codee

Shift Left Performance with Codee

Enabling Faster Drug Discovery using Machine Learning

Executive Summary

"*Shift left*" is a software development approach that moves the emphasis towards the earlier stages of the development life cycle. We demonstrate the use of the Codee tool to shift left performance on a drug discovery application. In a 1.5 hour session the tool identified 8 actions, of which we applied 6, resulting in a 2x speedup of the code.

The Challenge: Fast Virtual Molecule Screening

Virtual Molecule Screening (VMS) is a computational technique used in drug discovery that uses machine learning to predict if a chemical compound is likely to bind to a drug target. It evaluates millions of molecules in so-called virtual screens in order to identify candidate molecules that affect proteins associated with diseases. IMEC's Exascience Lab has developed the SMURFF software, which provides high-quality models to process industry-scale datasets of many millions of molecules. Speeding up the VMS process on modern low-power hardware platforms requires going through a time-consuming manual code inspection process focused on performance.

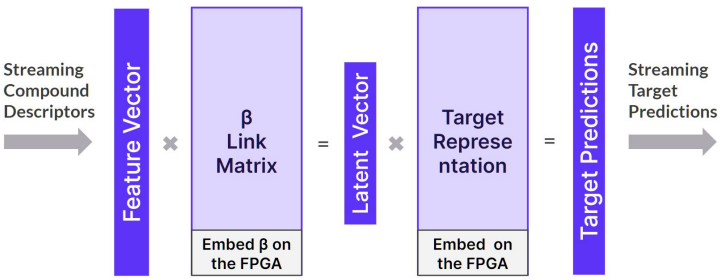

Fig .1. Virtual Molecule Screening (VMS) Pipeline

Shifting Left Performance with Codee

"*Shift left*" is a software development approach that moves the emphasis towards the earlier stages of the development life cycle. It is important because it enables automated code inspection to identify, debug and fix issues while the code is being developed, reducing the technical debt and the likelihood of delivering software products with hidden code errors. Shift-left has been successfully applied to bug-catching, coding standard enforcement or security. For the first time, Codee makes it possible to "shift left performance". This use case shows clear evidences of the benefits of a static code analyzer specifically designed to boost the performance of time-critical software running on modern low-power hardware.

> *"I really like this new tool because it takes zero time to complete the analysis of the application code, in contrast to hours or days of manual code inspection"*
>
> *" The tools points you into directions you would not think of to increase performance. Definitely, it is worth to give it a try!"*

Tom Vander Aa
IMEC, Exascience Lab, Belgium

Shifting Left Performance of Virtual Molecule Screening

A time-limited live technical session was conducted to assist the developer of a VMS application code, assuming no previous knowledge about the usage of Codee's performance optimization report. The experimental setup was as follows:

- **Time:** dedicated technical session of 1.5 hours (90 minutes).
- **Developer:** the owner of the VMS code, with performance optimization expertise in the scope of the hotspot loops for processing compound descriptors and producing target predictions`, as shown in Figure 1.
- **Platform:** Linux Ubuntu 20.04 with gcc 9.3.0 and clang 12.0.1 on an Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz (Haswell with avx2 and fma).

The developer went through the stages of the Codee technical evaluation process, with the following outcomes:

- **Discover:** produced the Codee's performance optimization report of the VMS code identifying 8 actions in the hotspots loops.
- **Adopt:** explored the 6 actions related to sequential performance optimizations and vectorization, deciding to implement the code changes corresponding to PWR022 (remove loop invariant conditions of the loop body) and OPP002 (explicit loop vectorization).

The experimental results obtained with the optimized VMS code are as follows:

- **Speedup** of the VMS code is ~2x using GCC and CLANG.
- **Code size** increase is +6 lines of code in total.
- **Effort** required by the developer is ~2 hours, including the 1.5 hours session and additional 0.5 hours to finish the work.

Overall, this use case shows clear evidences of the benefits of the "shift left performance" approach in the software development process, making it accessible to a wider community of developers.

	GCC	CLANG
PWR022	+5 lines of code	
OPP002	+1 lines of code	
Initial time	12.76 secs	8.62 secs
Fastest time	6.58 secs	4.30 secs
Speedup	1.94	2.00

About IMEC

Interuniversity Microelectronics Centre is an international research & development organization, active in the fields of nanoelectronics and digital technologies. It employs around 4,000 researchers from more than 90 countries and has facilities dedicated to semiconductor processing. This deep-tech knowhow is combined with software and system knowledge to build up advanced technologies that open the door to smart, sustainable solutions in domains such as healthcare, clean energy and Industry 4.0.

About EPEEC

The EPEEC project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement Nº 801051.