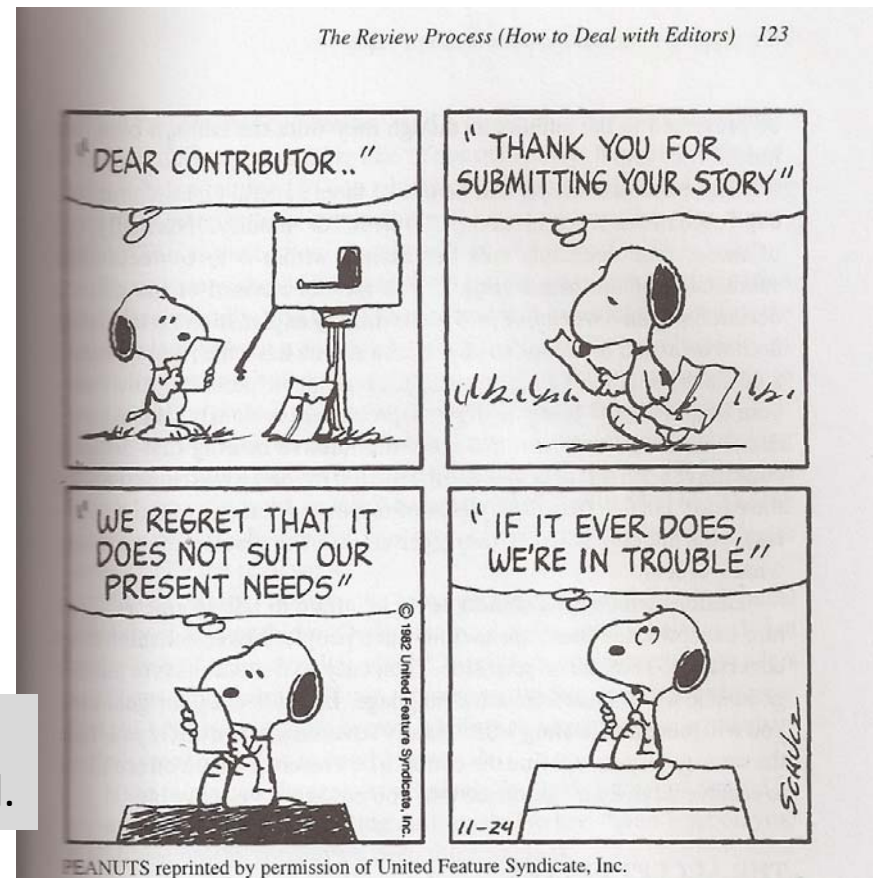# What is a Good PhD?

— some «common sense» and personal views

Lasse Natvig
Professor in computer architecture

Lasse@computer.org

The Review Process (How to Deal with Editors) 123

"DEAR CONTRIBUTOR.."

"THANK YOU FOR SUBMITTING YOUR STORY"

"WE REGRET THAT IT DOES NOT SUIT OUR PRESENT NEEDS"

"IF IT EVER DOES, WE'RE IN TROUBLE"

PEANUTS reprinted by permission of United Feature Syndicate, Inc.

