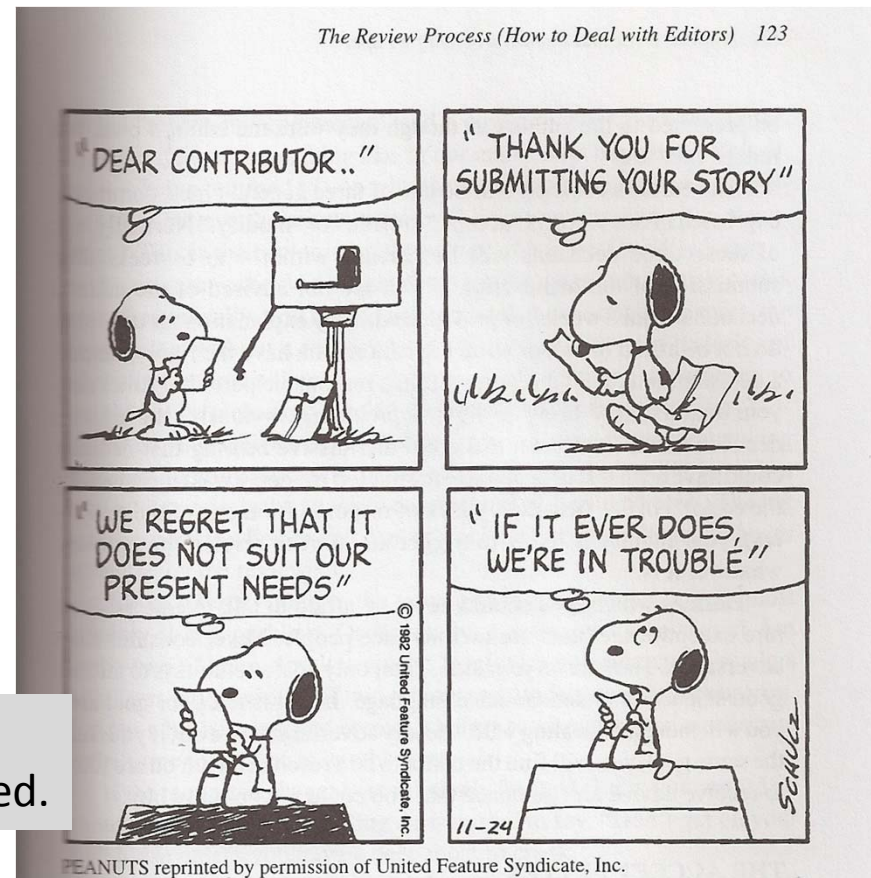


What is a Good PhD?

— some «common sense» and personal views

Lasse Natvig
Professor in computer architecture
Lasse@computer.org



From: How to Write and Publish a Scientific Paper, Robert A. Day 5th ed.

Presentation Overview



- What is a Good PhD?
 - Context
 - PhD theses I have supervised (6 + 3?) or evaluated (13)
 - Quality → Importance of focus
 - Research group / Supervisor / PhD student
 - Reproducibility and testing (Method)
 - Quantity
 - 6 papers — «The Reidar Model»
 - From NTNU regulations
 - Surprise

From the official NTNU regulations

Underlining and coloring by Lasse

- From *Guidelines for the Assessment of Candidates for Norwegian Doctoral Degrees*, Section 3.2 Assessment of the thesis [NTNU12b]:
 - A Norwegian doctoral degree is awarded as proof that the candidate's research qualifications are of a certain standard
 - ... the academic standard and quality of the work submitted
 - ...the candidate must satisfy the *minimum requirements to qualify as a researcher* – demonstrated through requirements related to the formulation of research questions, precision and logical stringency, originality, a good command of current methods of analysis and be able to reflect on their possibilities and limitations.
 - ... thesis must contribute new knowledge to the discipline and be of an academic standard appropriate for publication as part of the scientific literature in the field
 - ...
 - And more! 😊

