

# How to Write Research Papers

## Part 1 – General Principles



SAMUEL GINN  
COLLEGE OF ENGINEERING

**Xiao Qin**

*Department of Computer Science and  
Software Engineering*

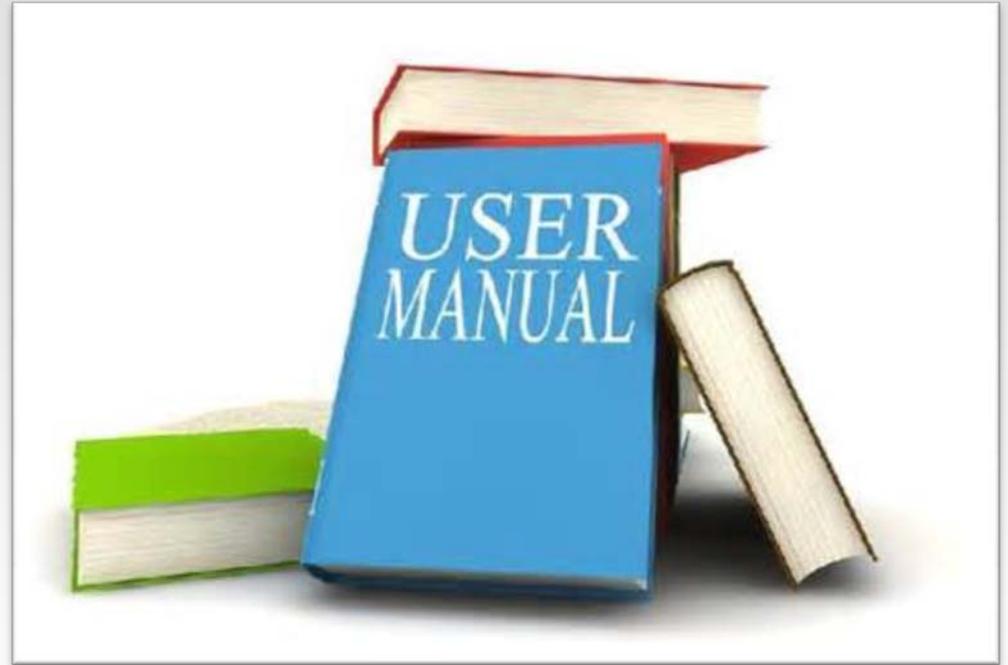
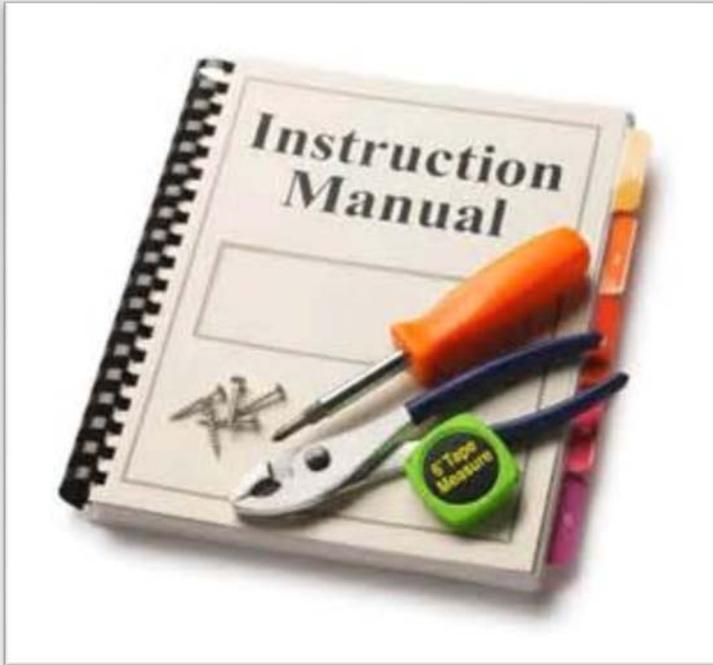
*Auburn University*

