

Publication patterns in HEP computing

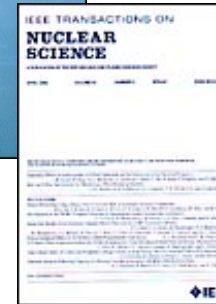
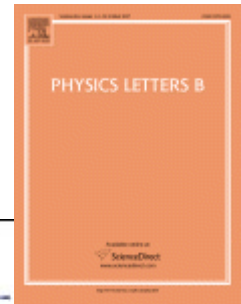
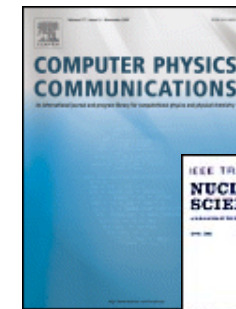
M. G. Pia¹, T. Basaglia²,
Z. W. Bell³, P. V. Dressendorfer⁴

¹INFN Genova, Genova, Italy

²CERN, Geneva, Switzerland

³ORNL, Oak Ridge, TN, USA

⁴IEEE, Piscataway, NJ, USA



NSS 2012
Anaheim, CA



Analysis topics

● General tools

- Geant4
- ROOT

● HEP experiments

- LEP
 - ALEPH, DELPHI, L3, OPAL
- CDF
- BaBar
- LHC
 - ALICE, ATLAS, CMS, LHCb, TOTEM

● Grid computing

- LCG

- What they publish
- How much
- Where
- Citations
- Technology vs physics
- Software vs hardware
- Software/DAQ-trigger

**Representative
Not exhaustive**

No time to report all the results

Data sources

● Thomson-Reuters: ISI Web of Knowledge

- CERN subscription: since 1970, conference database not included
- Search by keywords, collaboration name

● Journal web sites

- IEEE TNS
- NIM, Comp. Phys. Comm. (*Elsevier*)
- JINST (*IOP/SISSA*)
- Full-text searches

● CERN databases

- CERN Document System
- Greybook

● Years: 1982-2011 (*LEP*), 1992-2011 (*BaBar*, *LHC*)

- Reproducible sample for citation analysis

● Publication analysis extended to 30 September 2012

Data sample

● Contamination

- Non-pertinent entries in the data sample

● Omission

- Pertinent papers are not included in the data sample

⇒ Cross-checks

- WoS/CDS, WoS/publishers' web sites

● WoS inconsistencies and errors

- Total number of citations includes Conference database
- Proceedings papers: false classifications and omissions
- ⇒ Manually corrected whenever possible

● Automated analysis (whenever possible)

● Manual evaluation: abstracts and full-text papers

- Some degree of **subjectivity**

Geant 4

S. Agostinelli et al.

Geant4: a simulation toolkit

NIM A, vol. 506, no. 3, pp. 250-303, 2003

3301 citations

(20 October 2012)

Most cited CERN publication in WoS

(excluding Rev. Part. Properties)

J. Allison et al.

Geant4 Developments and Applications

IEEE Trans. Nucl. Sci., vol. 53, no. 1, pp. 270-278, 2006

665 citations

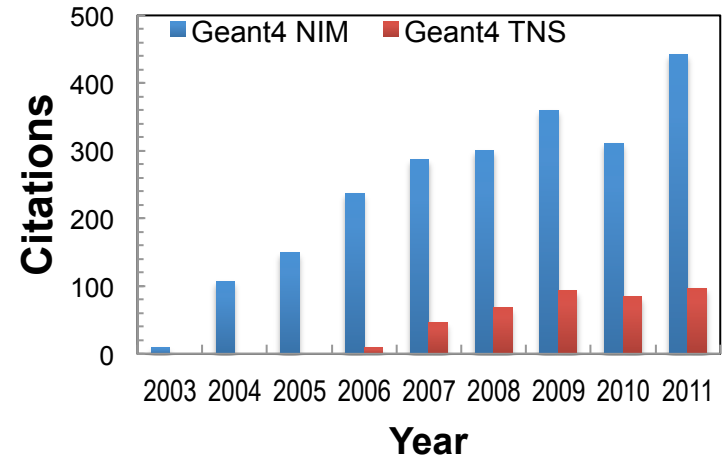
(20 October 2012)

