BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY
Office of Science

COMPUTING SCIENCES SUMMER PROGRAM 2023

Literature Surveys and Reviews: Where do we stand?

Jean Luca Bez
Scientific Data Division – LBNL

BERKELEY LAB

Bringing Science Solutions to the World

Jean Luca Bez

Data Management Research Scientist
Scientific Data Division
jlbez@lbl.gov
jeanbez.github.io

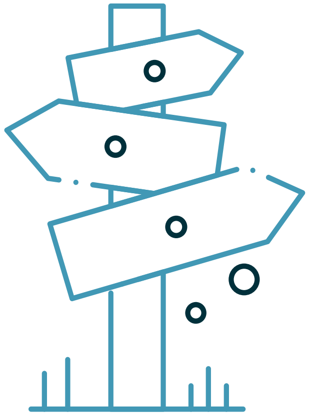
Jean Luca is a Career-Track Researcher in the Scientific Data Management Group at Lawrence Berkeley National Laboratory (LBNL), USA. He is passionate about High-Performance I/O, Parallel I/O, Education, and Competitive Programming. His research focuses on optimizing the I/O performance of scientific applications at the middleware level by exploring I/O Forwarding, I/O Scheduling, and Automatic Tuning and Reconfiguration using Machine Learning techniques.


"If I have seen further, it is by standing on the shoulders of Giants"
Isaac Newton

What are we covering?

Reviewing Manuscripts

Positioning to Related Work



Writing Literature Surveys

Reviewing Manuscripts

Reviewing Manuscripts

What's this all about?

- Allows you to **form** your own scientific opinion
- Develop **critical thinking** skills
- Get access to super fresh research firsthand
- Gain **insight** into other authors' argument structure
- **Improve** your writing and communication skills

- It definitely takes **effort** and **time**, but in the end it will be **worth it**!
- What are the main steps?
  a. Accepting a paper to review
  b. Reviewing a paper
  c. Submitting a review

Reviewing Manuscripts

Accepting an Invite

- Check if it **belongs** to your field of expertise
  - If out of your scope, most probably, your judgment will not be fair
  - You will spend longer time to understand and review
- Check for **conflict of interest**
- Check if you do have **time**
  - Review process often happens in a tight schedule
- To keep in mind:
  - Put yourself in the author's position!
  - **What type of review would you like to receive?**


ATTENTION
FROM NOW ON, ALL PAPER REVIEWS WILL BE DONE WITH FACEBOOK EMOTICONS
Cham, et al. submitted a New Paper
Oct. 24 at 6:07am
3 people reviewed this
LIKE
"I LIKED YOUR PAPER BUT I'M NOT EXCITED ABOUT IT. IT'S GOOD WORK BUT NOT GROUNDBREAKING.
HEART
"I LOVED YOUR PAPER AND WANT TO SEE IT PUBLISHED RIGHT AWAY!"
LOL
"DID YOU SERIOUSLY THINK THIS PAPER WOULD GET ACCEPTED IN THIS CONFERENCE OR JOURNAL? LOL."
SAD FACE
"THIS PAPER IS SO BAD, I'M LOSING MY FAITH IN ACADEMIA."
ANGRY FACE
"GRRR, I WISH I HAD WRITTEN THIS PAPER."
SURPRISE
"WHOA. YOU GOT FUNDING TO DO THIS??"
JORGE CHAM, CARLOS R-M ©2016
WWW.PHDCOMICS.COM

Reviewing Manuscripts

Reviewing a Paper

A review is **not** about whether one **likes** a certain piece of work, but whether the research is **valid** and tells us something **new!**

"A thoughtful, well-presented evaluation of a manuscript, with tangible suggestions for improvement and a recommendation that is supported by the comments, is the most valuable contribution that you can make as a reviewer, and such a review is greatly appreciated by both the authors of the manuscript and the editors of the journal." — American Chemical Society Reviewer Lab

Reviewing Manuscripts

Reviewing a Paper

- Don’t start when you are not in a good mood
- Be responsible and do it in time
- Be **direct** and **concise**
- Be open to new ideas, think about the **impact** of the paper in the field
- Complicated papers are not necessarily of good quality
- Complicated and colorful graphs are not an indication of good results
- Your writing style is yours!
- Do **not** be **rude**! It is ok to be tough sometimes!
- Focus on the **content** and **idea** (not the authors, institution, or country)
- Fast screening review (presentation, quality, references) and detail review
- Be **specific** (especially if you ask revisions) and don’t give general comments
- Check the citation of the references

Reviewing Manuscripts

How should I structure my review?