*<http://www.eng.auburn.edu/~xqin>*

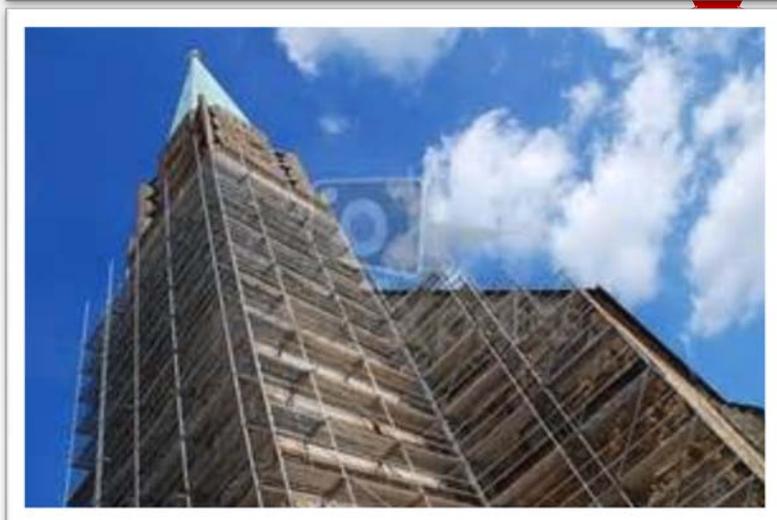
*[xqin@auburn.edu](mailto:xqin@auburn.edu)*

**Some slides are adapted from notes by Simon Peyton Jones (Microsoft Research, Cambridge)**

# My First Paper



# How to write a good research paper?





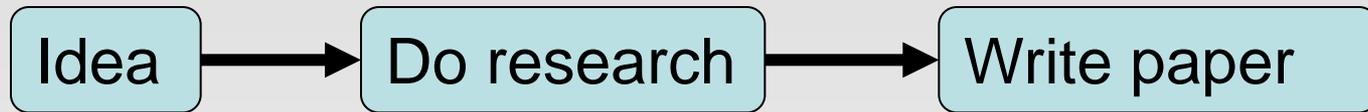
# Writing Papers is a Skill

- Many papers are badly written
- Good writing is a skill **you can learn**
- It's a skill that is worth learning:
  - You will get more brownie points (more papers accepted etc)
  - Your ideas will have more impacts
  - You will have better ideas

Increasing importance

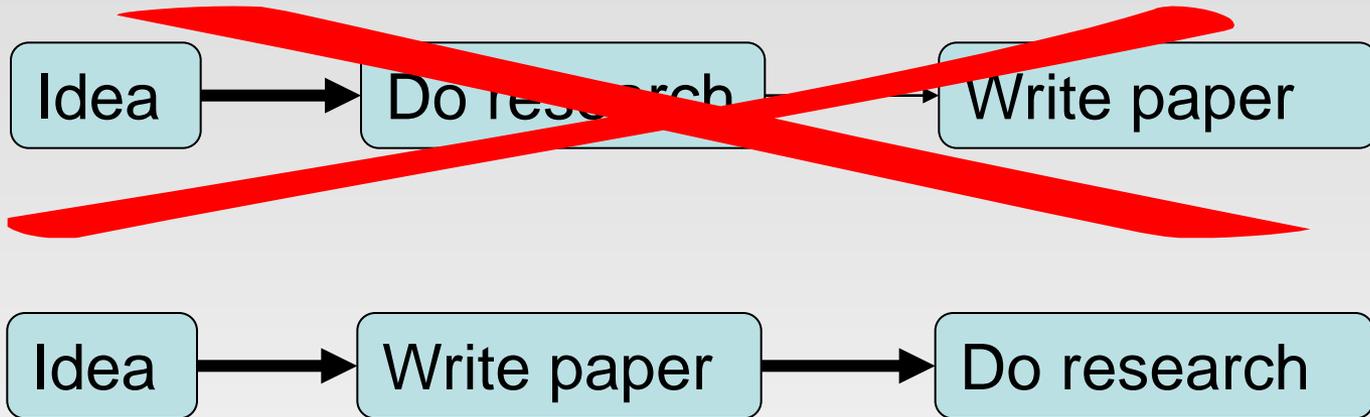


# Writing Papers: Model 1



When I was a doctoral student 7 years ago, this is my model of writing papers.