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

Norwegian University of Science and Technology

Presentation Overview

- What is a Good PhD?
  - Context
    - PhD theses I have supervised (6 + 3) or evaluated (15)
  - <u>Quality</u> ➔ Importance of focus
    - Research group / Supervisor / PhD student
  - Reproducibility and testing (Method)
  - Quantity
    - 6 papers — «The Reidar Model»
  - From NTNU regulations
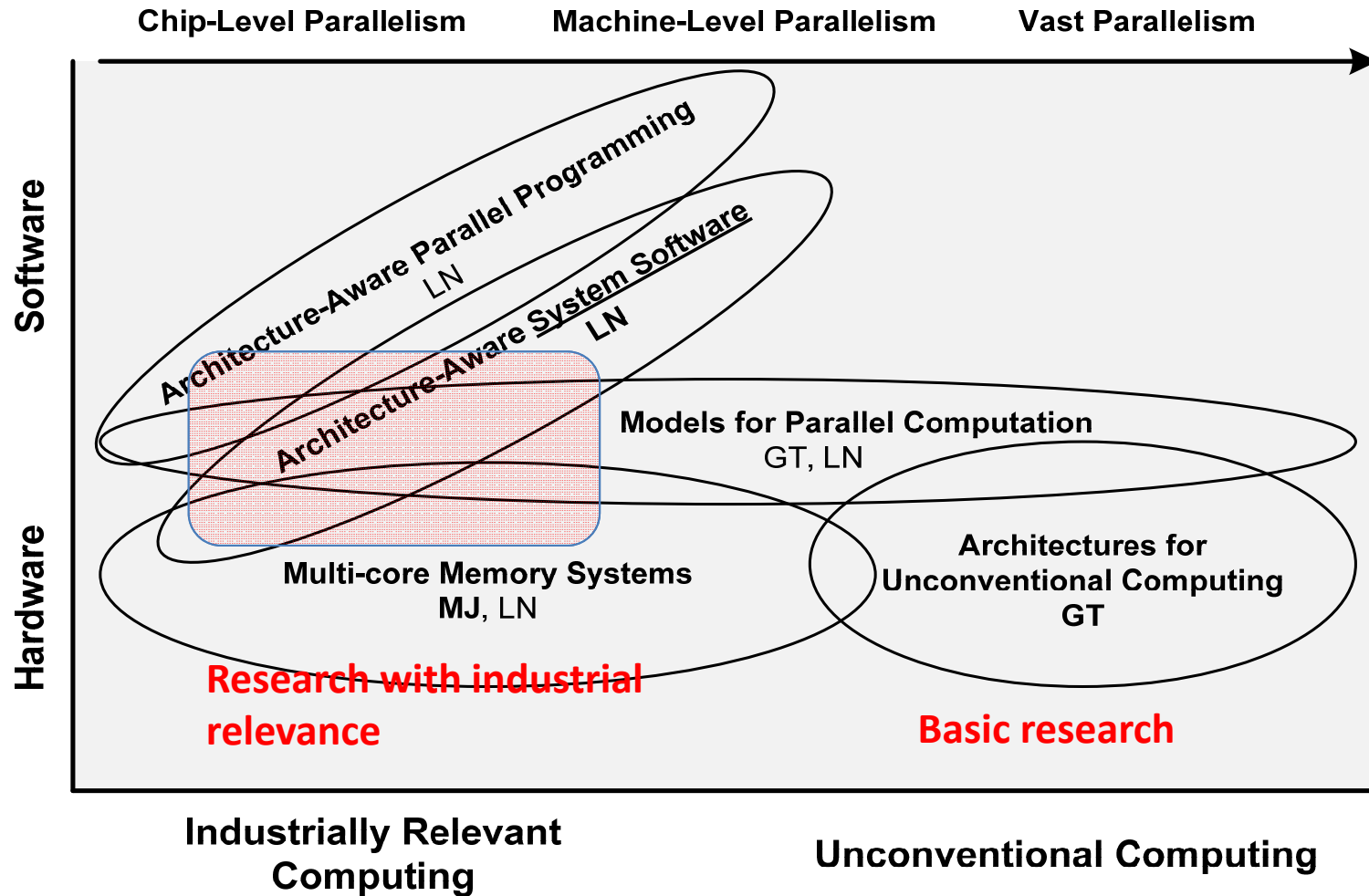 - Surprise

From the official NTNU regulations

- 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 <u>the candidate's research qualifications</u> are of a certain standard

  – … the academic <u>standard and quality</u> of the work submitted

  – …the candidate must satisfy the *minimum requirements to qualify as a researcher* – demonstrated through requirements related to the formulation of <u>research questions,</u> <u>precision and logical stringency</u><u>, originality, a good command of current methods of analysis and be able to reflect on their possibilities and limitations</u>…