Many papers cite the NIM paper, but they omit citing the TNS one, even though both are indicated in <http://cern.ch/geant4>

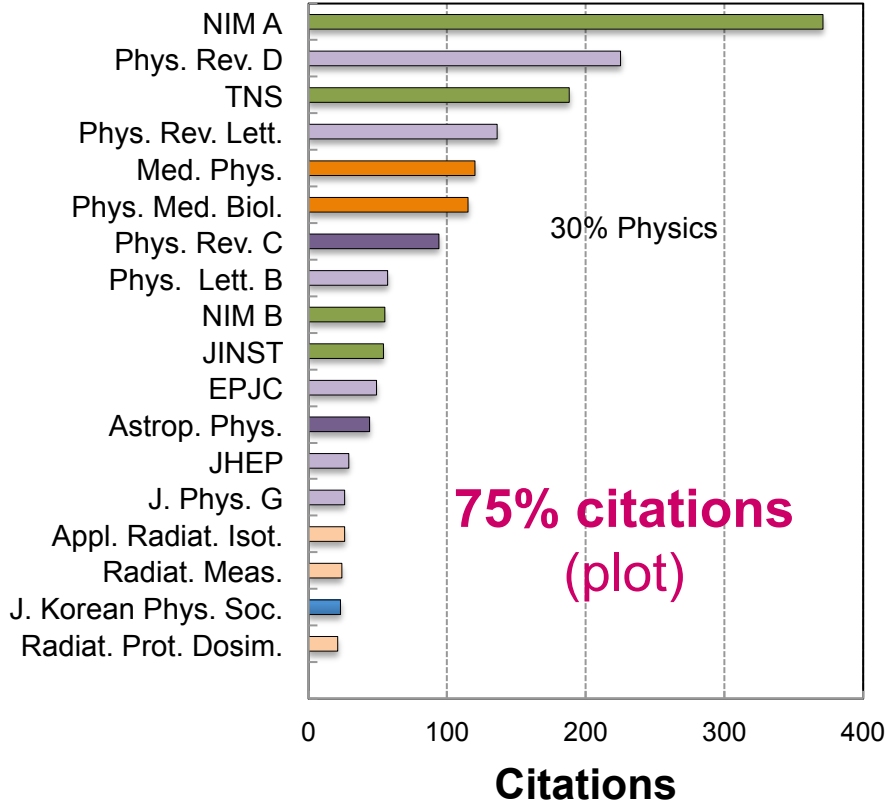
Many papers that use Geant4 do not cite either reference

Geant 4

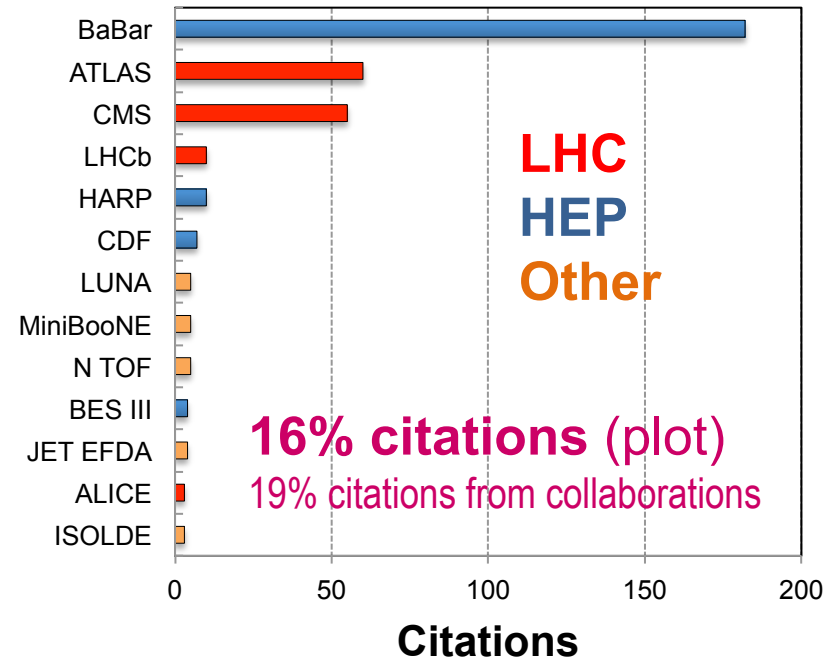
Born from LHC experimental requirements
Multidisciplinary sources of citations