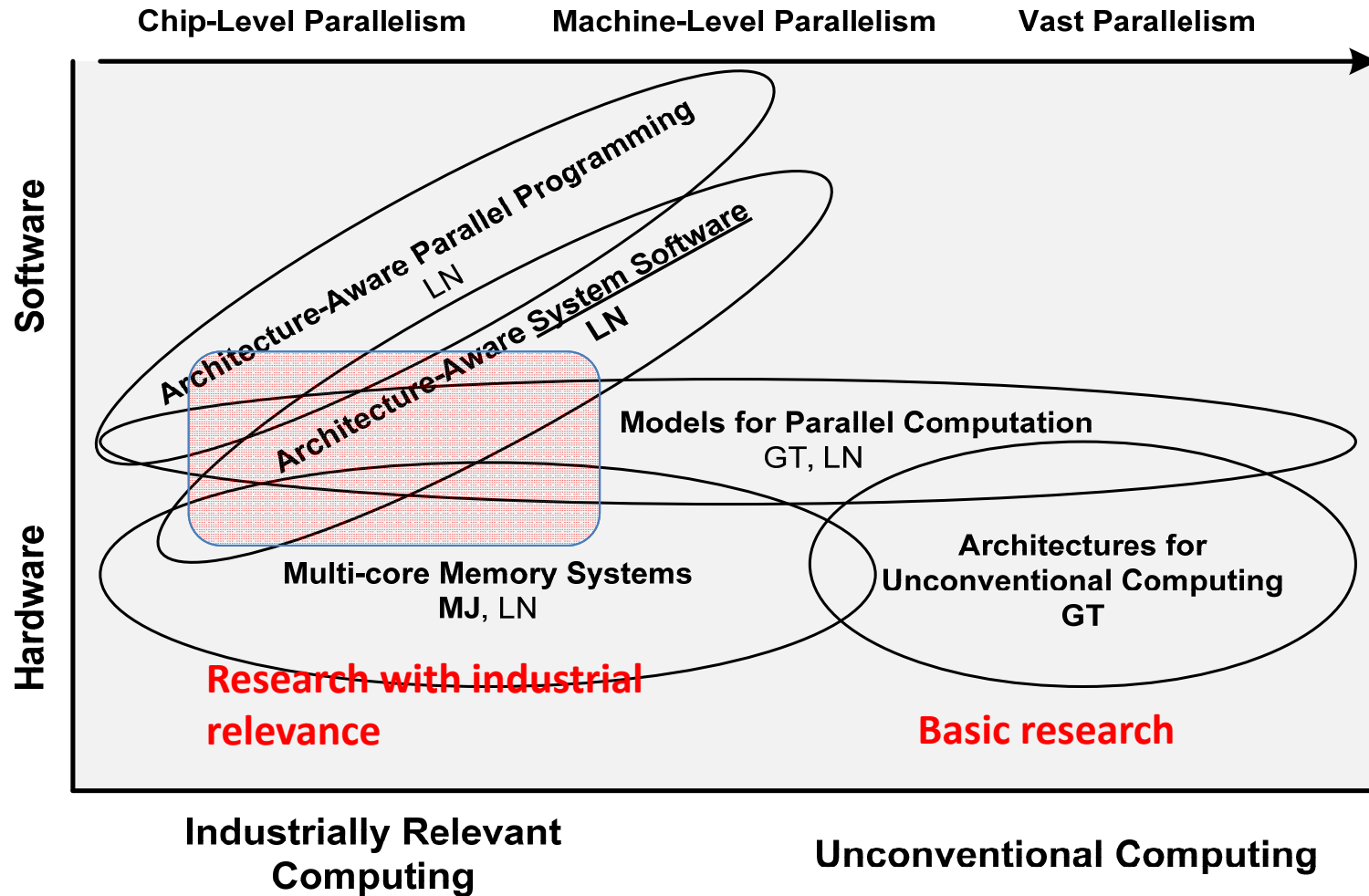


CONTEXT AND FOCUS

What is Computer Architecture?

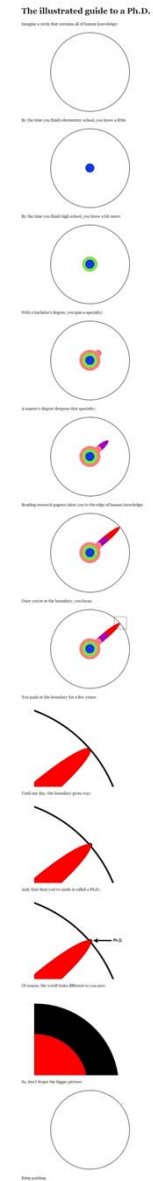
- Computer architecture “is a specification detailing how a set of software and hardware technology ...interact to form a computer system determining the needs of the user/system/technology, and creating a logical design and standards based on those requirements”
[Techop]
 - + performance evaluation
 - Includes parallel processing (My personal interest in 30 years)
- Broad knowledge vs. deep knowledge
- There is an old saying, “Architects know a little about almost everything and an engineer knows a lot about almost nothing.”[Career]

How to focus within architecture?