- **Summary**
  - Demonstrate you understood the work you read
- **Strengths and Weaknesses**
- **Detailed Comments**
  - Major issues (methodology, analysis, conclusions, etc)
  - Minor issues (formatting, figures, etc)
- **Decision**
  - Remember: the scores can vary from venue to venue
  - Additional factors can be scored (relevance, novelty, presentation, etc)

Positioning to Related Work

Positioning to Related Work

What are the steps?

Step 1 Define your **research scope**

Step 2 Identify the literature

Step 3 **Critically analyze** the literature

Step 4 **Categorize** your resources

Step 5 Write your literature review

Positioning to Related Work

Step 1 Define your research scope

- The topic needs to be **identified** and **defined** as **clearly** as possible
- What is the specific **research question**?
- Identify **keywords** that you will be using to search for relevant research

Positioning to Related Work

Step 2 Identify the literature

- Use a range of **keywords** to search databases
  - ACM Digital Library, IEEE Xplore, Google Scholar, Scopus, etc.
- Keep in mind that **peer-reviewed** articles are considered to be the “gold standard”
  - Prefer peer-reviewed papers from high quality conferences
- Read through **titles** and **abstracts**
  - Abstracts will help save time while looking for relevance
- **Select** and obtain articles
- Make sure you cover the **latest publications** as well!
  - These are often the most important ones

Positioning to Related Work

Step 3 Critically analyze the literature

- **Takes notes, take notes, and take notes!**
  - Rely on tools to help index (Mendeley, Zotero, EndNote, etc)
  - Make sure your notes and keywords are **searchable**!
  - It will save you time (now and in the future)
- Read and **summarize**
  - Coverage, methodology, and relationship to other works
  - Analyze relationships, major themes, and any **critical gaps**
- Sample (non-exhaustive) questions:
  - What was the research question of the study you are reviewing?
  - What were the authors trying to discover?
  - What were the research methodologies?
  - What further questions does it raise?
  - Are there are conflicting studies?
  - Has this study been cited or reproduced?

Positioning to Related Work

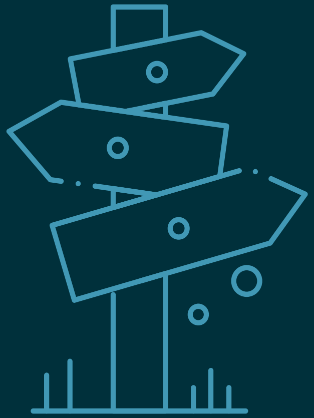
Step 4 Categorize your resources

- How can you **group** related work?
  - Rely on tools to help index (Mendeley, Zotero, EndNote, etc)
  - Make sure your notes and keywords are **searchable**!
- Look for **patterns** and by developing **subtopics** to categorize:
  - Chronology, theme, methodology, theoretical/experimental approach
  - Findings that are common/contested
  - Important trends in the research
  - The most influential theories
  - The existing limitations and gaps
  - Develop headings/subheadings that reflect the major themes and patterns

Positioning to Related Work

Step 5 Write your literature review

- Write a one or two sentence statement summarizing the work
- **Prioritize analysis over description**
- A good review of the literature does **NOT**:
  - Simply reference and list all of the material you have cited in your paper
  - Present material that is not directly relevant to your study
  - Starting a literature review with “A number of scholars have studied…” and simply listing who has studied the topic and what each scholar concluded is not going to strengthen your paper
- Make sure you:
  - **Summarize** the most relevant and important aspects
  - **Synthesize** what has been done in this area of research and by whom
  - **Highlight** what previous research indicates about a topic
  - Identify potential **gaps** and areas of disagreement in the field
  - **Make you clearly state how you compare to each work you cited!**



Writing
Literature Surveys

Writing Literature Surveys

What are the steps?

Step 1 Define your **scope** and **methodology**

Step 2 Identify the literature

Step 3 **Critically analyze** the literature

Step 4 **Categorize** your resources

Step 5 Write your literature survey

Writing Literature Surveys

A little bit more about the process…

“Ten Simple Rules for Writing a Literature Review”

1. Define a Topic and Audience
2. Search and “Re-search” the Literature
3. Take Notes While Reading
4. Choose the Type of Review You Wish to Write
5. Keep the Review Focused, but Make It of Broad Interest
6. Be Critical and Consistent
7. Find a Logical Structure
8. Make Use of Feedback
9. Include Your Own Relevant Research, but Be Objective
10. Be Up-to-Date, but Do Not Forget Older Studies

Pautasso M. Ten simple rules for writing a literature review. PLoS Comput Biol. 2013;9(7):e1003149. doi: 10.1371/journal.pcbi.1003149. Epub 2013 Jul 18. PMID: 23874189; PMCID: PMC3715443.