Geant4 NIM: Citing Journals

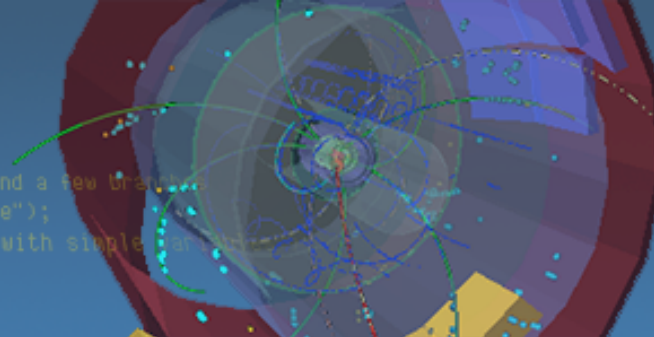


G4 NIM: Citing Collaborations



ROOT

```
//create the file, the Tree and a few branches
TFile f("tree1.root","recreate");
TTree t1("t1","a simple Tree with simple branches");
t1.Branch("px",&px,"px/F");
t1.Branch("py",&py,"py/F");
```



R. Brun and F. Rademakers

ROOT - An object oriented data analysis framework

NIM A, vol. 389, no. 1-2, pp. 81-86, 1997

584 citations

(20 October 2012)