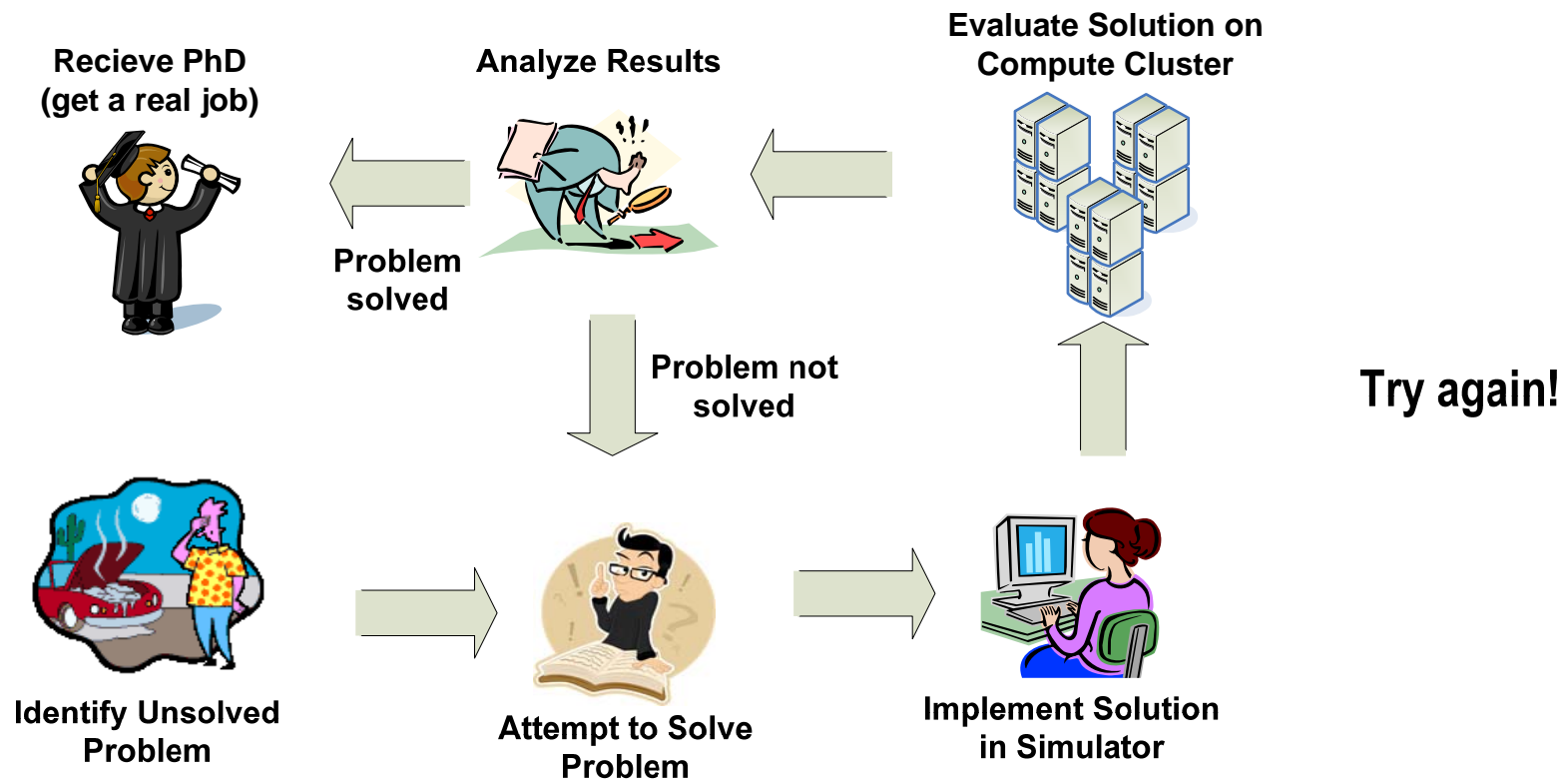


A PhD student must focus even more!

- JUMP to [The illustrated guide to a Ph.D](#) by [Matt Might](#)



Research Workflow

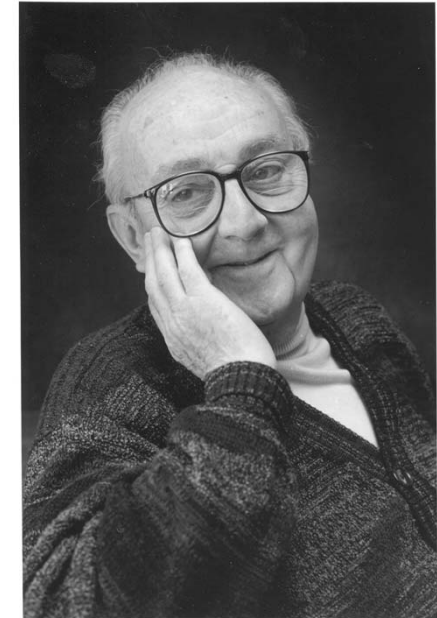


From: How to Write a Computer Architecture Paper,
lecture about miniproject report writing in TDT4260 comp.arch [Jahre-14]