  – … thesis must <u>contribute new knowledge</u> to the discipline and be of an academic standard <u>appropriate for publication </u>as part of the scientific literature in the field …

  – And more! ☺

NTNU

CONTEXT AND FOCUS
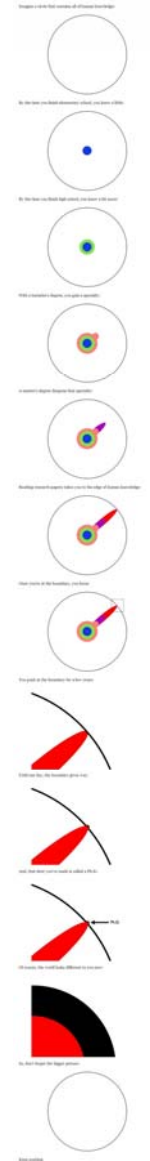
How to focus within architecture?

Chip-Level Parallelism **Machine-Level Parallelism** **Vast Parallelism**

Software

Hardware

Architecture-Aware Parallel Programming
LN

Architecture-Aware System Software
LN

Models for Parallel Computation
GT, LN

**Architectures for
Unconventional Computing
GT**

**Multi-core Memory Systems
MJ**, LN

**Research with industrial
relevance**

Basic research

**Industrially Relevant
Computing**

Unconventional Computing

A PhD student must focus even more!

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

http://matt.might.net/articles/phd-school-in-pictures/

Research Workflow

**Recieve PhD
(get a real job)**

Analyze Results

**Evaluate Solution on
Compute Cluster**

**Problem
solved**

**Problem not
solved**

Try again!

**Identify Unsolved
Problem**

**Attempt to Solve
Problem**

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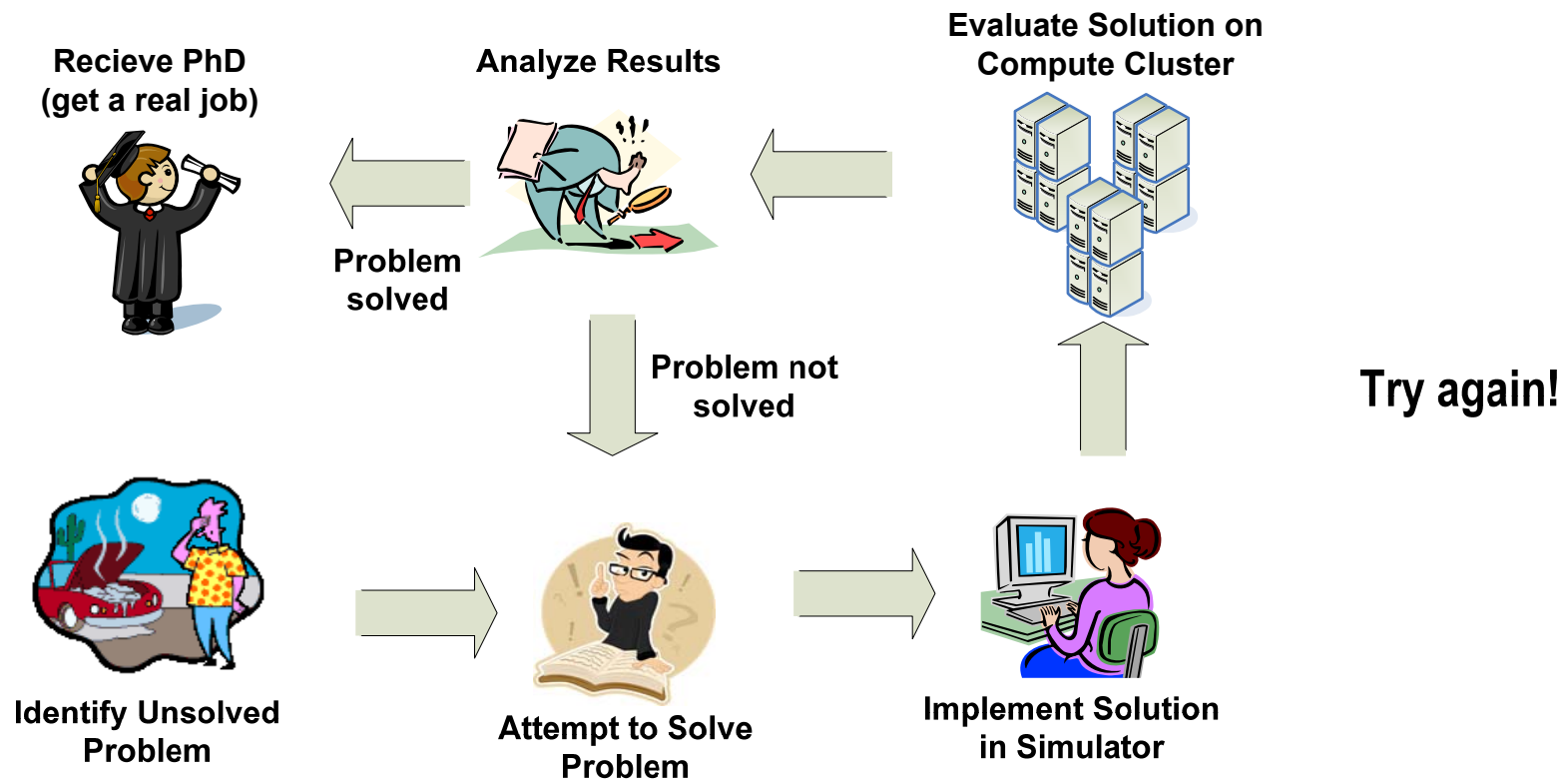
From: How to Write a Computer Architecture Paper,
lecture about miniproject report writing in TDT4260 comp.arch [Jahre-14]

