|epcc|

THE UNIVERSITY OF EDINBURGH

Domain Decomposition: Computational Fluid Dynamics

May 24, 2015

1 Introduction and Aims

This exercise takes an example from one of the most common applications of HPC resources: Fluid Dynamics. We will look at how a simple fluid dynamics problem can be run on a system like ARCHER and how varying the number of processes it runs on and the problem size affect the performance of the code. This will require a set of simulations to be run and performance metrics plotted as a graph.

The CFD program differs from the more straightforward task farm in that the problem requires more than source-worker-sink communications. Here the workers are in regular communication throughout the calculation.

This exercise aims to introduce:

- Grids
- Communications – Halos
- Performance metrics

2 Fluid Dynamics

Fluid Dynamics is the study of the mechanics of fluid flow, liquids and gases in motion. This can encompass aero- and hydro- dynamics. It has wide ranging applications from vessel and structure design to weather and traffic modelling. Simulating and solving fluid dynamic problems requires large computational resources.

Fluid dynamics is an example of continuous system which can be described by Partial Differential Equations. For a computer to simulate these systems, the equations must be discretised onto a grid. If this grid is regular, then a finite difference approach can be used. Using this method means that the value at any point in the grid is updated using some combination of the neighbouring points.

Discretisation is the process of approximating a continuous (i.e. infinite-dimensional) problem by a finite-dimensional problem suitable for a computer. This is often accomplished by putting the calculations into a grid or similar construct.

2.1 The Problem

In this exercise the finite difference approach is used to determine the flow pattern of a fluid in a cavity. For simplicity, the liquid is assumed to have zero viscosity which implies that there can be no vortices (i.e. no whirlpools) in the flow. The cavity is a square box with an inlet on one side and an outlet on another as shown below.

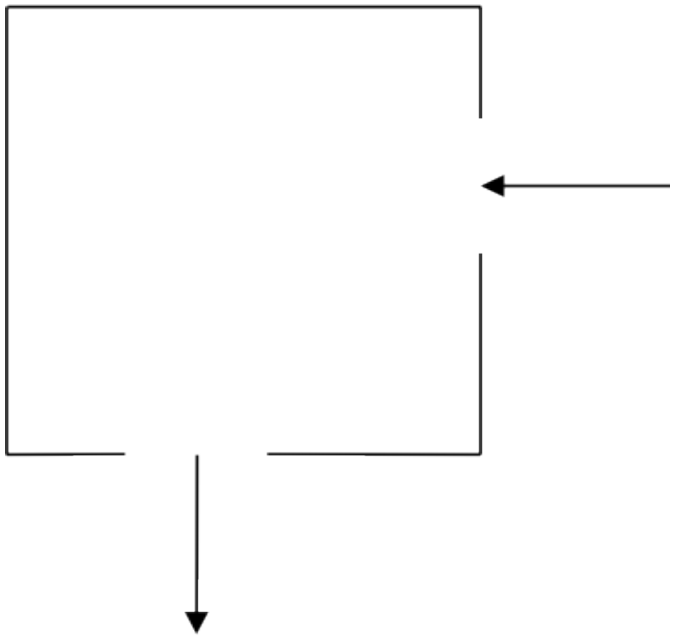

Figure 1: The Cavity

2.2 A bit of Maths

In two dimensions it is easiest to work with the *stream function* Ψ (see below for how this relates to the fluid velocity). For zero viscosity Ψ satisfies the following equation:

$$\nabla^2\Psi = \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} = 0$$

The finite difference version of this equation is:

$$\Psi_{i-1,j} + \Psi_{i+1,j} + \Psi_{i,j-1} + \Psi_{i,j+1} - 4\Psi_{i,j} = 0$$

With the boundary values fixed, the stream function can be calculated for each point in the grid by averaging the value at that point with its four nearest neighbours. The process continues until the algorithm converges on a solution which stays unchanged by the averaging process. This simple approach to solving a PDE is called the Jacobi Algorithm.

In order to obtain the flow pattern of the fluid in the cavity we want to compute the velocity field $\tilde{u}$. The x and y components of $\tilde{u}$ are related to the stream function by

$$u_x = \frac{\partial\Psi}{\partial y} = \frac{1}{2}(\Psi_{i,j+1} - \Psi_{i,j-1})$$

$$u_y = -\frac{\partial\Psi}{\partial x} = -\frac{1}{2}(\Psi_{i+1,j} - \Psi_{i-1,j})$$

This means that the velocity of the fluid at each grid point can also be calculated from the surrounding grid points.