REPRODUCIBILITY

Abstraction/Models & Reproducibility

- Model of a system
 - Model the interesting parts with high accuracy
 - Model the rest of the system with sufficient accuracy
- “The Danger of Abstraction”
 - [George E. P. Box](#):
 - “All models are wrong but some are useful”
 - “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful”
- Abstractions and simplifications
 - Even more important for small countries/groups!
- Hm....,
how to get people to trust our research?
 - 100% precise documentation!
 - **Reproducibility**



Give "all" experimental details

	Crossbar Based Architecture			Ring Based Architecture		
	4-core	8-core	16-core	4-core	8-core	16-core
ITRS Year of Production	2007	2010	2013	2007	2010	2013
Feature Size (nm)	65	45	32	65	45	32
Shared Cache Size (MB)	8	16	32	8	16	32
Memory Bus Channels	1, 2 or 4	1, 2 or 4	1, 2 or 4	1, 2 or 4	1, 2 or 4	1, 2 or 4
Interconnect Latency (End-to-End/Per Hop)	8/-	16/-	30/-	-4	-4	-8

Table III
CACHE PARAMETERS

Cache	Size (4-core/8-core/16-core)	Associativity	Access Latency (cycles)	Cycle Time (cycles)	MSHRs / WB (per bank)	Banks	Area (mm ²)
Level 1 Private Cache	64KB	2	3/2/2	2	16MSHRs/4WB	1	2.3/1.1/0.5
Level 2 Private Cache	1 MB	4	9/6/5	4/3/2	16	1	14.6/7.0/3.6
Level 2/3 Shared Cache	8/16/32 MB	16	16/12/12	4	16/32/64	4	94.0/91.9/84.7

Table IV
PROCESSOR CORE PARAMETERS

Parameter	Value
Clock frequency	4 GHz
Reorder Buffer	128 entries
Store Buffer	32 entries
Instruction Queue	64 instructions
Instruction Fetch Queue	32 entries
Load/Store Queue	32 instructions
Issue Width	4 instructions/cycle
Functional units	4 Integer ALUs, 2 Integer Multiply/Divide, 4 FP ALUs, 2 FP Multiply/Divide
Branch predictor	Hybrid, 2048 local history registers, 4-way 2048 entry BTB

Table V
INTERCONNECT AND DRAM INTERFACE

Parameter	Value
Crossbar Interconnect	8/16/30 cycles end-to-end transfer latency, 32 entry request queue, Pipelined (2/4/6 pipe stages)
Ring Interconnect	4/4/8 cycles per hop transfer latency, 1/1/2 pipe stages per hop, 32 entry request queue, 1/2/2 request rings, 1 response ring
Point to Point Link	4/3/2 transfer latency, 32 entry request queue
Main memory	DDR2-800, 4-4-4-12 timing, 64 entry read queue, 64 entry write queue, 1 KB pages, 8 banks, FR-FCFS scheduling [21], Closed page policy

From: *A Quantitative Study of Memory System Interference in Chip Multiprocessors*, Jahre et al., HPC09

Reproducibility

Ten Simple Rules for Reproducible Computational Research, by Geir Kjetil Sandve et.al. [SNTH13]

- 1: For Every Result, Keep Track of How It Was Produced
- 2: Avoid Manual Data Manipulation Steps
- 3: Archive the Exact Versions of All External Programs Used
- 4: Version Control All Custom Scripts
- 5: Record All Intermediate Results, When Possible in Standardized Formats
- 6: For Analyses That Include Randomness, Note Underlying Random Seeds
- 7: Always Store Raw Data behind Plots `Matplotlib, gnuplot`
- 8: Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
- 9: Connect Textual Statements to Underlying Results
- 10: Provide Public Access to Scripts, Runs, and Results



Parallel computers using random numbers might execute non-deterministically

More on reproducibility

- 4'th Int'l Workshop on Adaptive Self-tuning Computing Systems [ADAPT'14]
 - Two papers got the quality mark **reproducible**
- 1st ACM SIGPLAN Workshop on Reproducible Research Methodologies and New Publication Models in Computer Engineering [TRUST14]

See also:

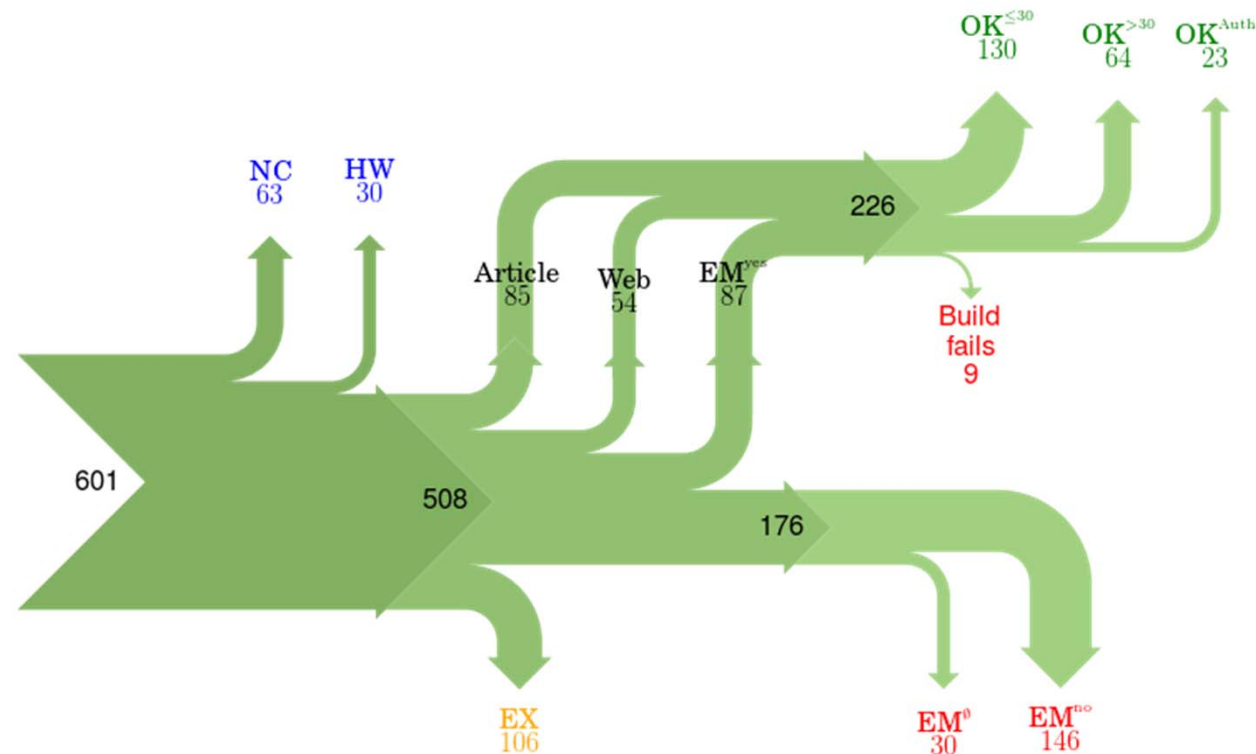
<http://ctuning.org/reproducibility>
(Grigori Fursin)

Validated
by the community



More on reproducibility

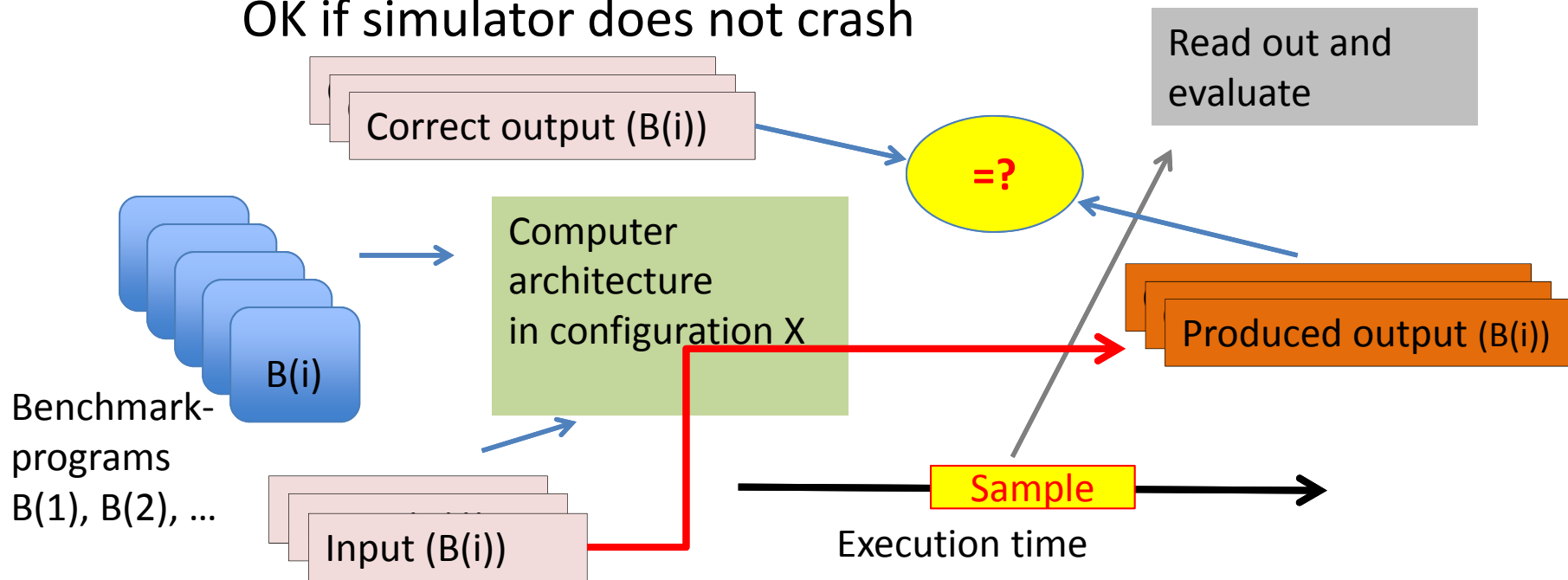
- *Repeatability* in Computer Science
- Techn. Report (68 pages)
- <http://reproducibility.cs.arizona.edu/>



