge, “Okanagan College and Vancouver Island University educational joint projects results,” in Proceedings of the 14th Western Canadian Conference on Computing Education, 2009, pp. 65–69.

[5] Y. Khmelevsky and V. Voytenko, “Cloud computing infrastructure prototype for university education and research,” in WCCCE 10 Proceedings of the 15th Western Canadian Conference on Computing Education Article# 8.—ACM New York, NY, USA, 2010.

[6] Y. Khmelevsky, L. Burge, M. Govorov, and G. Hains, “Distance learning components in CS and GIS courses,” in Proceedings of the 16th Western Canadian Conference on Computing Education, ser. WCCCE ’11. New York, NY, USA: ACM, 2011, pp. 17–21. [Online]. Available: http://doi.acm.org/10.1145/1989622.1989627

[7] Y. Khmelevsky, V. Ustimenko, G. Hains, C. Khuka, E. Ozan, and D. Syrotovskiy, “International collaboration in SW engineering research projects,” in Proceedings of the 16th Western Canadian Conference on Computing Education, ser. WCCCE ’11. New York, NY, USA: ACM, 2011, pp. 52–56. [Online]. Available: http://doi.acm.org/10.1145/1989622.1989637

[8] Y. Khmelevsky, “Research and teaching strategies integration at post-secondary programs,” in Proceedings of the 16th Western Canadian Conference on Computing Education, ser. WCCCE ’11. New York, NY, USA: ACM, 2011, pp. 57–60. [Online]. Available: http://doi.acm.org/10.1145/1989622.1989638

[9] R. Grmek, Y. Khmelevsky, and D. Syrotovskiy, “Automated inventory tracking system prototype in cloud,” in High Performance Computing and Simulation (HPCS), 2011 International Conference on, July 2011, pp. 435–441.

[10] Y. Khmelevsky, “Ten Years of Capstone Projects at Okanagan College: A Retrospective Analysis,” in Proceedings of the 21st Western Canadian Conference on Computing Education, ser. WCCCE ’16. New York, NY, USA: ACM, 2016, pp. 71–76. [Online]. Available: http://doi.acm.org/10.1145/2910925.2910949

[11] Y. Khmelevsky, G. Hains, and C. Li, “Automatic code generation within student’s software engineering projects,” in Proceedings of the Seventeenth Western Canadian Conference on Computing Education, ser. WCCCE ’12. New York, NY, USA: ACM, 2012, pp. 29–33. [Online]. Available: http://doi.acm.org/10.1145/2247569.2247578

[12] Y. Khmelevsky and V. Voytenko, “Strategies for teaching mobile application development,” 2013. [Online]. Available: http://wccce2013.ca/wp-content/uploads/2013/05/wccce2013_submission_8.pdf

[13] T. Alstad, R. Dunkin, R. Bartlett, A. Needham, G. Hains, and Y. Khmelevsky, “Minecraft computer game simulation and network performance analysis,” in In Second International Conferences on Computer Graphics, Visualization, Computer Vision, and Game Technology (VisioGame 2014), November 2014.

[14] Y. Khmelevsky and V. Voytenko, “Hybrid cloud computing infrastructure in academia,” in WCCCE 2015 - the 20th Western Canadian Conference on Computing Education, At May 8-9, 2015. Vancouver Island University (VIU), Nanaimo, British Columbia, Canada., 2015.

[15] N. Feamster and A. Gray, “Can great research be taught?: Independent research with cross-disciplinary thinking and broader impact,” SIGCSE Bull., vol. 40, no. 1, pp. 471–475, Mar. 2008. [Online]. Available: http://doi.acm.org/10.1145/1352232.1352294

[16] L. Barker, “Student and faculty perceptions of undergraduate research experiences in computing,” Trans. Comput. Educ., vol. 9, no. 1, pp. 5:1–5:28, Mar. 2009. [Online]. Available: http://doi.acm.org/10.1145/1513593.1513598

[17] T. Alstad, J. Riley Dunkin, T. Delior, B. French, H. Caswell, Z. Ouimet, Y. Khmelevsky, and G. Hains, “Game network traffic simulation by a custom bot,” in Systems Conference (SysCon), 2015 9th Annual IEEE International, April 2015, pp. 675–680.

[18] Y. Khmelevsky, M. Rinard, and S. Sidiroglou-Douskos, “A source-to-source transformation tool for error fixing,” in Proceedings of the 13th Conference of the Center for Advanced Studies on Collaborative Research, ser. CASCON ’13. Riverton, NJ, USA: IBM Corp., 2013, pp. 147–160. [Online]. Available: http://dl.acm.org/citation.cfm?id=2555523.2555540

[19] G. Hains, C. Li, D. Atkinson, J. Redly, N. Wilkinson, and Y. Khmelevsky, “Code generation and parallel code execution from business UML models: A case study for an algorithmic trading system,” in Science and Information Conference (SAI), 2015, pp. 84–93.

[20] Z. Ouimet, H. Caswell, Y. Khmelevsky, R. Bartlett, and A. Needham, “Game servers deployment automation case study,” in 2016 Annual IEEE Systems Conference (SysCon). IEEE, 2016, pp. 1–7.

[21] N. Mcdonald, D. Atkinson, C. Frank, Y. Khmelevsky, and S. McMillan, “Biometric data encryption and encryption for sport wearable devices (a case study),” in 2015 Annual IEEE Systems Conference (SysCon). IEEE, 2016, pp. 1–6.

[22] N. Mcdonald, D. Atkinson, Y. Khmelevsky, and S. McMillan, “Sport wearable biometric data encrypted emulation and storage in cloud,” in 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2016, pp. 1–4.
