



# ARCHER

## Software Carpentry Workshop

Instructors:

Emmanouil (Manos) Farsarakis (EPCC)  
Kevin Stratford (EPCC)

Helpers:

Matthew Withers (Portsmouth)  
Dida Markovic (Portsmouth)  
Daniel Goddard (Portsmouth)  
Kyle Westfall (Portsmouth)

University of Portsmouth, December 2015

# Reusing this material



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

[http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en\\_US](http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en_US)

This means you are free to copy and redistribute the material and adapt and build on the material under the following terms: You must give appropriate credit, provide a link to the license and indicate if changes were made. If you adapt or build on the material you must distribute your work under the same license as the original.

Note that this presentation contains images owned by others. Please seek their permission before reusing these images.

# Who am I?

Hi! I'm Manos! :)



# Who sent me?

A large blue banner features a woman's face in profile on the left, a map of Europe in the center, and the word "PRACE" surrounded by stars on the right. Overlaid on the map is the text "PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE".

EPCC's PRACE Advanced Training Centre



# Six PATCs

**Hubs for world-class HPC training  
for researchers in Europe**

Barcelona Supercomputing Center (Spain)

CINECA - Consorzio Interuniversitario (Italy)

CSC - IT Center for Science Ltd (Finland)

EPCC at the University of Edinburgh (UK)

Gauss Centre for Supercomputing (Germany)

Maison de la Simulation (France)



PARTNERSHIP  
FOR ADVANCED COMPUTING  
IN EUROPE



<http://www.training.prace-ri.eu>





|epcc|



<http://www.archer.ac.uk>  
[support@archer.ac.uk](mailto:support@archer.ac.uk)



# ARCHER

## in a nutshell...

- › UK National Supercomputing Service
  - › Replacement for HECToR
- › Cray XC30 Hardware
  - › Nodes based on 2 x Intel Ivy Bridge 12-core processors
  - › 64GB (or 128GB) memory per node
  - › 4920 compute nodes (118,080 cores)
  - › Linked by Cray Aries interconnect (dragonfly topology)
- › Cray Application Development Environment
  - › Cray, Intel, GNU compilers
  - › Cray Parallel Libraries (MPI, SHMEM, PGAS)
  - › DDT Debugger, Cray Performance Analysis Tools

# ARCHER

## Performance

- › **HECToR #50 in top 500 with 830TFLOP/s**
- › **ARCHER**
  - › Designed to provide 3-4 times scientific throughput of HECToR
  - › #25 in November 2014 top 500 with 1.65PFLOP/s
- › **Extract the best performance possible from the hardware**

# Optimisation is risky business

(without necessarily achieving it!)

“More computing sins are committed in the name of efficiency than for any other single reason...  
...Including blind stupidity.” (Wulf)



# Optimisation is risky business

**“We should forget about small efficiencies, say about 97% of the time:**

**...Premature optimisation is the root of all evil.” (Knuth 1974)**

# Optimisation

What is preferable?

› An **optimised** code that produces  
**incorrect** results?

or...

› A **non-optimised** code that produces  
**correct** results?

Which is the most efficient use of  
resources?

# Optimisation

- > How do we prevent ourselves from introducing bugs when we're optimising and parallelising?
- > Isn't our time more valuable than a computer's time?
- > **Hardware life < Software life < Social life**

# Three rules of optimisation:

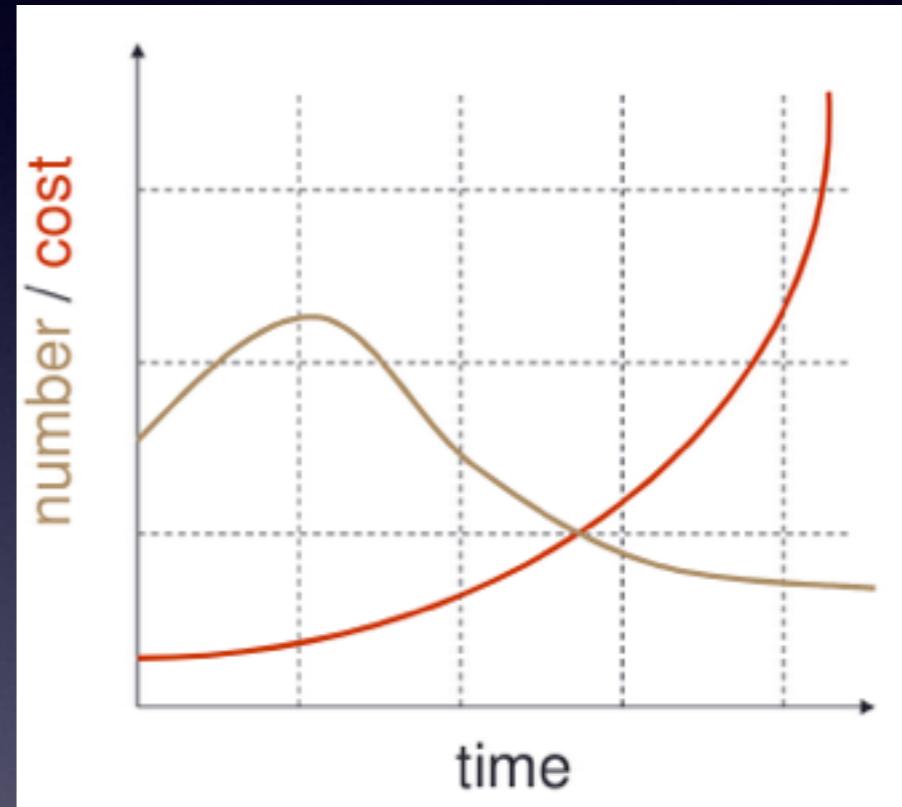
- > Optimised code is not readable or maintainable code
- > Designing code is not the same as optimising code
- > Profile first - anything else is just a guess!