Writing Literature Surveys

Outline your literature structure

- **Chronological**
  - Trace the development of the topic over time
  - Be careful to avoid simply listing and summarizing sources in order
  - Focus on patterns, turning points and key debates that have shaped the direction of the field
- **Thematic**
  - Recurring central themes
  - Organize your literature review into subsections that address different aspects of the topic
- **Methodological**
  - Focus on comparing the results and conclusions that emerge from different approaches
  - Look at what results have emerged in qualitative versus quantitative research
  - Discuss how the topic has been approached by empirical versus theoretical research
- **Theoretical**
  - Discuss various theories, models, and definitions of key concepts
  - Argue for the relevance of a specific theoretical approach

Writing Literature Surveys

file system comprises two types of servers with distinct roles: the data servers and the metadata servers. The latter handles information about the files (e.g., sizes and permissions) and their location in the system. Lustre [80, 210], IBM Spectrum Scale (previously known as GPFS) [192], BeeGFS [94], etc. are commonly used parallel file systems on large-scale HPC systems. To achieve high performance, these file systems harness parallelism by using *data striping* [207] which consists of partitioning the files and distributing the data into fixed-size chunks across multiple storage nodes. Finally, the PFS servers provide a logical file system abstraction over diverse storage devices such as Hard Disk Drives (HDDs), Solid State Drives (SSDs), or Redundant Array of Independent Drives (RAID).

Summary #1

The multi-layered software and hardware HPC I/O stack is complex. To access data in HPC systems, applications issue requests that, while traversing the I/O stack, are reshaped via a series of data transformations. These originate from distinct abstractions and mappings between the data models used in each layer combined with optimization techniques applied before reaching the file system and eventually the storage hardware.

In the following sections, we discuss the I/O access patterns observed in the HPC stack's layers, from application data models and their I/O requests percolating through the underlying layers until the file systems handle them.

3 APPLICATION DATA MODELS AND ACCESS PATTERNS

Scientific applications often use data abstractions provided by high-level libraries (e.g., HDF5, NetCDF, ADIOS) to express data structures more naturally to a problem and domain. HPC simulations often describe their data objects using multi-dimensional data or meshes, arbitrary subsets, points and curves, and key values [125, 200]. Mesh data objects, in particular, can be further represented by structured rectilinear, non-uniform rectilinear, grid-less points, structured (curvilinear), arbitrary polyhedral, constructive solid geometry (CSG), unstructured zoo (UCD), and adaptive mesh refinement (AMR) meshes. In Figure 2, we show these most common high-level data models used by HPC scientific applications.

a dataspace. In our experiments, for a shared file, IOR defines the start offset as `offset` module `segmentSize`, count as one, and a stride and a block equal to the transfer size, i.e., 4MB. However, once the requests reach the MPI-IO layer, they are further broken down by the four collective aggregators into a larger number of 1MB POSIX requests, considering the underlying parallel file system striping configuration before sending them to each storage device. We have defined Lustre to use 1MB stripes over eight servers to make it easier to visualize. Once we delve into lower levels of the I/O stack, we are to lose contextual information from the applications and start to observe the effect of natural interference in this shared storage infrastructure. For instance, if we glance at one OST, the requests arrive at the storage servers in an interleaved fashion, coming from the two applications that the file system is unaware of. At this point, the original contiguous requests issued by the application using 4MB requests arrive at the server much smaller (in 1MB requests) and with a different spatiality (non-contiguous).

Furthermore, it is essential to highlight the inter-application interference caused by other applications sharing those data servers. Figure 15 clearly depicts how two identical applications that started simultaneously begin to diverge in time towards the end of our experiment. Such observation also highlights the importance of taking into account temporal features when discussing access patterns.

Summary #8

Different tools extract and visualize I/O access patterns from coarse-grain profilers to fine-grain traces as I/O requests pass through the stack. However, we could not find a complete solution that allows observing patterns and all of their transformations in the context of each layer. Because of the complexity of the current stack, this gap might not easily reflect the root causes of bottlenecks.

10 CONCLUSION

The HPC I/O stack has been complex due to multiple layers of hardware and software, their various tuning options, and inter-dependencies among the layers. This survey discussed extensively the overloaded "I/O access pattern" terminology used to describe how accesses are done from the major layers of the HPC I/O stack, covering the high-level models used by scientific applications, how those are represented by high-level I/O libraries and translated by middleware libraries before reaching the parallel file system. We have also highlighted I/O benchmarks and kernels employed to exercise access patterns in different levels, alongside existing tools to visualize those patterns using profiling and tracing.