# Writing Papers: Model 2



- Forces to be clear, focused
- Crystallises what we don't understand
- Opens the way to dialogue with others: reality check, critique, and collaboration

# Do not be Intimidated

**Fallacy** You need to have a fantastic idea before you can write a paper. (Everyone else seems to.)

Write a paper,  
and give a talk, about  
**any idea,**  
no matter how weedy and insignificant it  
may seem to you

# Do not be intimidated

Write a paper, and give a talk, about any idea, no matter how insignificant it may seem to you

- **Writing a paper is how you develop an idea in the first place**
- It usually turns out to be more interesting and challenging than it seemed at first

# The Purpose of Your Papers



AUBURN

UNIVERSITY

SAMUEL GINN  
COLLEGE OF ENGINEERING

**Xiao Qin**

*Department of Computer Science and  
Software Engineering*

*Auburn University*

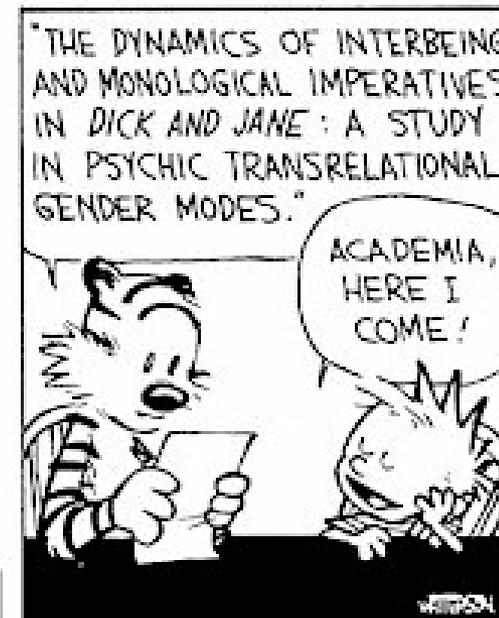
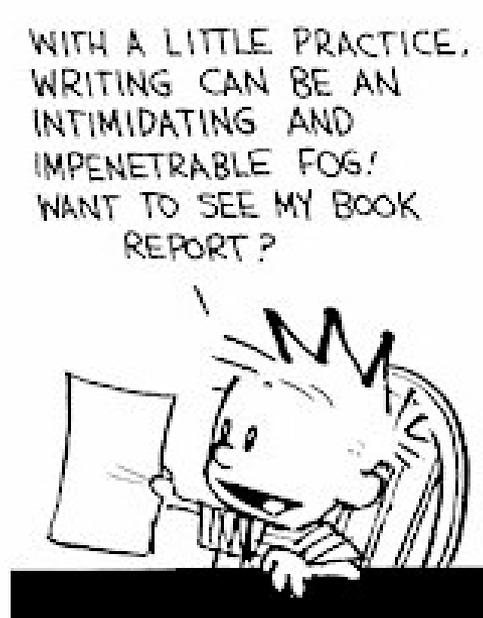
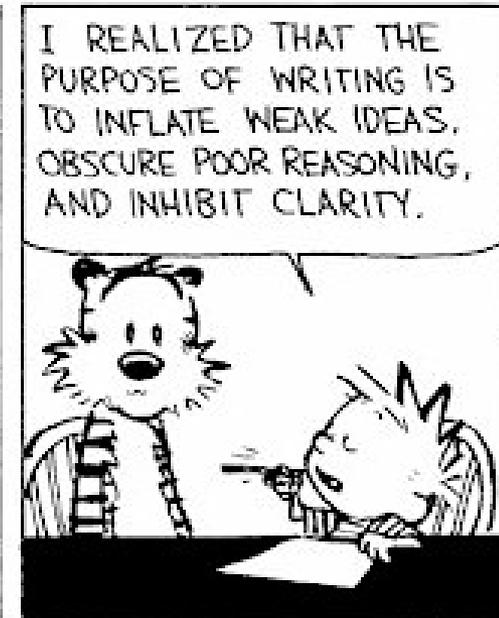
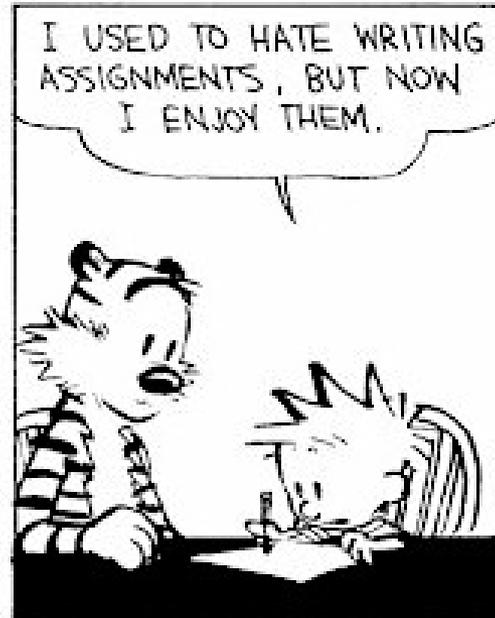
*<http://www.eng.auburn.edu/~xqin>*