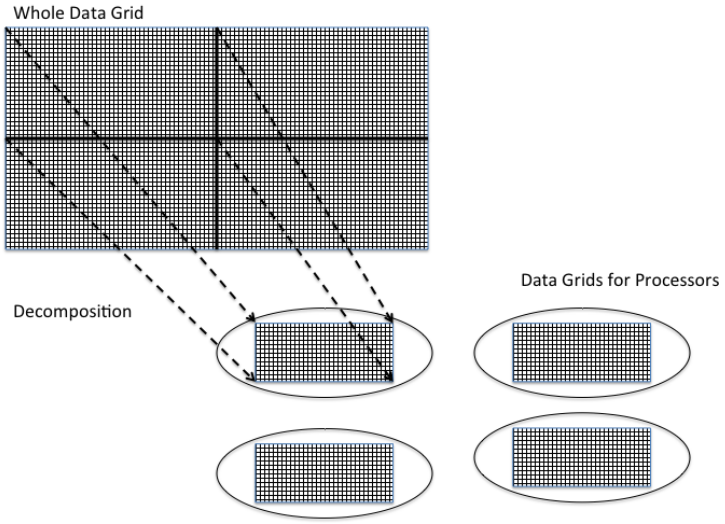

Figure 2: Breaking Up the Big Problem

2.3 An Algorithm

The outline of the algorithm for calculating the velocities is as follows:

```
Set the boundary values for Ψ and ũ
while (convergence= FALSE) do
   for each interior grid point do
      update value of Ψ by averaging with its 4 nearest neighbours
   end do
   check for convergence
end do
for each interior grid point do
   calculate u_x
   calculate u_y
end do
```

For simplicity, here we simply run the calculation for a fixed number of iterations; a real simulation would continue until some chosen accuracy was achieved.

2.4 Broken Up

The calculation of the velocity of the fluid as it flows through the cavity proceeds in two stages:

- Calculate the stream function Ψ.
- Use this to calculate the x and y components of the velocity.

Both of these stages involve calculating the value at each grid point by combining it with the value of its four nearest neighbours. Thus the same amount of work is involved to calculate each grid point, making it ideal for the regular domain decomposition approach. Figure 2 shows how a two dimension grid can be broken up into smaller grids for individual processes. This is usually known as **Decomposition**.

This process can hold for multiple other cases, where slices or sections of grids are sent to individual processes and the results can be collated at the end of a calculation cycle.

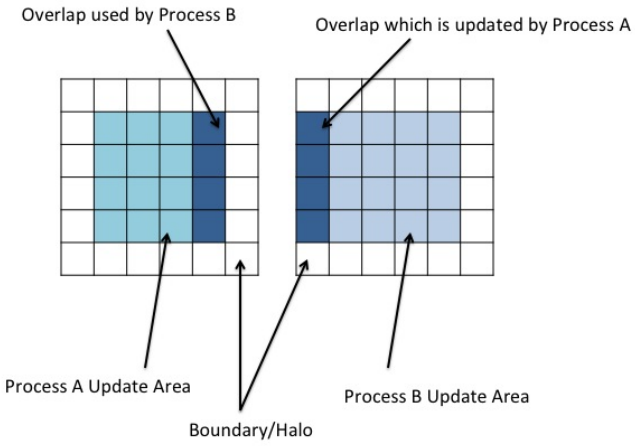

Figure 3: Halo: Process A and Process B

2.5 Halos

Splitting up the big grid into smaller grids introduces the need for interprocess communications. Looking at how each point is being calculated, how does the system deal with points on the edge of a grid? The data a process needs has been shipped off to a different process.

To counter this issue, each grid for the different processes has to have a boundary layer on its adjoining sides. This layer is not updated by the local process: it is updated by another process which in turn will have a boundary updated by the local process. These layers are generally known as *halos*. An example of this is show in Figure 3.

In order to keep the halos up to date, a halo swap must be carried out. When an element in process B which adjoins the boundary layer with process A is updated and process A has been updating, the halo must be swapped to ensure process B uses accurate data. This means that a communication between processes must take place in order to swap the boundary data. This halo swap introduces communications that if the grid is split into too many processes or the size of data transfers is very large, the communications can begin to dominate the runtime over actual processing work. Part of this exercise is to look at how the number of processes affects the run-time for given problem sizes and evaluate what this means for speed up and efficiency.