TESTING

The importance of testing

- (Industry typically use 50% of work force for testing)
 - They cannot afford low quality
- Running benchmarks in computational comp.arch.
 - Common practice has not been perfect: Assumed OK if simulator does not crash



From ADEPT workshop January 2014

Presented by David Black-Schaffer, Uppsala [SHBS14]:

David Black-Schaffer Uppsala University / Department of Information Technology 1/21/14 | 24

An Aside: the Importance of Verification

30B OoO + vFF

Benchmark	Verifies in Reference	Verifies using VFF	Verifies when Switching
400.perlbench 401.bzip2 416.gamess	Yes	Yes	Yes
433.milc 453.povray 456.hmmer			
458.sjeng 462.libquantum 464.h264ref			
471.omnetpp 481.wrf 482.sphinx3			
483.xalancbmk			
410.bwaves 434.zeusmp 435.gromacs	No	Yes	Yes
436.cactusADM 444.namd 459.GemsFDTD			
470.lbm			
445.gobmk 450.soplex 454.calculix	Fatal Error ¹	Yes	Yes
429.mcf 473.astar	Fatal Error ²	Yes	Yes
437.leslie3d	Fatal Error ³	Yes	Yes
403.gcc	Fatal Error ⁴	Yes	Yes
447.dealII	Fatal Error ⁵	Yes	No
465.tonto	Fatal Error ⁶	Yes	Yes
	13/29 verified, 9/29 fatal	29/29 verified	28/29 verified

Only 13 verify when doing a 30B instruction OoO simulation.

1. Simulator gets stuck.
2. Triggers a memory leak causing the simulator crash.
3. Terminates prematurely for unknown reason.
4. Fails with internal error. Likely due to unimplemented instructions.
5. Benchmark segfaults due to unimplemented instructions.
6. Terminated by internal benchmark sanity check.

QUANTITY

When is 6 papers good enough?

- First/main author of most
 - “If the thesis consists primarily of papers, the candidate must normally be the main author or first author of at least half the papers” [NTNU12a]
- At least 2 - 4 in high quality conferences or good journals
- All in acceptable journals, conferences or good workshops
 - IDI Relevant Conferences (357), A and B rating (can have weaknesses) [IDI-AB]
 - 1 (or maybe 2) can be in state submitted, if ...
- **Watch out!**
 - There are “fake conferences” and “bogus journals” (and websites)
 - Accepting papers written by paper-automata
 - You can easily get papers published that NEVER should have been published
 - Your and (your supervisors) responsibility

PhD as a collection of papers

- If the thesis consists of several interrelated minor pieces of work, the candidate must document the integrated nature of the work and the assessment committee must decide whether the content comprises a coherent entity. In such cases, the candidate must compile a separate part of the thesis that not only summarizes but also compares the research questions and conclusions presented in the separate pieces... [NTNU12b]

Haakon Dybdahl [Dybd07]

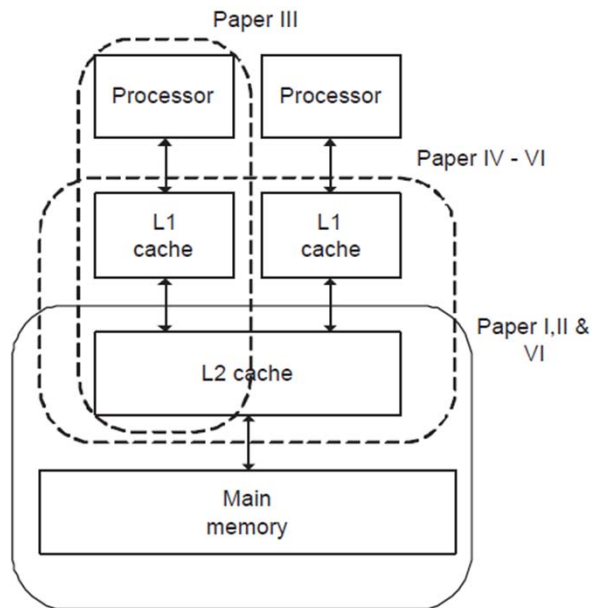
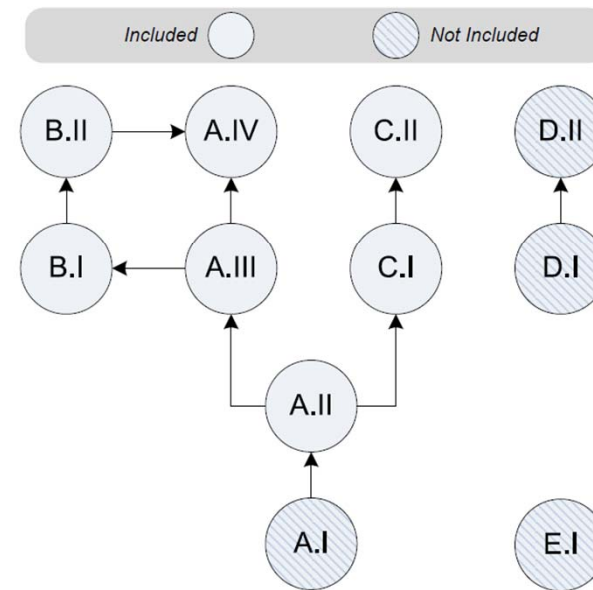