*[xqin@auburn.edu](mailto:xqin@auburn.edu)*

# Why bother?

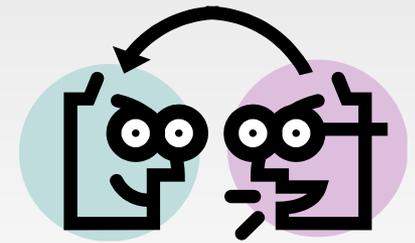
## Fallacy

we write papers and give talks mainly to impress others, gain recognition, and get promoted



# Papers Communicate Ideas

- Your goal: to infect the mind of your reader with **your idea**, like a virus
- Papers are far more durable than programs (think Mozart)



The greatest ideas are (literally) worthless if you keep them to yourself

# The Idea

## Idea

A re-usable insight,  
useful to the reader

- Figure out what your idea is
- Make certain that the reader is in no doubt what the idea is. Be 100% explicit:
  - “The main idea of this paper is....”
  - “In this section we present the main contributions of the paper.”
- Many papers contain good ideas, but do not distil what they are.



# The Idea – An example

## Abstract

In this paper, we present aggressive, proactive mechanisms that tailor file system resource management to the needs of I/O-intensive applications. In particular, we show how to use application-disclosed access patterns (hints) to expose and exploit I/O parallelism, and to dynamically allocate file buffers among three competing demands: prefetching hinted blocks, caching hinted blocks for reuse, and caching recently used data for unhinted accesses. Our approach estimates the impact of alternative buffer allocations on application execution time and applies cost-benefit analysis to allocate buffers where they will have the greatest impact. We have implemented informed prefetching and caching

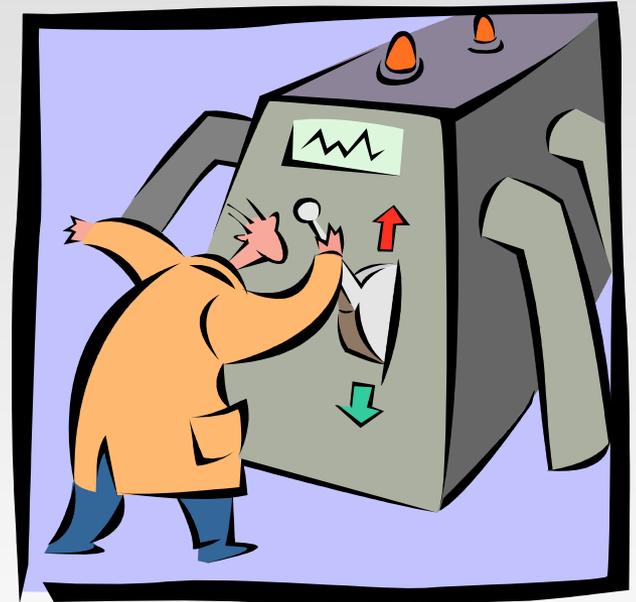
# One Ping

- Your paper should have just one “ping”: one clear, sharp idea
- Read your paper again: can you hear the “ping”?
- You may not know exactly what the ping is when you start writing; but you must know when you finish
- If you have lots of ideas, write lots of papers

Thanks to Joe Touch for “one ping”

# The purpose of your paper is not...

To describe the  
WizWoz  
system



- Your reader does not have a WizWoz
- Readers are **primarily interested in re-usable** brain-stuff, not executable artefacts

# Your Narrative Flow

- Here is a problem
- It's an interesting problem
- It's an unsolved problem
- **Here is my idea**
- My idea works (details, data)
- Here's how my idea compares to other people's approaches

I wish I knew how to solve that!

I see how that works. Ingenious!