REPRODUCIBILITY

NTNU

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"

- Hm…,
  how to get people to trust our research?
  - 100% precise documentation!
  - **Reproducibility**

Validated by the community

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 Multipy/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., HPCC09*

NTNU

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

CAUTION

Parallel computers using random numbers might execute non-deterministically

More on reproducibility (Anno 2014)

- 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/** and
http://www.dividiti.com/
(Grigori Fursin & Anton Lokhmotov)

More on reproducibility

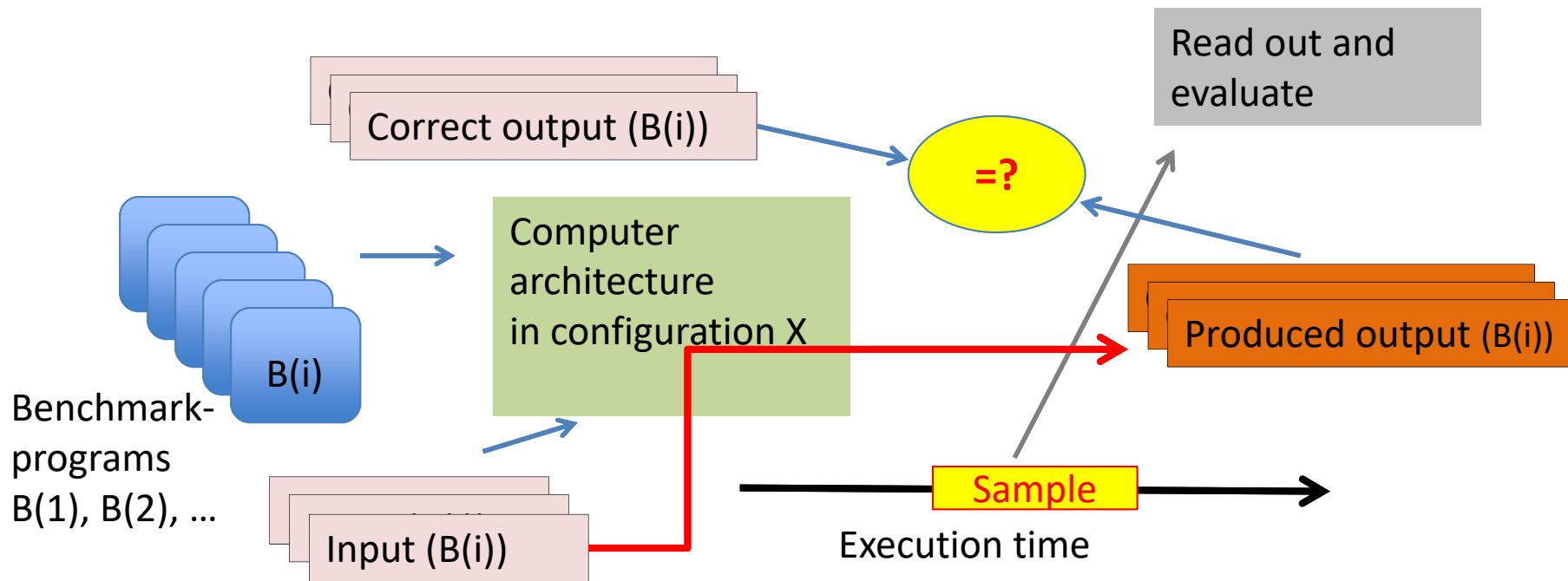- *Repeatability* in Computer Science
- Techn. Report (68 pages)
- http://reproducibility.cs.arizona.edu/
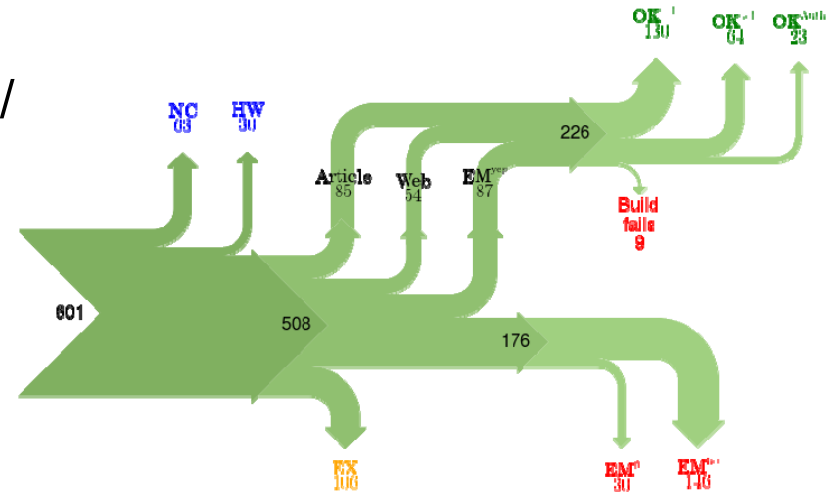
Legend					
Classification		**Code Location**		**Build Results**	
BC	Paper where the results are backed by code.	Article	Code is found from link in the article itself.	OK$^{\leq 30}$	We succeed in building the system in ≤ 30 minutes.
NC	Paper excluded due to results not being backed by code.	Web	Code is found from a web search.	OK$^{>30}$	We succeed in building the system in >30 minutes.
HW	Paper excluded due to replication requiring special hardware.	EM^{yes}	Code is provided by author after email request.	OK$^{>Author}$	We fail to build, but the author says the code builds with reasonable effort.
EX	Paper excluded due to overlapping author lists.	EM^{no}	Author responds that the code cannot be provided.	Fails	We fail to build, and the author doesn't respond to survey or says code may have problems building.
		EM$^{\varnothing}$	Author does not respond to email request within 2 months.		

NTNU

TESTING

The importance of testing

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

Correct output (B(i))

Read out and evaluate

=?