Harnessing the I/O community's knowledge over the last 20 years, we surveyed 146 papers from ACM DL and IEEE Xplore to propose a baseline taxonomy that could define an application's I/O access patterns. Our effort targets bringing

Writing Literature Surveys

Interface	Opportunities	Challenges
POSIX	• Portability • Strong consistency guarantees • Shallow learning curve (wide adoption)	• Not designed for HPC • Strong consistency vs. scalability • Collective access (locking) • Optimizations (lack of whole application view)
MPI-IO	• Designed for HPC • Relaxed consistency • Flexibility (express patterns natural to applications) • High-level I/O optimizations	• Hard adoption (source-code changes) • Complex tunning
STDIO	• Simple and buffered stream interface	• Not designed for HPC • Scalability • Optimizations (lack of whole application view)

Table 2. Opportunities and challenges presented by each I/O interface when used in the context of HPC.

```
"query": {
    AllField: (("I/O access pattern" OR "I/O characterization" OR "I/O characteristic" OR "I/O signature") AND "HPC"))
}
"filter": {Publication Date: (01/01/2000 TO 12/31/2021)}, {ACM Content: DL}
```

Fig. 11. ACM DL query and filter parameters used in this survey.

```
("Full Text & Metadata":"I/O access pattern") AND ("Full Text & Metadata":"HPC") OR
("Full Text & Metadata":"I/O characterization") AND ("Full Text & Metadata":"HPC") OR
("Full Text & Metadata":"I/O characteristic") AND ("Full Text & Metadata":"HPC") OR
("Full Text & Metadata":"I/O signature") AND ("Full Text & Metadata":"HPC")          Filters Applied: 2000 - 2021
```

Fig. 12. IEEE Xplore query and filter parameters used in this survey.

Name	Synthetic	Kernel(s)	Data	Metadata	Write	Read	Request Size	Independent	Collective	Temporal	Shared File	File-per-process	Synchronous	Asynchronous	Interface
IOR [95]	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	POSIX, MPI-IO, HDF5, HDFS, S3, NCMPI, IME, MMAP, RADOS
MADbench2 [26]	✗	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	POSIX, MPI-IO
IFER [247]	✓	✗	✓	✗	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	MPI-IO
S3D [40]	✗	✓	✓	✗	✓	✓	✗	✗	✓	✗	✓	✗	✓	✓	PnetCDF
NAS BT-IO [168]	✓	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	✓	✗	MPI-IO
S3aSim [43]	✗	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	MPI-IO
h5bench [131]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	HDF5
HACC-IO [227]	✗	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✗	POSIX, MPI-IO
MACSio [125]	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	STDIO, MPI-IO, HDF5
MPI Tile I/O [188]	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	POSIX, MPI-IO

Table 4. Summary of access pattern features exercised by each benchmark and I/O kernel. The check in orange indicates that h5bench does support asynchronous operations, however, it requires the HDF5 ASYNC VOL Connector [215] to be available and enabled.

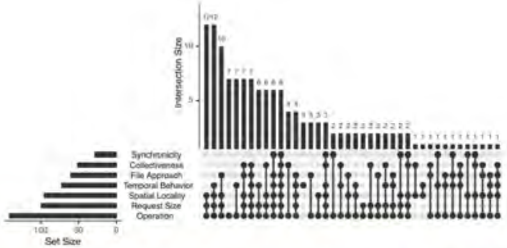


Fig. 14. Analysis of the set of features used to describe an I/O access pattern in ACM DL and IEEE Xplore research papers.

BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY
Office of Science

COMPUTING SCIENCES SUMMER PROGRAM 2023

Literature Surveys and Reviews: Where do we stand?

Jean Luca Bez
Scientific Data Division – LBNL

Resources

https://www.science.org/content/article/how-review-paper

https://www.journals.elsevier.com/applied-soft-computing/news/tips-and-advice-when-you-review-a-scientific-paper

https://mitcommlab.mit.edu/broad/commkit/peer-review

https://authorservices.wiley.com/Reviewers/journal-reviewers/how-to-perform-a-peer-review/step-by-step-guide-to-reviewing-a-manuscript.html

https://www.sae.org/participate/volunteer/reviewer-resources/writing-guide

https://libguides.uwf.edu/c.php?g=215199&p=1420520

https://guides.lib.uoguelph.ca/c.php?g=130964&p=5000948

https://psychology.ucsd.edu/undergraduate-program/undergraduate-resources/academic-writing-resources/writing-research-papers/writing-lit-review.html#6.-Incorporate-the-literature-r

https://lib.arizona.edu/research/write-cite/lit-review

https://blog.wordvice.com/how-to-write-a-literature-review/

https://www.nature.com/articles/d41586-022-01878-7

https://libguides.northwestern.edu/howtochoose

https://owl.purdue.edu/owl/research_and_citation/conducting_research/writing_a_literature_review.html

https://www.scribbr.com/methodology/literature-review/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715443/

https://www.ncbi.nlm.nih.gov/books/NBK481583/

https://equity.ucla.edu/know/implicit-bias/