# Structure (Conference Paper)

- Title (1000 readers)
- Abstract (4 sentences, 100 readers)
- Introduction (1 page, 100 readers)
- The problem (1 page, 10 readers)
- My idea (2 pages, 10 readers)
- The details (5 pages, 3 readers)
- Related work (1-2 pages, 10 readers)
- Conclusions and further work (0.5 pages)



# The Abstract

- I usually write the abstract last
- Used by program committee members to decide which papers to read
- Four sentences [Kent Beck]
  1. State the problem
  2. Say why it's an interesting problem
  3. Say what your solution achieves
  4. Say what follows from your solution

# Example

1. Many papers are badly written and hard to understand
2. This is a pity, because their good ideas may go unappreciated
3. Following simple guidelines can dramatically improve the quality of your papers
4. Your work will be used more, and the feedback you get from others will in turn improve your research

# Structure

- Abstract (4 sentences)
- **Introduction** (1 page)
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- Related work (1-2 pages)
- Conclusions and further work (0.5 pages)



# The introduction (1 page)

1. Describe the problem
2. State your contributions

...and that is all

**ONE PAGE!**

# Describe the Problem

Nowadays security is of critical importance for a wide range of real-time applications on clusters [4][5][6][11][27][36]. For example, in a real-time stock quote update and trading system, incoming requests from business partners and outgoing responses from an enterprise's back-end application have deadlines and security requirements, which have to be dealt with by a cluster located between the business partners and enterprise back-end applications [13]. Unfortunately, since clusters are built to execute a broad spectrum of unverified user-implemented applications from a vast number of different users, both applications and users can be sources of security threats to clusters [47]. For example, the vulnerabilities of applica-

Use an example to introduce the problem



# State Your Contributions

- Write the list of contributions first
- **The list of contributions drives the entire paper:** the paper substantiates the claims you have made
- Reader thinks “gosh, if they can really deliver this, that’s be exciting; I’d better read on”

# State Your Contributions

PRE-BUD has the goal of dynamically fetching data sets with the highest energy-savings into buffer disks. To accurately prefetch data blocks, information concerning future disk requests is indispensable. PRE-BUD can deal with both offline and online situations. In the offline case, PRE-BUD is provided with a priori knowledge of the list of disk requests. In the online case, PRE-BUD employs the look-ahead technique [14] that can furnish a window of future disk requests.

This research offers the following contributions. **First**, we are among the first to examine how to prefetch data blocks with maximum potential energy savings into buffer disks, thereby reducing the number of power-state transitions and increasing the number of standby periods to improve energy efficiency. **Second**, we build a new energy-saving prediction model, based on which an energy-saving calculation module was implemented for parallel I/O systems with buffer disks. Energy savings measured by the prediction model represent the importance and priority of prefetching blocks in a buffer disk to efficiently conserve energy in the disk system. **Third**, we developed an energy-efficient prefetching algorithm in the context of two buffer disk configurations. A greedy prefetching module was implemented to fetch blocks that have the highest energy savings. **Finally**, we construct models to theoretically and experimentally analyze the energy efficiency and performance of PRE-BUD. We quantitatively compared PRE-BUD with three existing techniques employed in parallel I/O systems.

Bulleted list of contributions

**Do not leave the reader to guess what your contributions are!**

# Bulleted List and Forward References

The main contributions of this work are listed below:

- We describe an offloading framework that is able to be applied to either a newly developed or an existing data-intensive application with merely slight effort in Sec. II.
- We discuss in details a couple of design and implementation issues of an offloading application in Sec. IV.
- We apply the offloading framework to a group of widely used applications in Sec. V.
- We evaluate experimental results to show that they perform much better than original versions in Sec. VI.

# Structure

- Abstract (4 sentences)
- Introduction (1 page)
- ~~Related work~~
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- Related work (1-2 pages)
- Conclusions and further work (0.5 pages)

# Contributions should be refutable

NO!	YES!
<p>We describe the PRE-BUD system. It is really cool.</p>	<p>...to examine how to prefetch data blocks with maximum potential energy savings into buffer disks, thereby reducing the number of power-state transitions and increasing the number of standby periods to improve energy efficiency.</p>
<p>We study an energy consumption model.</p>	<p>We build a new energy-saving prediction model, based on which an energy-saving calculation module ... (see Section 4.2)</p>
<p>We developed a prefetching algorithm.</p>	<p>We developed an energy-efficient prefetching algorithm in the context of two buffer disk configurations. A greedy prefetching module ... (see Section 4.1)</p>

# No “rest of this paper is...”

- Not: 

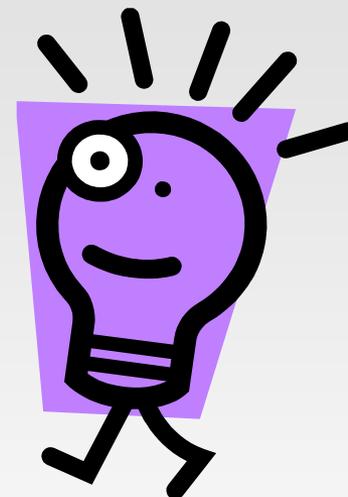
“The rest of this paper is structured as follows. Section 2 introduces the problem. Section 3 ... Finally, Section 8 concludes”.
- Instead, **use forward references from the narrative in the introduction.**  
The introduction (including the contributions) should survey the whole paper, and therefore forward reference every important part.