Figure 3.2: The research focus for the different papers.

34

Chapter 4. Research Process

Magnus Jahre [Jahre10]



(a) Logical Structure

... more examples of research process

16

Chapter 3. Research Summary

Morten Hartmann [Hart05]

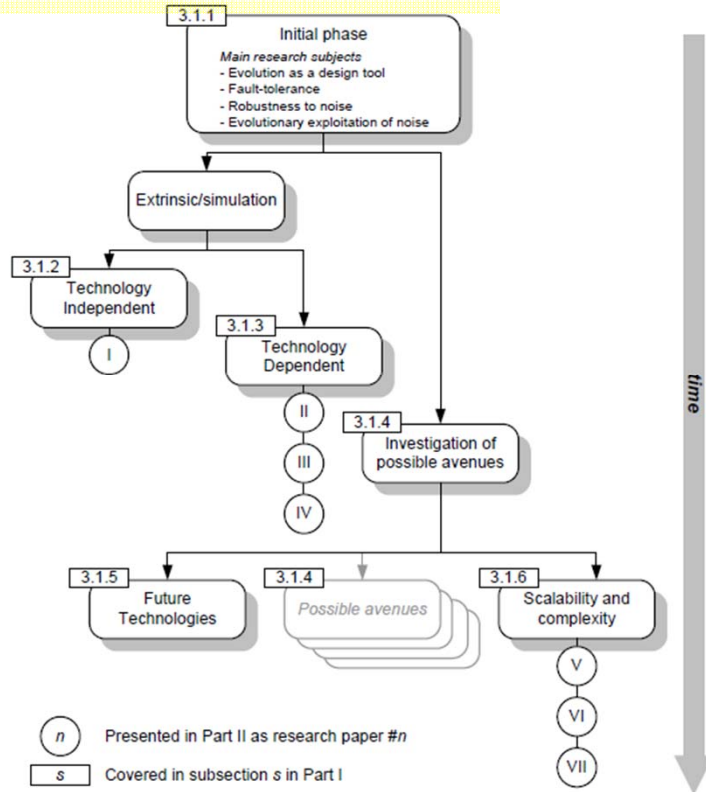


Figure 3.1: A conceptual illustration of the research process and relevant contributions

Asbjørn Djupdal [Djup08]