Computer architecture in configuration X

Produced output (B(i))

B(i)

Benchmark-programs B(1), B(2), …

Input (B(i))

Sample

Execution time

NTNU

From ADEPT workshop January 2014

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

QUANTITY

NTNU

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 <u>high quality</u> 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 <u>document the integrated nature of the work and the assessment committee must decide whether the content comprises a coherent entity</u>. 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]
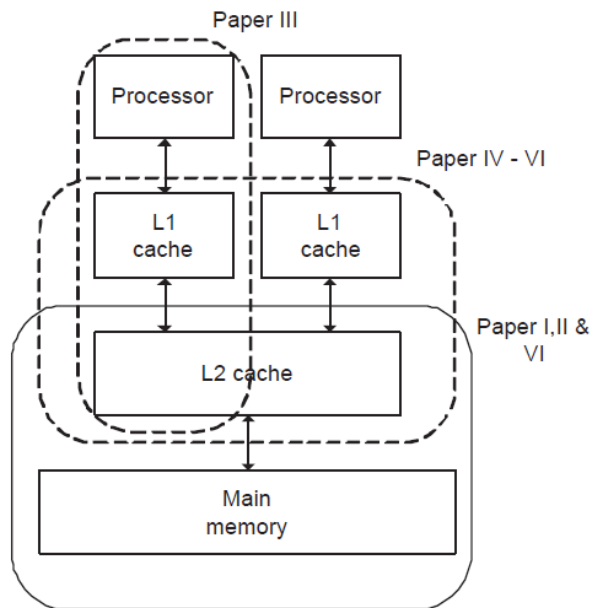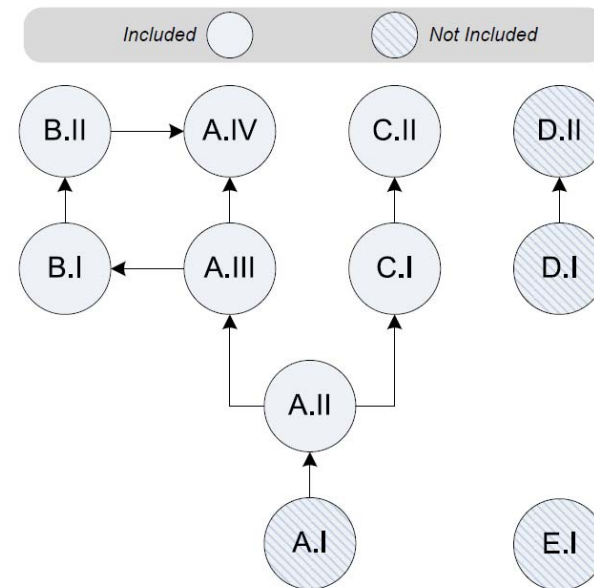
Haakon Dybdahl [Dybd07]

Figure 3.2: The research focus for the different papers.

Magnus Jahre [Jahre10]

(a) Logical Structure

… more examples of research process

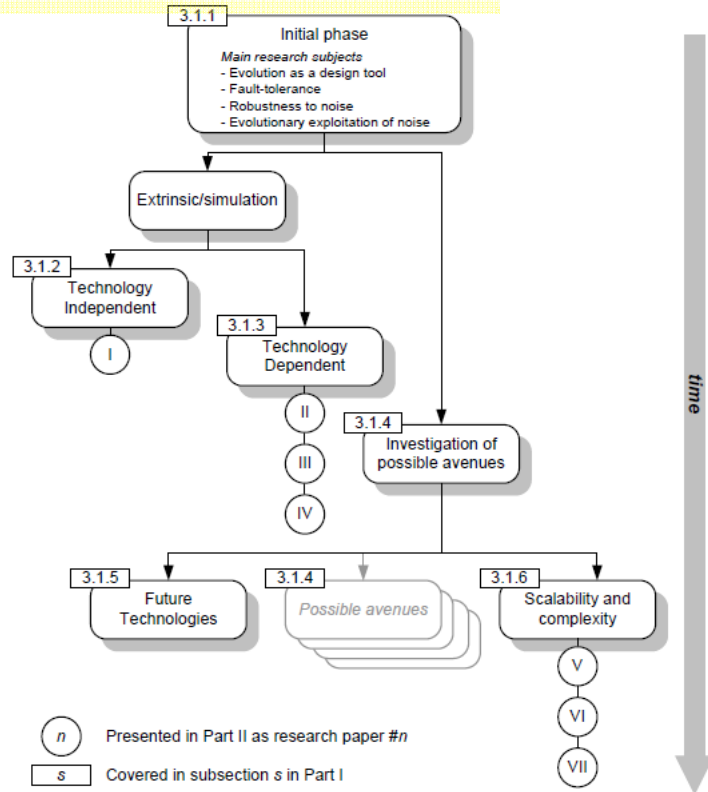Morten Hartmann [Hart05]