# No related work yet!



Your reader

Related  
work



Your idea

Energy efficient prefetching was explored by Papathanasiou and Scott [20]. Their techniques relied on changing prefetching and caching strategies within the Linux kernel. PB-LRU is another energy efficient cache management strategy [32]. This strategy focused on providing more opportunities for underlying disk power strategies to save energy. Flash drives have also been proposed for use as buffers for disk systems [4].

# No related work yet

- **Problem 1:** the reader knows nothing about the problem yet; so your (carefully trimmed) description of various technical tradeoffs is absolutely incomprehensible
- **Problem 2:** describing alternative approaches gets between the reader and your idea

I feel stupid



I feel tired



AUBURN  
UNIVERSITY

SAMUEL GINN  
COLLEGE OF ENGINEERING

# For readers who know your field very well...

- Abstract (4 sentences)
- Introduction (1 page)
- Related work (1-2 pages)
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- ~~Related work~~
- Conclusions and further work (0.5 pages)

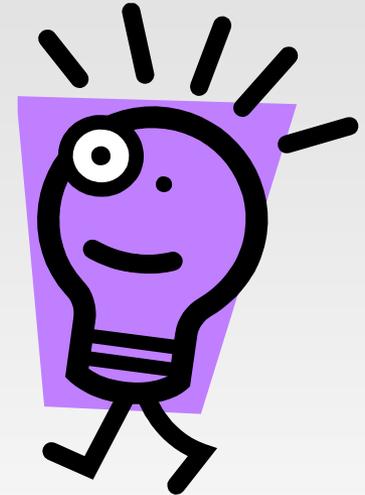
# Related work first!



Your reader

## Related Work

Is your idea novel?  
What are your contribution?



Your idea

Energy efficient prefetching was explored by Papathanasiou and Scott [20]. Their techniques relied on changing prefetching and caching strategies within the Linux kernel. PB-LRU is another energy efficient cache management strategy [32]. This strategy focused on providing more opportunities for underlying disk power strategies to save energy. Flash drives have also been proposed for use as buffers for disk systems [4].

# Where should you put the related work?

- Abstract (4 sentences)
- Introduction (1 page)
- **Put related work here?**
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- **Put related work here?**
- Conclusions and further work

# Where should you put the related work?

## A Suggested Principle

- Abstract (4 sentences)
- Introduction (1 page)
- **Put related work here?**
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- **Put related work here?**
- Conclusions and further work

You want to convince reviewers in the first place.

You feel this is a novel idea



# Structure

- Abstract (4 sentences)
- Introduction (1 page)
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- Related work (1-2 pages)
- Conclusions and further work (0.5 pages)

# Presenting the idea

## 3. The idea

Consider a bifurcated semi-lattice  $D$ , over a hyper-modulated signature  $S$ . Suppose  $p_i$  is an element of  $D$ . Then we know for every such  $p_i$  there is an epi-modulus  $j$ , such that  $p_j < p_i$ .

- Sounds impressive...but
- Sends readers to sleep
- In a paper you **MUST** provide the details, **but FIRST convey the idea**

# Presenting the idea

- Explain it as if you were speaking to someone using a whiteboard
- **Conveying the intuition is primary**, not secondary
- Once your reader has the intuition, she can follow the details (but not vice versa)
- Even if she skips the details, she still takes away something valuable

# Putting the reader first

- **Do not** recapitulate your personal journey of discovery. This route may be soaked with your blood, but that is not interesting to the reader.
- Instead, choose the most direct route to the idea.

