

Practical 6: odds and ends

The main objectives in this practical are to learn about:

- how to have a main code compiled with g++ and use CUDA only for routines called by it
- how to build a library file
- how to use C++ templates

What you are to do is as follows:

1. Click on the link in the course webpage to the Google Colab notebook.
2. Carefully follow the instructions in the notebook.
3. `prac6` and `prac6a` are both based on the same two source files `main.cpp` and `prac6.cu`. `main.cpp` has the main code and is compiled by the default C++ compiler which is g++.

`prac6.cu` has the CUDA routines and is compiled by `nvcc`. The difference between the two executables is that in the first case `prac6.cu` is compiled and the resulting object file is linked to `main.o` to create `prac6`, whereas in the second case `prac6.cu` is compiled into a library, and this is later linked to `main.o`.

4. `prac6b` is generated by the modified code `prac6b.cu` which uses C++ templates to generate two versions of the kernel code, one for floats and one for ints.

Study how they are used and ask questions if anything is not clear. Run the code to see the output it produces.

Modify `prac6b.cu` to use a third version of the kernel routine for double precision variables.

5. `prac6c` is generated by the code `prac6c.cu`. This also uses C++ templates to generate two versions of a different kernel code with different sizes for a fixed size array which will be mapped to registers by the compiler.

Study how they are used and ask questions if anything is not clear. Run the code to see the output it produces. You can check that the first thread (`tid=0`) computes the correct value.

This is a useful technique to generate multiple “instances” (or “instantiations”) of a kernel as an alternative to dynamically-sized arrays which can only be put in shared memory.

6. Look at the documentation for the NVCC compiler available at http://docs.nvidia.com/cuda/pdf/CUDA_Compiler_Driver_NVCC.pdf