# Optimise ourselves

50-80% of errors in  
15-20% of modules  
(Davis 1995 quoting Endres  
1975, Weinberg 1992)

50-80% of modules are  
error free (Boehm and Basili  
2001)

Technical debt  
1-10-100 rule



# design errors >> # coding errors  
(Boehm et al. 1975)

# What is this course all about?

Good programming practice

# Leaving here you will have learned to:

- Write robust code
- Minimize repetitive tasks
- Appreciate the value of “sharable” code
- Share your work
- Protect your work
- Collaborate on other people’s work

# Programming language or programmer?

“The number of lines of code a programmer can write in a fixed period of time is the same independent of the language used.” - Corbato’s Law

Quantity

“Regardless of whether one is dealing with assembly language or compiler language, the number of debugged lines of source code per day is about the same!” (Corbato 1969)

Correctness

“Performance variability that derives from differences among programmers of the same language ... is on average as large or larger than the variability found among the different languages.” (Prechelt 2000)

Performance



epcc



software carpentry

# What about correctness?

“Chang was horrified to discover that a homemade data-analysis program had flipped two columns of data”

The screenshot shows the Wikipedia article for Geoffrey Chang. The page title is "Geoffrey Chang". Below the title, it says "From Wikipedia, the free encyclopedia". The main text discusses his work at the University of California, San Diego's Skaggs School of Pharmacy and Pharmaceutical Sciences and Department of Pharmacology. It mentions his research on the structural biology of integral membrane proteins, particularly exploring X-ray crystallography techniques. The text notes that Chang and coauthors retracted five previously published papers describing the structures of three multidrug transporter proteins after another research group published a widely differing structure, which led to the discovery of a critical bug in the Chang group's custom software tools. The "Retracted papers" section is highlighted with a red box.

Miller, G. “A Scientist's Nightmare: Software Problem Leads to Five Retractions”, Science 314(5807), pp1856- 1857. DOI: 10.1126/science.314.5807.1856

# What about correctness?

Create account Log in

Article [Talk](#) Read Edit View history Search Q

 WIKIPEDIA  
The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikimedia Shop

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export

## Geoffrey Chang

From Wikipedia, the free encyclopedia.

**Geoffrey Chang** is a professor at the [University of California, San Diego's Skaggs School of Pharmacy](#) and [Pharmaceutical Sciences](#) and [Department of Pharmacology](#), School of Medicine. His laboratory focuses on the [structural biology](#) of [integral membrane proteins](#), particularly exploring [X-ray crystallography](#) techniques for solving the [tertiary structures](#) of membrane proteins that are notoriously resistant to [crystallization](#). The laboratory has specialized in structures of [multidrug resistance](#) transporter proteins in [bacteria](#). In 2001, while a faculty member of [The Scripps Research Institute](#), Chang was awarded a [Beckman Young Investigators Award](#), designed to support researchers early in their academic careers, for his work on the structural biology of multidrug resistance.<sup>[1]</sup> Chang announced a move from Scripps to neighboring UC San Diego in 2012.<sup>[2]</sup>

In 2007, Chang and coauthors [retracted](#) five previously published papers describing the structures of three multidrug transporter proteins after another research group published a widely differing structure, which led to the discovery of a critical [bug](#) in the Chang group's custom software tools.<sup>[3]</sup> Since that time, however, Chang has published other papers in the field of structural biology,<sup>[4][5]</sup> and has been awarded a EUREKA grant, "for exceptionally innovative research projects that could have an extraordinarily significant impact on many areas of science," from the National Institutes of Health.<sup>[6]</sup>

**Contents [hide]**

- [1 Retracted papers](#)
- [2 Recent work](#)
- [3 References](#)
- [4 External links](#)

**Retracted papers** [edit]

Chang and coauthors published papers on the structures of multidrug resistance transporters known as [EmrE](#), [MsbA](#), and [NorM](#) between 2001 and 2010. Although the initial structures were widely considered puzzling in the field due to their unexpected placement of their [ATP binding sites](#) in the assembled dimer,<sup>[7]</sup> the publication of an

# Behind every great piece of science...

```
if((!($trait[$x] eq $trait[$y])) && (abs($pos[$x] - $pos[$y]) <= 500000) && (exists($legArrayPos{$pos[$x]})) && (exists($legArrayPos{$pos[$y]})) {
    my $snp1ArrayPos = "";
    my $snp2ArrayPos = "";
    my $snp1All = "";
    my $snp2All = "";

    #create output file for this SNP pair
    my $filename = "ConditionedResults2/$chr[$x].$pos[$x]-$pos[$y].EHH.GBR.2.txt";
    print "$filename\n";
    unless (-e $filename) {
        open(OUT, ">$filename");

        ##### CHANGE THESE IF NOT FOCUSING ON SECOND SNP #####
        my $start = $pos[$y]-500000;
        if ($start < 1) {
            $start = 1;
        }
        my $end = $pos[$y]+500000;
        if ($end > $chrLengths{$chr[$x]}) {
            $end = $chrLengths{$chr[$x]};
        }
    }
}
```

Courtesy of Carole Goble

Take a breath  
and we'll get started shortly...