# The payload of your paper

Introduce the problem, and  
your idea, using

**EXAMPLES**

and only then present the  
general case

# Using examples

## 3. Motivational Example

For a simple motivational example that demonstrates the utility of the buffer disk architecture, we present a scenario that is depicted in Fig. 1. Each horizontal bar represents the time a particular disk is busy or idle. Fig. 1 presents requests for individual disks that are represented by the specific colors

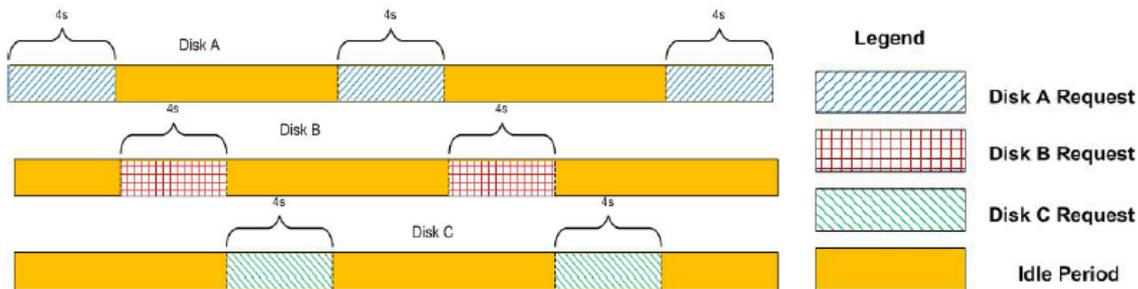


Fig. 1 Sample Disk Trace

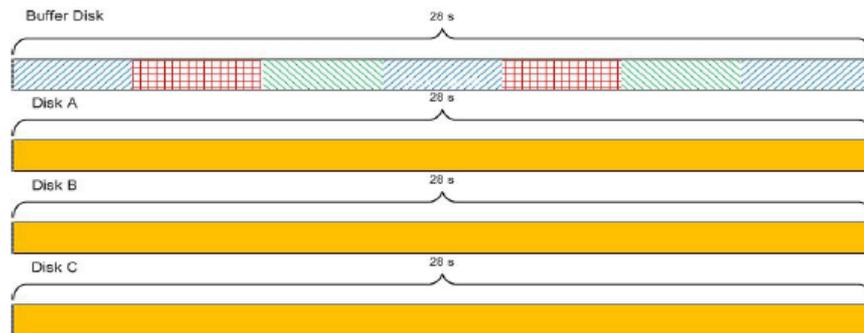


Fig. 2 Buffer Disk Added to Architecture

Example  
right away

# The details: evidence

- Your introduction makes claims
- The body of the paper provides **evidence to support each claim**
- Check each claim in the introduction, identify the evidence, and forward-reference it from the claim
- Evidence can be: analysis and comparison, theorems, measurements, case studies

# Structure

- Abstract (4 sentences)
- Introduction (1 page)
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- **Related work** (1-2 pages)
- Conclusions and further work (0.5 pages)

# Related work

**Fallacy** To make my work look good, I have to make other people's work look bad

# The truth: credit is not like money

Giving credit to others does not diminish the credit you get from your paper

- Warmly acknowledge people who have helped you
- Be generous to the competition. “In his inspiring paper [Foo98] Foogle shows.... We develop his foundation in the following ways...”
- Acknowledge weaknesses in your approach



# Credit is not like money

Failing to give credit to others  
can kill your paper

If you imply claim that an idea is yours, and the referee knows it is not, then either

- You don't know that it's an old idea (bad)
- You do know, but are pretending it's yours (very bad)

If reviewers pointed out that your idea is not novel, what can you do?

# Structure

- Abstract (4 sentences)
- Introduction (1 page)
- The problem (1 page)
- My idea (2 pages)
- The details (5 pages)
- Related work (1-2 pages)
- Conclusions and further work (0.5 pages)

# Conclusions and further work

- Be brief.

# Future Work

## Fallacy

(1) Do not let others compete against me using my future work ideas.

(2) Point out future directions that seems to be a dead-end.

(3) A future work I believe I can complete sooner than the others.

# Future Work (cont.)

- Refer to the weakness of my work
- Encourage other researchers to continue working on your project
- Issues that you plan to address

There are three possible **future research directions** for extending PRE-BUD. **First**, we will improve the scalability of PRE-BUD by adding more than one buffer disk to the parallel I/O system. This can be implemented by considering a buffer disk controller which manages various buffer disks each responsible for a set of data disks. In this work we investigate the relationship between buffer disks and data disks, to improve the parallelism of PRE-BUD **we need to investigate** the relationship between a buffer disk controller and the buffer disks. The number of buffer disks will have to be increased as the scale of the disk system is increased. Second, PRE-BUD will be integrated with the dynamic speed control or DRPM [9] for parallel disks. **Last but not least**, we will quantitatively study the reliability impacts of PRE-BUD on parallel I/O systems.