AIHENP Workshop **proceedings paper**

I. Antcheva et al.

**ROOT - A C++ framework for petabyte data storage,
statistical analysis and visualization**

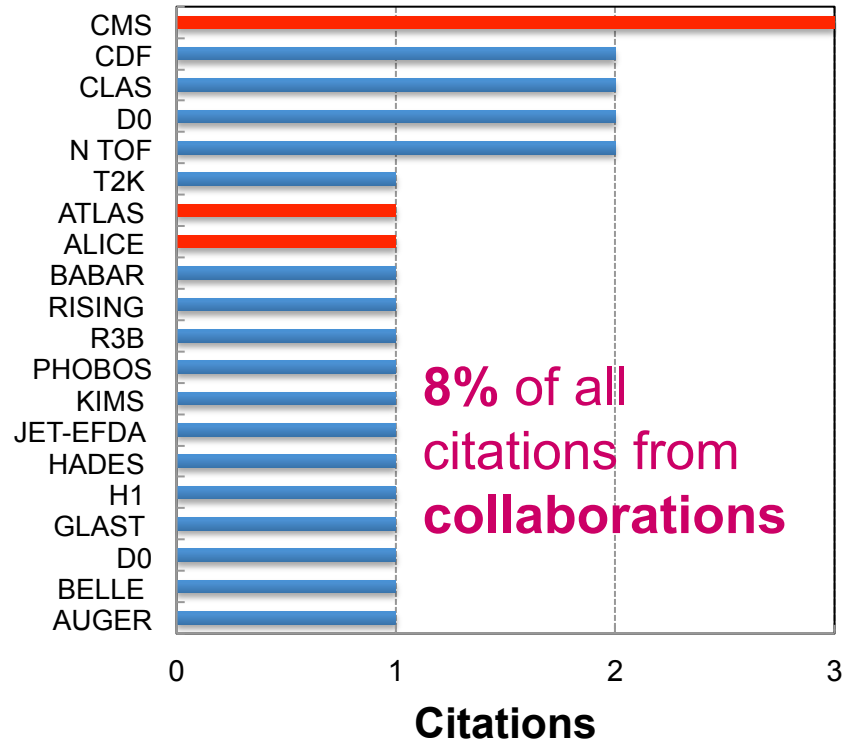
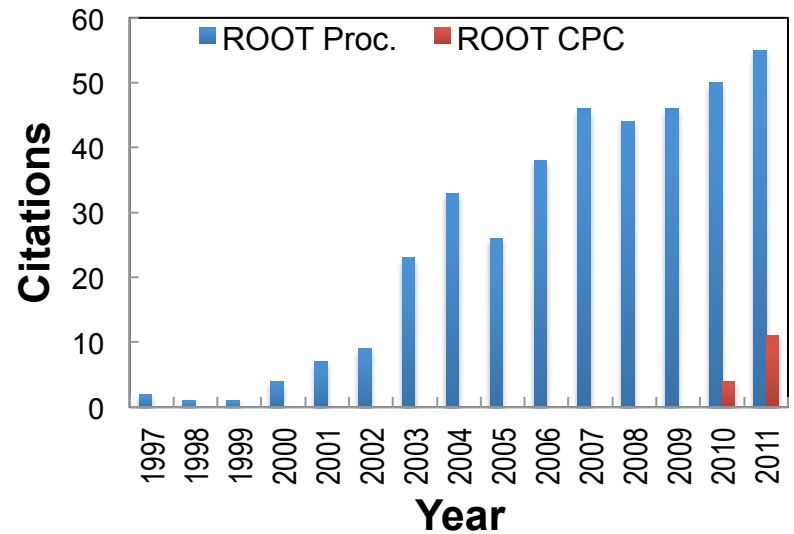
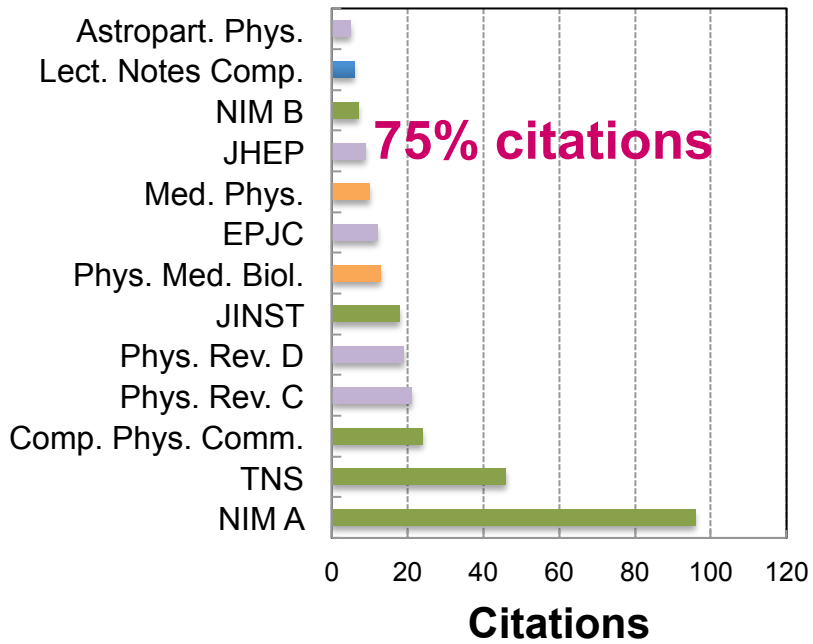
Comp. Phys Comm., vol. 180, no. 12, pp. 2499-2512, 2009

32 citations

(20 October 2012)



ROOT Proc.: Citing Journals



| Field of citing journals | | |
|--------------------------|----------|--------|
| | Geant4 % | ROOT % |
| Technology | 30.3 | 49.6 |
| Physics | 29.9 | 18.2 |
| BioMedical | 13.9 | 6.0 |

