

Contribution ID: 624 Contribution code: **contribution ID 624**Type: **Oral**

Managing Heterogeneous Device Memory using C++17 Memory Resources

Monday, November 29, 2021 6:40 PM (20 minutes)

Programmers using the C++ programming language are increasingly taught to manage memory implicitly through containers provided by the C++ standard library. However, many heterogeneous programming platforms require explicit allocation and deallocation of memory, which is often discouraged in “best practice” C++ programming, and this discrepancy in memory management strategies can be daunting for C++ developers who are not already familiar with heterogeneous programming. The C++17 standard introduces the concept of **memory resources**, which allows the user to control how standard library containers allocate memory. By default, memory resources only support host memory, but their use can be extended to include device memory to relieve the aforementioned issue of ergonomics. In this talk, we present **vecmem**, a library of memory resources for heterogeneous computing, which allows efficient and user-friendly allocation of memory on CUDA, HIP, and SYCL devices through standard C++ containers. In addition, it provides a number of STL-like classes, which present a familiar programming interface in device code for accessing, and even modifying, the contents of various types of containers. We investigate the design and use cases of such a library, the potential performance gains over naive memory allocation, and the limitations of this memory allocation model.

Significance

The memory management described in the talk is being developed for one of the ATLAS Experiment’s GPU demonstrator projects, meant to define how code for managing memory on accelerators would be handled in the future. The technique described in the talk is not used in HEP code anywhere yet, as far as we can tell.

This would be the first conference where we can show first results from this R&D project.

References

Speaker time zone

Compatible with Europe

Primary authors: KRASZNAHORKAY, Attila (CERN); GESSINGER, Paul (CERN); SWATMAN, Stephen Nicholas (University of Amsterdam (NL))

Presenter: SWATMAN, Stephen Nicholas (University of Amsterdam (NL))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research