# Software Tools

- Latex
  - MikTex
  - TexMaker
- Figures: Matlab
- Diagram: Matlab, visio, google docs

# Summary

If you remember nothing else:

- **Identify your key idea**
- **Make your contributions explicit**
- **Use examples**
- **Download the slides from**

<http://www.slideshare.net/xqin74/how-to-write-papers-part-1-principles>

A good starting point:

“Advice on Research and Writing”

<http://www-2.cs.cmu.edu/afs/cs.cmu.edu/user/mleone/web/how-to.html>

Google: **Xiao Qin**

<http://www.eng.auburn.edu/~xqin>

Google

Xiao Qin

× Search

About 4,170,000 results (0.27 seconds)

[Advanced search](#)

[Xiao Qin, Associate Professor of Computer Science, Auburn University](#) ★

Sep 3, 2010 ... UNL CSE graduate **Xiao Qin** received an NSF CAREER award in 2009 to investigate parallel disks that put substantial multicore computing power ...

[www.eng.auburn.edu/~xqin/](http://www.eng.auburn.edu/~xqin/) - [Cached](#) - [Similar](#)

# My webpage

<http://www.eng.auburn.edu/~xqin>

Home Teaching **Research** Publications Students Software Useful Links

- Current Projects
- Completed Projects
- Research Funding
- Research Sponsors
- Presentations**

**Xiao Qin**     

**Associate Professor of Computer Science**

If you are wondering how to pronounce my name click here

Shelby Center for Engineering Technology, Suite 3101E  
Department of Computer Science and Software Engineering  
Samuel Ginn College of Engineering  
Auburn University, AL 36849-5347  
Office: 334-844-6327 Fax: 334-844-6329  
[xqin@auburn.edu](mailto:xqin@auburn.edu)

» [curriculum vitae](#)

**Latest News** [Archive >>](#)

**Focusing on Data Storage Systems**

**BUD: A Buffer-Disk Architecture for Energy-Efficient Parallel Disks**

**Jiang Defends and Heads to Earlham College**

# Download Slides at slideshare

<http://www.slideshare.net/xqin74>

Home

Teaching

Research

Publications

Students

Software

Useful Links

## Xiao Qin's Presentations



### Presentations

- Current Projects
- Completed Projects
- Research Funding
- Research Sponsors
- Presentations



Download our slides from  
**SLIDESHARE PRESENTATIONS**



Download the source code of  
our data placement module for

1. Energy-Efficient Data Storage Systems. [ [PPT](#) | [Slideshare](#) ]
2. HDFS-HC: A Data Placement Module for Heterogeneous Hadoop Clusters. Note: This presentation is based on our paper - [Improving MapReduce Performance via Data Placement in Heterogeneous Hadoop Clusters](#) published in Proc. 19th Int'l Heterogeneity in Computing Workshop, Atlanta, Georgia, April 2010. [ [PPT](#) ]
3. "An Application-Oriented Approach for Computer Security Education," Invited Talk at the Information Security and Computer Applications Conference, Feb. 25, 2011. [ [PPT](#) ]
4. Energy Efficient Prefetching with Buffer Disks for Cluster File Systems. IEEE

Xiao Qin

Analytics Newsfeed Presentations Documents Videos Following Favorites More



Username xqin74

- Post on wall Private message

Presentations 11 Public 11 Private 0

Tags (23)

application-oriented clusters computer security data placement disk systems education energy efficiency energy efficient going to graduate school hdfs heterogeneous cluster high-performance improve mapreduce papers read real-time parallel applications research papers resource management scheduling secure security-aware write

Presentations

Latest Most viewed



Improve energy efficiency... 2 months ago, 103 views Edit



Security-Aware Scheduling... 2 months ago, 80 views Edit



Going to graduate school 2 months ago, 51 views Edit



Energy efficient resource management 2 months ago, 96 views Edit



How to read papers 2 months ago, 56 views Edit



HDFS-HC: A Data Placement Module 2 months ago, 288 views Edit



How to write papers, Part 2 2 months ago, 52 views Edit



How to write papers, Part 1 2 months ago, 65 views Edit



An Application-Oriented Approach for Computer Security Education 2 months ago, 107 views Edit



Energy Efficient Prefetching 11 months ago, 263 views Edit



How to Succeed in the AU REU Program? 12 months ago, 105 views Edit

# Questions

