

History in Tools Development 1988-2010s

Prof. Dr. Thomas Ludwig Technische Universität München (1988-2001) Universität Heidelberg (2001-2009) Universität Hamburg (2009-today)

© Thomas Ludwig



Tool Development at Technische Universität München

- 1988
 - Realbug debugger for real time applications with hardware monitor
 - Developed by Thomas Bemmerl, start of the tools history
 - Bond-out processors and adapter for processor socket
- 1990s
 - Tool environment with debugger, performance analyzer, program flow visualizer, load balancer
 - For Intel iPSC parallel computer, later for Parsytec Transputer system (commercial)
- 1995 and later
 - Interface definition for interoperable tools (OMIS)
 - OMIS compliant monitoring system (OCM)
 - Now all in software
 - Main actors: Roland Wismüller (Univ. Siegen), Thomas Ludwig



Tool Development at Technische Universität München ...

- Research Ludwig
 - PhD thesis: load balancing via process migration based on on-line performance monitoring
 - Habilitation thesis: OMIS etc.
 - In addition since 1995
 - Focus on storage: emulation libraries for Intel's parallel I/O libraries



Slides were still in portrait mode...





The Tool-Set (1995)



Tool Development at Universität Heidelberg

Develop a tool to understand I/O of MPI programs



Project Goal

Design and implement a tool environment that visualizes three aspects:

- Client I/O activity
- Server I/O activity
- The relation of the two of them

Environment

- MPICH2 for parallel programming
- PVFS2 as a parallel file system



Idea

- Generate a trace for the servers
- Merge the server trace with the client trace
- Add arrows from MPI-IO calls to corresponding server activities
- Visualize everything together in Jumpshot



An Example: MPI_File_write triggers Trove write









David Bailey

Twelve Ways to Fool the Masses When Giving Performance Results on Parallel Computers Supercomputer Review, August 1991

(12) If all else fails, show pretty pictures and animated videos, and don't talk about performance



Tool Development at Universität Hamburg

Tools

- Still focus on I/O
- New focus on energy consumption
- Focus on I/O issues
 - Evaluation
 - Compression
 - Metadata performance
 - •

12



Bitwise Reproducibility in

Computational Climate Science

Prof. Dr. Thomas Ludwig German Climate Computing Center (DKRZ) University of Hamburg, Department for Computer Science (UHH/FBI)

© Thomas Ludwig

[Victoria Stodden: Reproducibility in High Performance Computing **Invited Plenary – SC15]**

\sim	• •	•	· ·	•
l rad				CIC
		ILY		212
10.715	0.416			

Washington's lawyer surplus

 \mathbf{W}

How to do a nuclear deal with Iran

Investment tips from Nobel acanomists Junk bonds are back The meaning of Sachin Tendulicar

Go

人間部

The

Economist

· Bantonia

ø

strengthen being many

L	os Angeles T	imes	BUSINESS	Announcement: Reducing our Irreproducibility : N	ature News & Comment			MC	
science	AL. U.S. WORLD BUSINESS SPORT ence has lost its way, a rchers are rewarded for splashy findings, ists looking for cures to diseases have bee	S ENTERTAINME at a big cost not for double-ch n building on ideas All Science Jour	ENT HEALTH STYLE TRAV st to humanity ecking accuracy. So many s that aren't even true. mais i filler back to accuracy parameters to	Home News & Commant Research Careers & Jobs Cu Audio & Video For Authors Archive Volume 498 Issue 7448 Ectional Archite	Soatch Ge 		GO		
Science The We	NEWS SCIENCE JOURNALS CAREERS MU	LTIMEDIA COLLECTI	ONS	NATURE EDITORIAL	< =	•	Nº 94		
Schence Home Guttent I Hutter > Science Negacite > 1	haua Pravious Issues Science Express Science Pro 12 January 2014 - McNutt, 343 (5168): 229	ducts Wy Science Ab	out the Journal	Announcement: Reducing our 24 April 2013	irreproducibilit	У			
Arsicle Views • Summary • Full Text	Science 17 January 2014: Vol. 343 no. 6168 p. 229 DDI: 10.1126/science.1250475 EDITORIAL		 Prex Table of Contents Next > Read Full Text to Comment (8) 	Over the past year, Nature has published a string of articles the published a string of articles the string of art	natur	e International o	which journal of science Advanced search	Search	G
r Fail Text (PDF) Article Tools	Reproducibility Marcia McNutt			Tensibility and reproductivity of published research (coverties)	archive - volume 483 -	issue 7391 - editor	rials - article		
Save to My Folders Dewnikad Cliation Alert Me When Article Is Cled Post to Cleat.kee E-mail This Page Rights & Permissions Commercial Respires	Marcia Mchutt is Editor-In-Chief of Science. Science advances on a foundation of trusted di approach that scientists use to gain confidence community was shaken by reports that a troub not reproducible. Because confidence in result community, we are announcing new initiatives Science. For preclinical studies (one of the targ recommendations of the U.S. National Institute			ntist ING INNOVATION	Must try hai Nature 483, 509 (29 M Published online 28 Ma	rder Aarch 2012) dol arch 2012	10.1038/483509a		r, m
Increasing transparency. Authors will indicate and E-Prints View FubMed Citation experimenter was blind to the conduct of the e guidelines.		The federal ag	uary 28, 2014	to improve the quality of scientific research.	Too many sloppy misi look more rigorously s	The Reprints 94 is takes are creepin at the data — and	Rights & permissions g into scientific pa f at themselves,	Article metric	es Is must