HEP experiments

Start of run

CDF: 1985/1988

LEP: 1989

BaBar: 1999

LHC: 2008

LEP

- ALEPH
- DELPHI
- L3
- OPAL

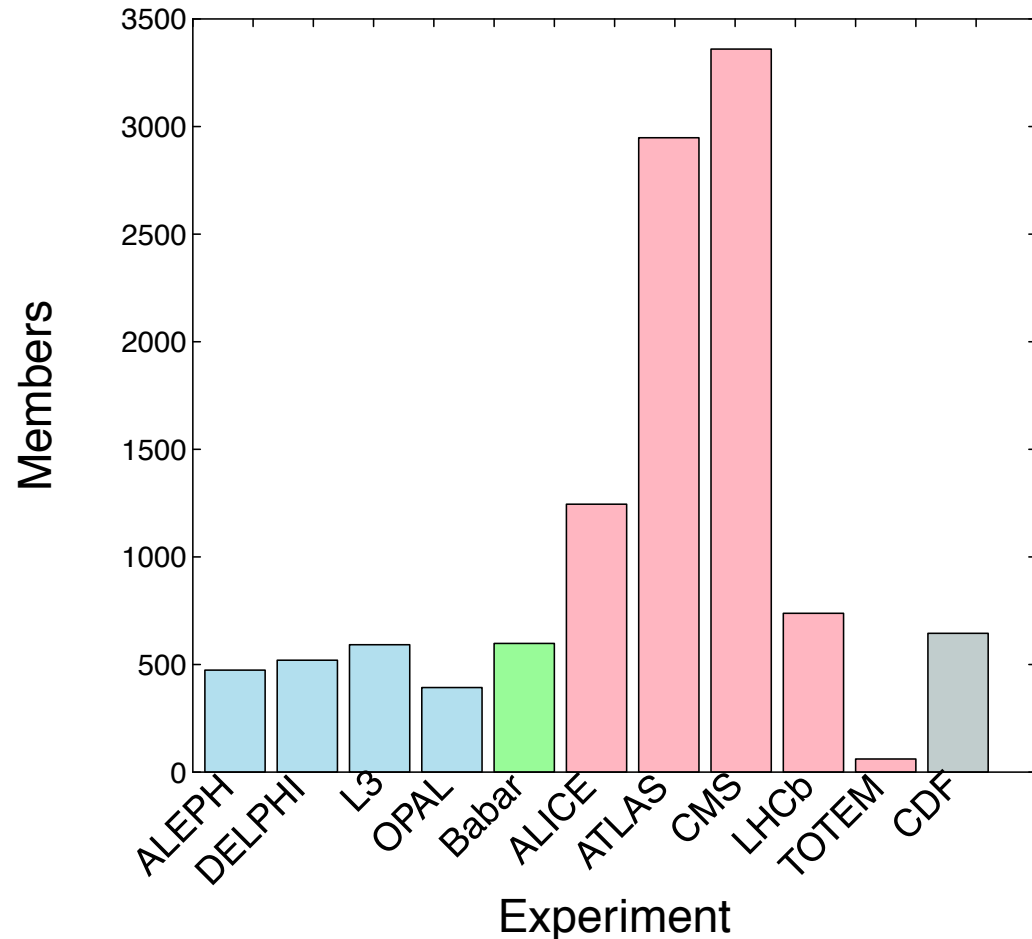
