Rainer Spurzem, spurzem@ari.uni-heidelberg.de

Introduction to GPU Accelerated Computing

Mon-Fri, Feb. 17-21, 2025, Uni Heidelberg, GPU Block Course

Tutors:

Vahid Amiri vahid.amiri@uni-heidelberg.de
Yash Mohan Sharma sharma@mpia.de
Navonil Saha navonil.saha@uni-heidelberg.de
Philip Cho philip.cho@stud.uni-heidelberg.de

Location:

Lecture: hybrid; PC Pool KIP INF227 1.401 and via zoom

Tutorials: PC Pool KIP INF 227 1.401 (Group 1); Seminar Room KIP INF227 2.402 (Group 2) and 2.403 (Group 3); online zoom (Group 4)

START Lecture: 10:15 a.m., Monday Feb. 17, 2025
(PC Pool KIP INF227 1.401)

1. zoom and chat access, general informations

Both Lecture (Hybrid: in person and online);
and online tutorial for Group 4 (Philip Cho):
https://eu02web.zoom-x.de/j/61046677162?wd=uyzpx921tESbQMy4YQpuPVcZ10yA13.1
Meeting ID: 610 4667 7162 ; Passcode: 363102

Time in lecture: Mon-Fri 10:15 – 13:00
Lunch Break: 13:00 – 14:15
Tutorials: 14:15 – 17:00 **(no afternoon session on Friday!)**
Home Work: Start on Thursday, Feb. 20, until March 2.
(can be done from home computer or in PC Pool Rooms)

There are rocket chats open:
General course chat:
https://uebungen.physik.uni-heidelberg.de/chat/group/WS24-GPU-Course
Also for each group:
https://uebungen.physik.uni-heidelberg.de/chat/group/WS24-GPU-G01

https://uebungen.physik.uni-heidelberg.de/chat/group/WS24-GPU-G02

https://uebungen.physik.uni-heidelberg.de/chat/group/WS24-GPU-G03

https://uebungen.physik.uni-heidelberg.de/chat/group/WS24-GPU-G04

```
(Note: this block course takes place in 2025, but it still belongs to
the winter semester 2024, therefore it is labelled with WS24)
```

Rocket Chat: Messages independent of zoom! Messages are permanent! In zoom: raise hand (preferred) and chat, but not permanent! Any questions raised in the general course rocket chat will be answered there, remain there, and/or discussed in the lecture.

Need to know: This course is not graded!

Certificate of successful participation only! (2 ECTS points)

Necessary condition for certificate:

Regular participation in course; missing one session (e.g. due to exam) will be no problem, all material comes online same day, but please tell your tutor. But all Mon-Wed exercises must be finally done on the GPU supercomputer (do not delete output files)! The last exercise is a homework, and will be prepared on Thursday. You have about one week time to turn it in. More time can be given on request if needed, just ask.

This is NOT a programming course ! 1:

You should have some knowledge of a higher level programming language (like python, c, c++, fortran, …). We will learn special CUDA extensions of the c-language for programming GPUs (graphical processing unit). This is called GPGPU – general purpose GPU programming.

This is NOT a programming course! 2:

You learn CUDA, but our learning here is by using examples, running them, explaining the background in the lecture, and only make small changes to the program, if any. The course is too short to teach you writing CUDA codes from scratch and many features of CUDA will not be covered.

Data Handling!

For the homework you produce some simple data, and the task is to plot them. You can use anything you like (jupyter notebook, gnuplot, idl, …); we try to help – but again – this is not part of the course.

2. Getting access to the bwUniCluster GPU computer

There are three steps to do, and a necessary condition is to have a valid uniID:

https://wiki.bwhpc.de/e/Registration/bwUniCluster
Contains general information about all steps. The registration process for the bwUniCluster is divided into three steps (links for all steps are given in the webpage given above; it is a bit complicated, try to read all informations carefully). These steps can be done before the course starts, but do not worry in case of problems we have time to solve them on the first day:

→ Step A: bwUniCluster Entitlement: You need to get the bwUniCluster Entitlement from your university/college, select me as your supervising professor, and I have to approve your entitlement via some online procedure. NOTICE: It may take about a day until your entitlement is properly registered for the next step. ***Please choose March 31, 2025 for the end of the entitlement, to have a reserve time for finishing the homework.***

→ Step B: You need to register for the service bwUniCluster (create an account on the cluster). For this you need to set a "service password" and initialize a two factor authentication (2FA), a cell phone app is needed for this (e.g. google authenticator, but possible other's are given in the web link).