[Victoria Stodden: Reproducibility in High Performance Computing]

Unpacking "Reproducibility"

"Empirical Reproducibility"

"Computational Reproducit

"Statistical Reproducibility"

na	atur	e	مواقعا كه اعتبارها ركامية الماد		
= Merr	-		 Advanced search 	Branch	60
archive -	volume 483 -	issue 7391 -	editorialis + article		1 A 1
NATURE	EDITORIA	L			: ⊞ ⊕
Mus	t try ha	irder		Î	
Natura 4 Publishe	43, 509 (29 d online 28 M	March 2012) Ianch 2012	do: 10.1036/403508a	5	Society for I
	& Citation	Bearlots	Rights & permissions	📓 Article metric	
Too man look mo	ny sloppy mi re rigoreusly	stakes are cri at the data -	eping into scientific p - and at themselves.	apers. Lab heads	SIAM NEWS >
oili	ty"				"Setting t Research June 3, 2013

nature News & Comment Research Careers & Jobs Curren Home Volume 506 > Issue 7487 Archive News Feature

NATURE | NEWS FEATURE

Scientific method: Statistical errors

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

Regina Nuzzo

12 February 2014

Industrial and Applied Mathematics

之國即

the Default to Reproducible" in Computational Science

Following a late-2012 workshop at the Institute for Computational and Experimental Research in Mathematics, a group of computational scientists have proposed a set of standards for the dissemination of reproducible research.

Victoria Stodden, Jonathan Borwein, and David H. Bailey



Renew SIAM · Contact Us · Site Map · Join SIAM

V. Stodden, IMS Bulletin (2013)



(Non-)Reproducibility with Computational Sciences

Many levels of abstractions

- Reproduce a scientific result (replicability)
 - Find same insight (different team, different system)
- Reproduce a computer based experiment (reproducibility)
 - Run code again and check result (different team, same system)
 - Good scientific practice
- Bitwise reproduce a single experiment (repeatability)
 - Run program with identical input data (same team, same system)

All levels are present at the same time!

ACM 2020 terminology



Reproducibility Contest at SC'16 (and following SCs)

SCC Reproducibility Initiative Winner

Replication and reproducibility of experimental computer science results in peer-reviewed paper is gaining relevance in the HPC community. SC, the leading conference in the field, wants to promote and support replication and reproducibility through a new initiative that aims to integrate aspects of past technical papers into the Student Cluster Competition (SCC). The SCC is excited to announce "A parallel connectivity algorithm for de Bruijn graphs in metagenomic applications" as the winning paper for the inaugural reproducibility initiative. This paper and accompanying application will be reproduced in the SCC at SC16. This is the first time that students have been challenged to reproduce a paper rather than run prescribed data sets. Although they are doing similar tasks from previous competitions, they are seeing it from an entirely new perspective, as a component to the scientific process. "We want students to understand, early in their careers, the important role reproducibility plays in research." explains the SCC Chair Stephen Harrell

(Follow-up activity to Stodden's talk at SC'15)

APART 2024, Obergurgl, Austria







What are the Problems Here?

In Climate Science

- Highly non-linear mathematics
- Simulations over long time periods

In Computer Science

- Associative property does not hold
 - a+(b+c) != (a+b)+c with limited number representation and rounding
- Machines are non-deterministic



Bitwise Reproducibility

What might harm bitwise reproducibility?

- Processors are exchanged
- Libraries are exchanged
- Compilers are exchanged
- Compiler options are changed
- Domain region is differently partitioned
- Number of compute nodes is changed
- Non-determinism in libraries with e.g. reduction operations

operators' responsibility

- Non-determinism in programs with e.g. load balancing
- This is a non exhaustive list
 - The order of items is non-deterministic ③

12

scientists' responsibility



Procedure for Climate Modelling

Example: ICON (icosahedral non-hydrostatic general circulation model)

Requirement

Must deliver bitwise reproducible results for all possible domain compositions (strong requirement!)