CDF

BaBar

LHC

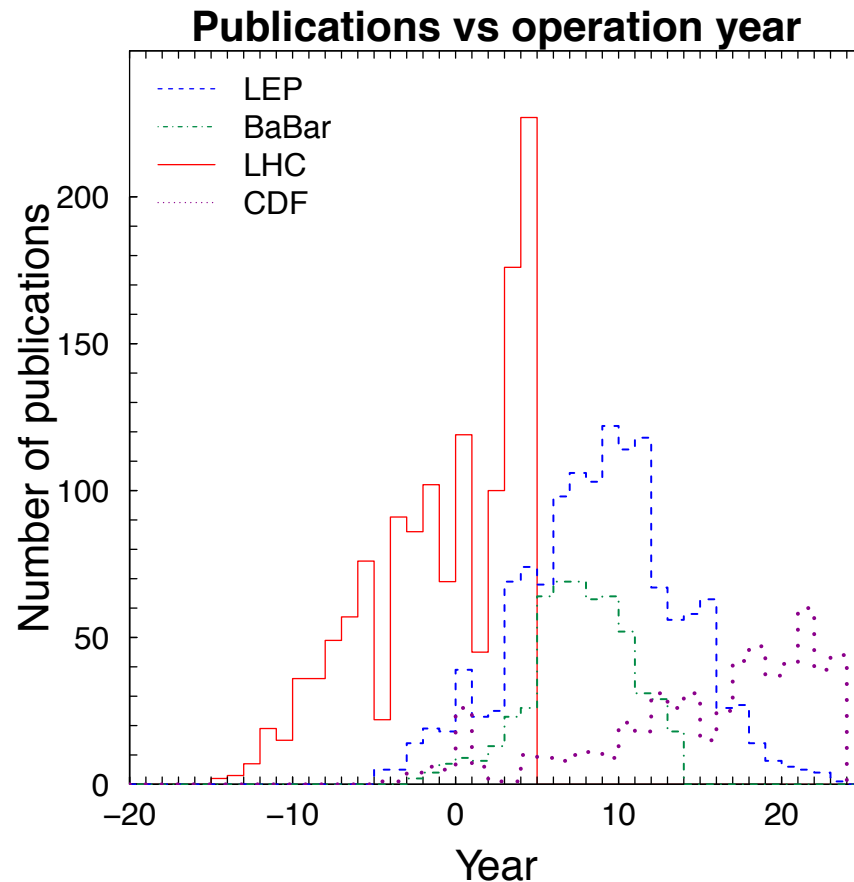
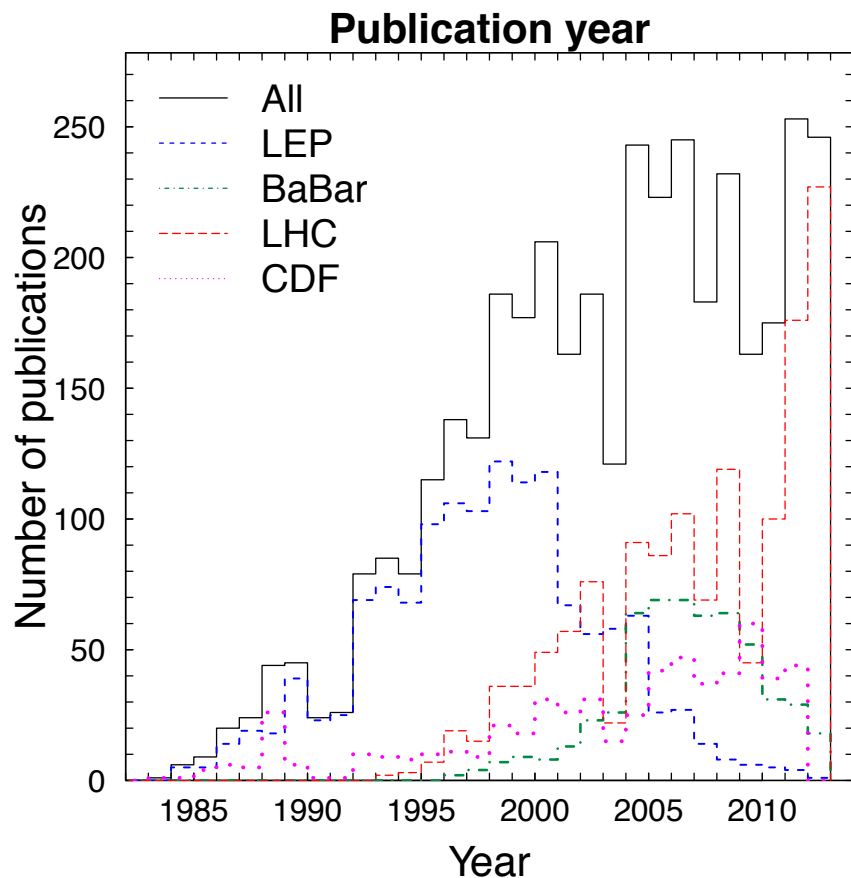
- ALICE
- ATLAS
- CMS
- LHCb
- TOTEM