→ Step C: You need to fill out the bwUniCluster questionnaire within 14 days.

In order to login to the bwUniCluster GPU supercomputer it is expected that you will use your own laptop to launch an ssh connection; in the PC pool you can use the terminals, your UniID serves as a login. See some information about the system here: https://wiki.bwhpc.de/e/BwUniCluster2.0
Login with:

```
ssh -Y hd_uniID@bwunicluster.scc.kit.edu
```

(where uniID has to be substituted by your own uniID). The -Y option is recommended, so you can send graphical windows back to your laptop. It will first ask for a one-time password (OTP) from your 2FA authentication (typically using a cell phone app) and thereafter for your service password (see Step B).

For Windows Users:

- you can use a terminal window, which works like a Linux command window, compare https://www.wikihow.com/Use-SSH
- Second Way: Use the putty client program for ssh; https://putty.org/
- Third Way – open a Linux Shell on Windows or use a WLS installation (ubuntu on Windows) – then you can follow the Linux informations above.

3. After login to the bwUniCluster GPU computer

```
ls -lta
```

shows your files in your home directory. The file named .bashrc should be edited, add three lines at the end:

```
module load compiler/gnu/10.2
module load devel/cuda/12.4
module load mpi/openmpi/default
```

To make the changes in .bashrc valid, you can logout and login again, or give a source .bashrc ; Thereafter you can copy the course files from the home directory of R. Spurzem:

```
cp -p -i ~hd_un119/gpu-course-master.tgz .
cp -p -i ~hd_un119/gworknb6.tar.gz .
```

(the 2nd file is not yet there, comes later for homework preparation on Thursday)

```
tar xvfz gpu-course-master.tgz
```

creates a subdirectory gpu-course in your home directory:

```
cd gpu-course ; ls -ltr
```

shows the course exercises, in several subdirectories (0_hello, 1_add, …), with:

```
module list
```

you can check that you have correctly edited your .bashrc file (see on top). The command should output a confirmation that the three modules listed on top are in fact loaded. They provide the necessary software for using the NVIDIA nvcc GPU compiler and MPI parallel computing.

4. How to run an exercise

Example of hello world program:

```
cd hello
nvcc -o hello.out hello.c
./hello.out
nvcc -o hello.gpu hello.cu
./hello.out
nvcc -o device.gpu device.cu
./device.out
```

(for those who like it: there is also a makefile). The -o option allows you to choose a name for your executable file (if not given Unix default name a.out). The first cuda program (hello.cu) produces output, saying there are no GPU devices. This is because the bwUniCluster computer has a slurm batch system, the login node has no GPU. Only worker nodes have the GPUs. We will use a shell script (gpu_script.sh) to submit batch jobs to the worker nodes:

```
sbatch gpu_script.sh
```

Information about the job and queue status:

```
squeue
sinfo_t_idle
```

The first one shows your own jobs in the system (if there are); the second one shows which partitions (queues) have idle nodes. We have a reservation for the course in the gpu_4 partition (reservation=ws-hd). Both is coded into the header of the gpu-script.sh file, so you do not need to worry about it. After some time the results of the batch job will appear on job.nnnnnnnn.out and job.nnnnnnnn.err (8 digit number nn… is a unique identifier for your job); look for it with

```
ls -ltr
```

For reading and editing files you can use the Linux programs:

```
cat  ; more ; vi ; nano ; pico ; …?
```

Transfer data files to your own computer (example for one job output):

```
scp -p hd_uniID@bwunicluster.scc.kit.edu:gpu-course/0_hello/job.nnnnnnnn.out .
```

(Note the . at the end; again uniID is your uiID, and nn… is an 8 digit individual number for every, job which you will find with the ls -ltr command given above)