2.6 One-dimensional Domain decomposition for CFD example

For simplicity, we only decompose the problem in one dimension: the y-dimension. This means that the problem is sliced up into a series of rectangular strips. Although for a real problem the domain would probably be split up in both dimensions as in Figure 2, splitting across a single dimension makes the coding significantly easier. Each process only needs to communicate with a maximum of two neighbours, swapping halo data up and down.

3 Exercises

3.1 Compilation

Use `wget` to copy the file `cfd.tar.gz` from the ARCHER web pages to `/work/` on ARCHER or your home directory on Morar, as you did for the previous exercises. Now unpack the file and compile the Fortran MPI CFD code as follows: after compilation, an executable file will have been created called **cfd**.

```
guestXX@archer:~> tar -zxvf cfd.tar.gz
cfd/
cfd/cfd.pbs
cfd/Makefile_archer
cfd/Makefile_morar
cfd/cfd.f90

guestXX@archer:~> cd cfd
```

To compile on **ARCHER**:

```
guestXX@archer:~/cfd> make -f Makefile_archer
ftn -g -c cfd.f90
ftn -g -o cfd cfd.o
```

To compile on **Morar**:

```
-bash-4.1$ module load mpich2-pgi
-bash-4.1$ make -f Makefile_morar
mpif90 -g -c cfd.f90
mpif90 -g -o cfd cfd.o
```

3.2 Run the program

ARCHER

Use emacs or your preferred editor to look at the `cfd.pbs` batch script:

```
#!/bin/bash --login

#PBS -l select=1
#PBS -l walltime=00:05:00
#PBS -A y14
#PBS -N cfd

#Change to directory that the job was submitted from
cd $PBS_O_WORKDIR

aprun -n 4 ./cfd 4 5000
```

The arguments to aprun have the following meaning:

- **-n 4**: run the code on 4 processes;
- **./cfd**: use the `cfd` executable in the current directory;

- **4 5000**: arguments to the cfd code to use a scale factor of 4, and run for 5000 iterations.

Morar

Use emacs or your preferred editor to look at the `cfd.sge` batch script.

You can modify `SCALE` and `ITERATIONS` to control the behaviour of the code:

```
#!/bin/bash
#$ -cwd -V
#  This is a template submission script for MPI batch jobs on morar

MPIEXE=`basename $REQUEST .sge`
SCALE=4
ITERATIONS=5000

echo "------------------------------------------------------------"
echo "Running MPI program <$MPIEXE> on" $NSLOTS "processes"
echo "------------------------------------------------------------"
echo

(time mpiexec -n $NSLOTS ./$MPIEXE $SCALE $ITERATIONS) 2>&1

echo
echo "-------------------------"
echo "Finished MPI program"
echo "-------------------------"
echo
```

To launch the job (for example on four processors) use:

```
bash-4.1$ qsub -pe mpi 4 cfd
```

The arguments you can modify have the following meaning:

- **SCALE=4**: argument for running the cfd code: use a scale factor of 4;
- **ITERATIONS=5000**: argument for running the cfd code: run for 5000 iterations.

Modifying the problem size

The mimimum problem size (scale factor = 1) is taken as a 32×32 grid. The actual problem size can be chosen by scaling this basic size, for example with a scale factor of 4 then it will use a 128×128 grid. Varying the number of processes and scale factor allows us to investigate Amdahl's and Gustafson's laws. The number of iterations is not particularly important as we are interested in the time per iteration. You can increase the number of iterations to ensure that the code does not run too fast on large numbers of processes, or decrease it so it is not too slow for large problem sizes. Two things to note:

- The code assumes the problem size decomposes exactly onto the process grid. If this is not the case (e.g. scale factor = 2 with 7 processes, since 7 is not a divisor of 64) it will complain and exit.
- If the output picture looks strange then you may not have used a sufficient number of iterations to converge to the solution. This is not a problem in terms of the performance figures, but it is worth running with more iterations just to check that the code is functioning correctly

Once you have run the job via the PBS batch system, the output file should look something like this (depending on the exact parameters):

```
Scale factor =   4, number of iterations =   5000
Running CFD on  128 x  128 grid using    4 processes

Starting main loop ...

completed iteration  1000
completed iteration  2000
completed iteration  3000
completed iteration  4000
completed iteration  5000

... finished

Time for   5000 iterations was  0.1763     seconds
Each individual iteration took  0.3527E-04 seconds

Writing output file ...
 ... finished

CFD completed
```