 In fact: sequential and parallel must be identical They really compare small data sets (single core vs. multiple core)



Benchmarking Problem During Procurement

Requirements

- Vendor is allowed to use optimal compiler options
- Vendor is not allowed to change the scientific result of the climate model programs

Consequences

- Bitwise reproduction cannot be guaranteed
- However, result must not be biased

14



Example 1: A Biased Result

Procedure

Vendors use -O2 and -O3 with vectorization

Results

- Exhibits differences between
 - -O2 novec and -O3 vec
 - -O3 AVX from run to run

AVX + FMA + dynamic memory management = non-deterministic

Thomas Ludwig, Beate Geyer; Reproduzierbarkeit. Informatik Spektrum **42**, 48–52 (2019)

t_{sim}=100d O2: 10min30sec -> O3: 8m41sec => (+17,3%)



© Thomas Ludwig

APART 2024, Obergurgl, Austria

> DKRZ

H Universität Hamburg



t_{sim}=100d O3 -x AVX - 4 runs with identical input



© Thomas Ludwig



Climate vs. Weather

climate is the statistics of weather – maybe the problem disappears?

© Thomas Ludwig

APART 2024, Obergurgl, Austria



Example 2: Ensemble Models

- Climate models use ensembles
 - Run the model n times
 - Slightly change starting condition, e.g. initial states, start time
 - Get n different results
 - Make a statistical merging of the individual results
- Why do we do this
 - Do some uncertainty quantification
 - Get insight concerning robustness of the models
 - Get better result than just having smaller grids
- Size of n? Several dozens, more are better, computers are limited



Experiment

- 1. Run a model in ensemble mode (result set A)
- 2. Take a random ensemble member and run it on different parallel computers (result set B)
- 3. Compare the variability of relevant variables in the two result sets

Beate Geyer, Thomas Ludwig, Hans von Storch: Limits of reproducibility and hydrodynamic noise in atmospheric regional modelling, *Nature Communications Earth & Environment* **2**, 17 (2021)



Experiment Setup

- Uses COSMO-CLM model
 - Has bitwise reproducibility for fixed execution environment
- Set A was produced on CSCS platform with six different initial states (6 ensemble members)
- Set B was produced with ONE initial state on SIX platforms
 - 2 at DKRZ (IBM, Intel), CSCS, LRZ, DWD, ZAMG

21





Fig. 1: Air temperature deviations from ensemble mean





Findings

- Variability due to ensemble computing and variability due to platform change have a very similar shape
- "Do not fear a change of platform, if you run in ensemble mode" (without real prove)

However, a change of platform is in general a problem



This paper and its results cannot be reproduced... ... because the machines do no longer exist. ③

But: only refers to ACM's repeatability and reproducibility not to replicability (different team, different system)



Example 3: Automotive Simulations

"Yeah, we know, we never get the same result – we do not care, it is not important."



Some Observations

- Bitwise reproducibility (repeatability) difficult to implement
- Load balancing not possible
- Usage of optimized libraries prohibited
- High compiler optimization impossible
- And still not a reproducible scientific result (because you will never have the same computer with the same libraries, compilers etc.)
- And presumably still not a reproducible (replicable) computer based experiment because of lack of description (a third party will not be able to reproduce it, not even on an identical environment)



Some Questions Related to Science in General

Let's assume you really can reproduce an HPC based program run in a bitwise manner

- Why would you as the program author want to do that?
- Answer from Max Planck Institute for Meteorology
 - Find errors in parallelization
 - E.g. races in OpenMP-based implementations
 - Understand influences of changes with
 - Compilers, libraries, global sums, ...



Some Questions Related to Science in General

Let's assume you really can reproduce an HPC based program run in a bitwise manner

- Why would you as the program author want to do that?
- Why would someone else want to do that?
- Who would want to pay for doing this, if it implies to copy e.g. the DKRZ infrastructure?
- Which progress in the history of mankind was based on being able to reproduce someone else's scientific result in an identical way?



A Personal Pre-Final Observation

- More questions than answers
 - Different answers from different people
- Exascale will intensify the problems
- Data intensiveness will intensify the problem
- Machine learning will add to the problem
- The production of new scientific results increases exponentially
- Science needs to investigate the issue of how to evaluate the correctness and validity of its results
- Fortunately, we see for about 5 years more and more workshops on this issue