16   Chapter 3. Research Summary

Figure 3.1: A conceptual illustration of the research process and relevant contributions
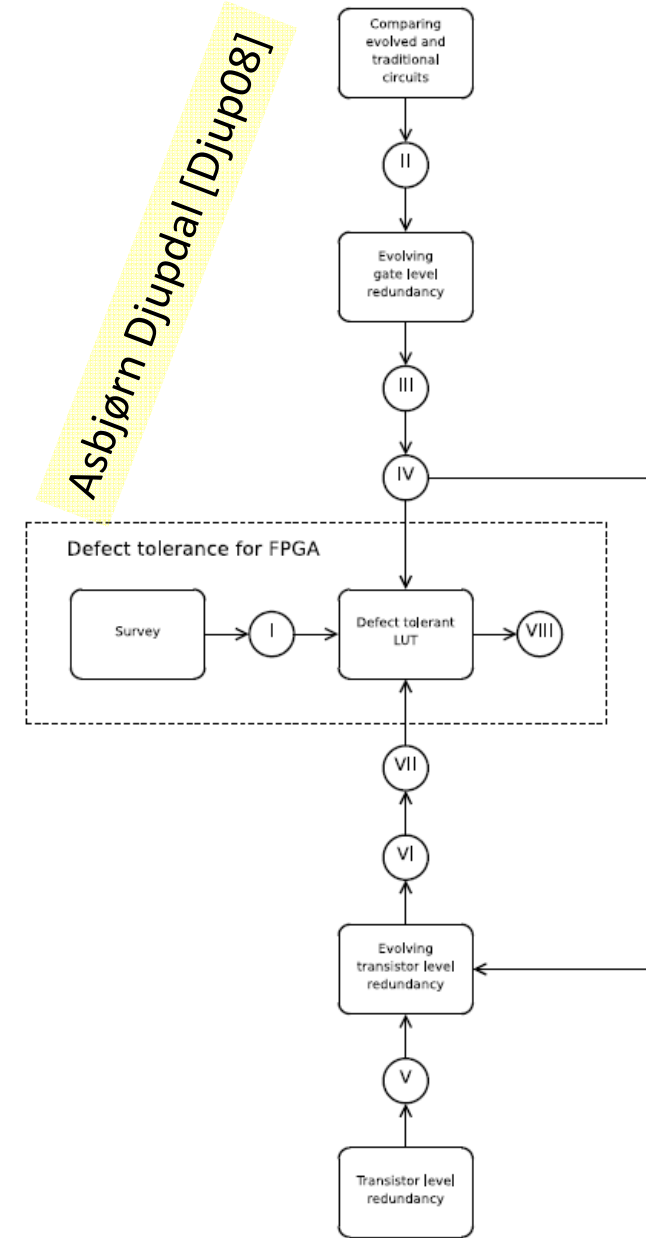
Asbjørn Djupdal [Djup08]

Figure 3.1: Research process and relation of papers

SURPRISE

NTNU

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"

NTNU

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 (→ jupyter.org ?)
  - 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 rapportskriving](#)

NTNU

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

[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.

NTNU

Questions

Contact:
Lasse.Natvig@ntnu.no

NTNU