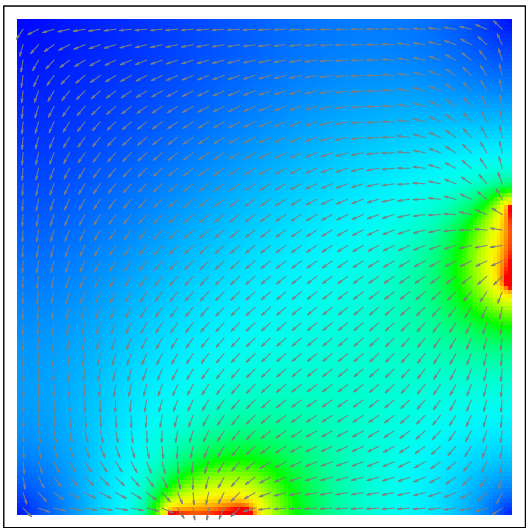

Figure 4: Output Image

The code also produces some graphical output in the form of a gnuplot file `cfd.plt`. You can view this as follows:

```
guestXX@archer:~> gnuplot -persist cfd.plt
```

which should produce a picture similar to Figure 4.

If the fluid is flowing down the right-hand edge then along the bottomm, rather than through the middle of the cavity, then this is an indication that the Jacobi algorithm has not yet converged. Convergence requires more iterations on larger problem sizes.

3.3 Performance Evaluation

The next part of this exercise will be to determine what the best configuration for a group of problems sizes in the CFD code would be. This will be worked out using two measures: speed-up and efficiency.

3.3.1 Speed-Up

The speedup of a parallel code is how much faster the parallel version runs compared to a non-parallel version. Taking the time to run the code on 1 process is T_1 and to run the code on P processes is T_P, the speedup S is found by:

$$S = \frac{T_1}{T_P} \tag{1}$$

3.3.2 Efficiency

Efficiency is how the well the esources (available processing power in this case) are being used. This can be thought of as the speed-up (or slow-down) per process. Efficiency E can be defined as:

$$E = \frac{S}{P} = \frac{T_1}{PT_P} \tag{2}$$

where $E = 1.0$ means 100% efficiency, i.e. perfect scaling.

3.3.3 Doing The Work

The two main evaluation points:

- How do the speed-up and efficiency of the program vary as the number of processes is increased?
- Does this change as the problem size is varied?

To investigate the speedup and parallel efficiency the code should be run using the same problem size but with varying numbers of processes. Calculate the speedup and efficiency (tables are provided overleaf for this) and plot a graph of the speedup against the number of processes. Is there any apparent pattern, e.g. does it follow Amdahl'slaw?

Now choose a different problem size and repeat the exercise. To increase the problem size increase the scale factor; to decrease the size, decrease the scale factor. For example, setting **scale factor** = 2 will give a problem size of 64x64; **scale factor** = 6 gives a size of 192x192.

What is the effect of problem size the parallel scaling of the code?

1. problem size (scalefactor) = _____ iterations =_____

No of processes	Time per iteration	Speed-up	Efficiency
1			
2			
4			
8			
16			
32			
64			
128			

2. problem size (scalefactor) = _____ iterations =_____

No of processes	Time per iteration	Speed-up	Efficiency
1			
2			
4			
8			
16			
32			
64			
128			

3. problem size (scalefactor) = _____ iterations =_____

No of processes	Time per iteration	Speed-up	Efficiency
1			
2			
4			
8			
16			
32			
64			
128			

4. problem size (scalefactor) = _____ iterations =_____

No of processes	Time per iteration	Speed-up	Efficiency
1			
2			
4			
8			
16			
32			
64			
128			

4 Compiler Investigation

We will use the CFD example to investigate how using different compilers and compiler options affects performance.

4.1 Changing compilers on ARCHER

On ARCHER, the Fortran compiler is always called `ftn`. However, what compiler this actually points to is determined by what module you have loaded.

For example, to switch from the default (Cray) compiler to the Intel compiler.

```
guestXX@archer:~> module switch PrgEnv-cray PrgEnv-intel
guestXX@archer:~> make clean
guestXX@archer:~> make
```

Here. `make clean` ensures that all compiled code is removed which means that the new code will be built with the new compiler. The GNU compiler module is called `PrgEnv-gnu`.

4.2 Exercises

Here are a number of suggestions:

- By default, the code is built with the `-g` debugging option. Edit the Makefile to remove this and recompile - what is the effect on performance?
- What is the difference between the performance of the code using the three different compilers (Cray, Intel and GNU) with no compiler options?
- It is not really fair to compare compiler performance using default options: one compiler may simply have higher default settings than another. Using the option suggested in "Useful compiler optipons" at http://www.archer.ac.uk/documentation/user-guide/development.php, compare the best performance you can get with each compiler.