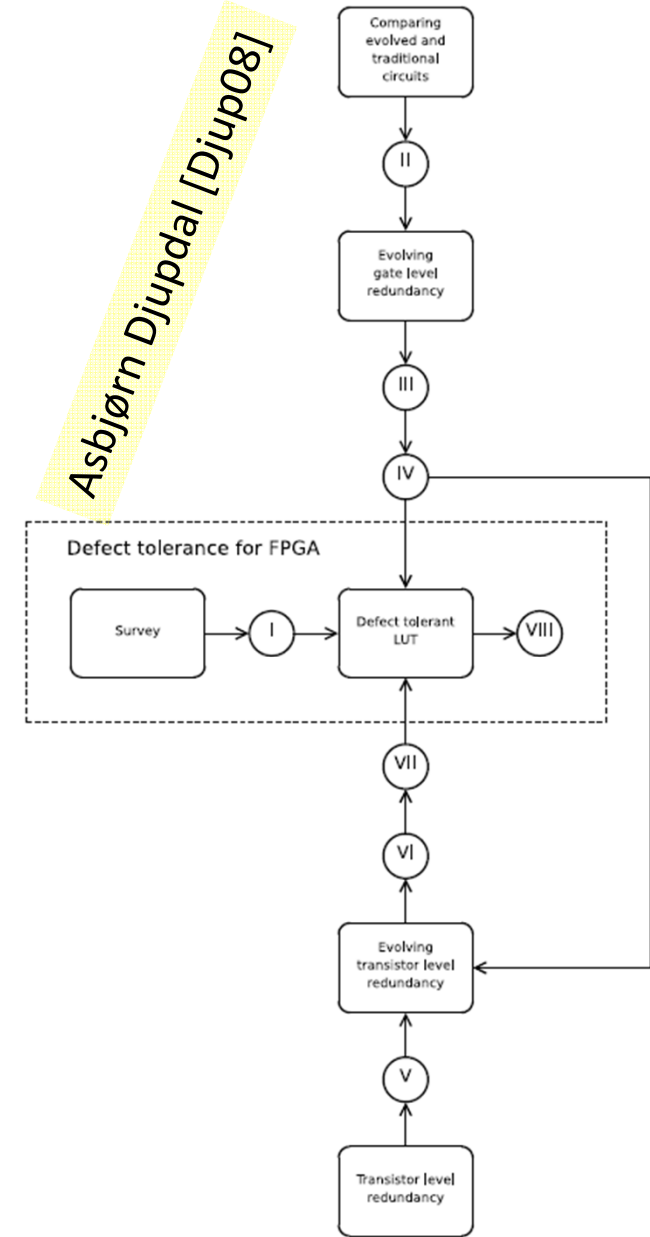


Figure 3.1: Research process and relation of papers

SURPRISE

How to supervise within a topic you do not know?

- ... or know only to some extent
- Case b) Change of main supervisor (not common)
- Case a) Your own student working efficiently and independently/self-driven
 - A normal case, or ideal case
 - How well can the PhD student answer your questions?
 - Clear and precise descriptions?
 - “General attitude”
 - from “maximum quality” to ... (worst case) “don’t care attitude”

Motivate your supervisor!

- Use the time with the supervisor efficiently
- Be prepared
 - Bring results, ideas, questions
- Take notes
- Give your supervisor time to prepare
- Help him/her supervise
 - Write readable
 - Use figures, visualizations
 - Use abstraction
 - Be precise and pedagogical
- You have one project, your supervisor might have 10-30 “projects”

Scientific writing, precision

- Notation/concepts
 - Often new concepts
 - Use best/most common terminology --- if it exist
 - Define your terminology precisely
 - Stick to it, be consistent!
- “help the reader”
- More (in Norwegian)
 - [Lasse's enkle tips om rapportskrivning](#)

References

Disclaimer: Some of these are “low-value references”
(All are incomplete, but contain hyperlinks)

- [ADAPT14] [The 4'th Int'l Workshop on Adaptive Self-tuning Computing Systems](#), Vienna, January 2014
- [Career] [What Is The Difference Between Architecture And Civil Engineering?](#)
- [Djup08] [Evolving Static Hardware Redundancy for Defect Tolerant FPGAs](#), PhD thesis by Asbjørn Djupdal
- [Dybd07] [Architectural Techniques to Improve Cache Utilization](#). Dr.ing. thesis by Haakon Dybdahl, 2007
- [Hart05] [Evolution of Fault and Noise Tolerant Digital Circuits](#), PhD thesis by Morten Hartmann, 2005
- [IDI-AB] [IDI Relevant Conferences](#) (list for travel grants, A and B rating)
- [Jahre10] [Managing Shared Resources in Chip Multiprocessor Memory Systems](#), PhD thesis by M. Jahre, 2010
- [Jahre14] [How to Write a Computer Architecture Paper](#), lecture about miniproject report writing in course TDT4260 comp.arch, given by Nico this spring
- [JN10] [Computational Computer Architecture Research at NTNU](#), ERCIM News April 2010
- [NTNU12a] [Regulations For The Philosophiae Doctor Degree \(PhD\) at NTNU](#), 23 January 2012.
- [NTNU12b] [Guidelines for the Assessment of Candidates for Norwegian Doctoral Degrees](#), NTNU 13 June 2012
- [SHBS14] [Full Speed Ahead: Detailed Architectural Simulation at Near-Native Speed](#),
Andreas Sandberg, Erik Hagersten, and David Black-Schaffer. March 2014, Tech.report 2014-005
- [SNTH13] [Ten Simple Rules for Reproducible Computational Research](#), Geir Kjetil Sandve et.al., 2013
- [Techop] [Computer Architecture, from Techopedia](#)
- [TRUST14] [1st ACM SIGPLAN Workshop on Reproducible Research Methodologies and New Publication Models in Computer Engineering](#), Edinburgh, 12 June 2014.

Questions



*Visit the EECS website:
<http://www.ntnu.edu/ime/eecs/>*

Contact:

Lasse.Natvig@idi.ntnu.no

<http://research.idi.ntnu.no/multicore/>