Collaboration members



Time distribution

Start of run {
CDF: 1985/1988
LEP: 1989
BaBar: 1999
LHC: 2008



Publication year

Rescaled w.r.t. year of start run

Time distribution

Start of run

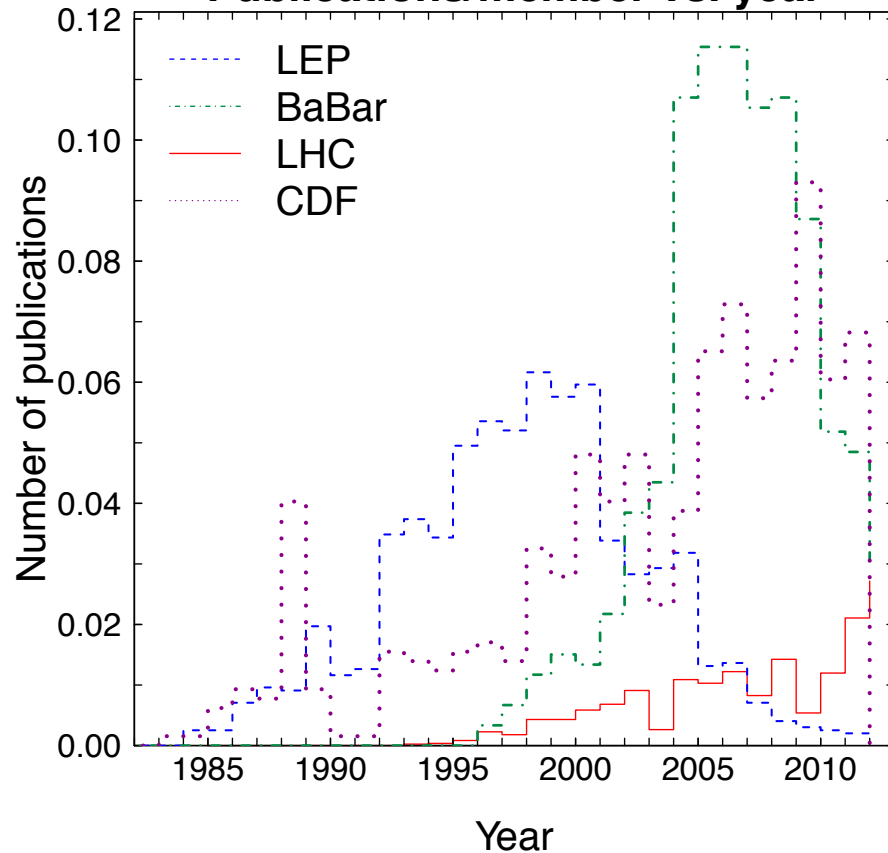
CDF: 1985/1988

LEP: 1989

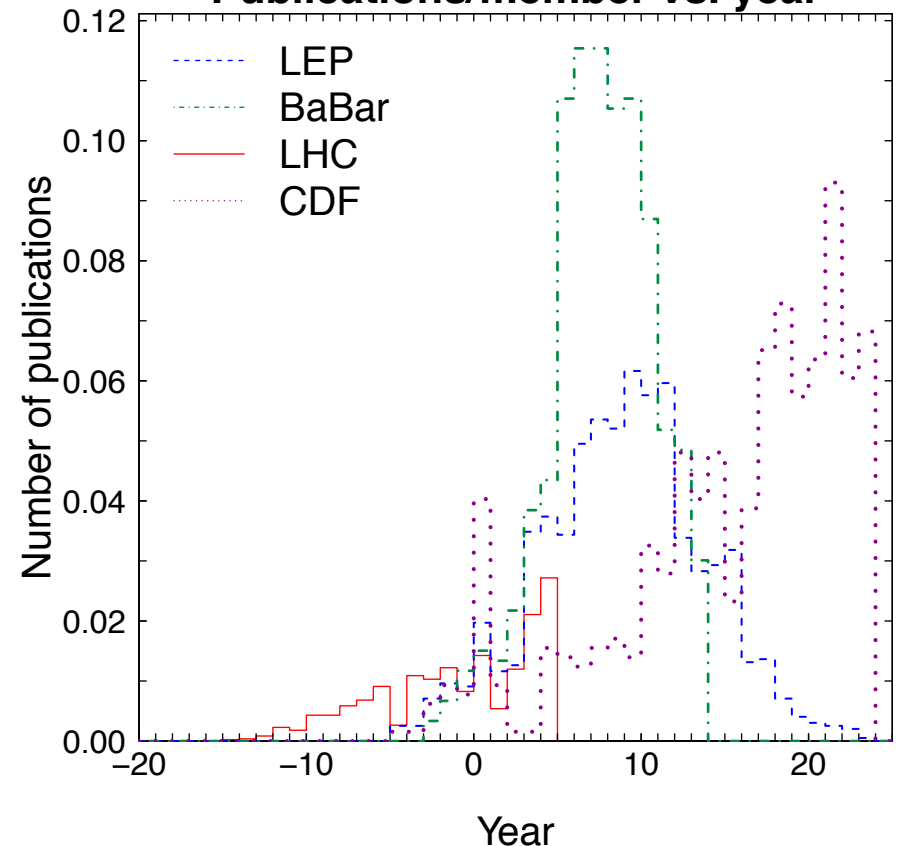
BaBar: 1999

LHC: 2008

Publications/member vs. year



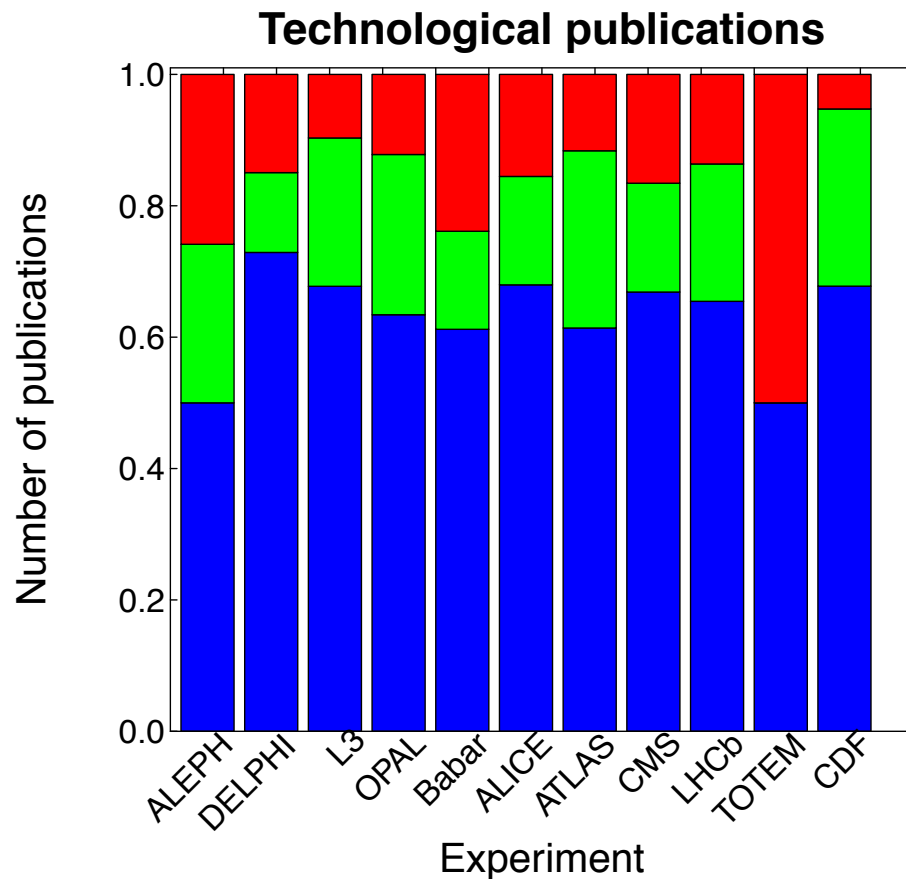
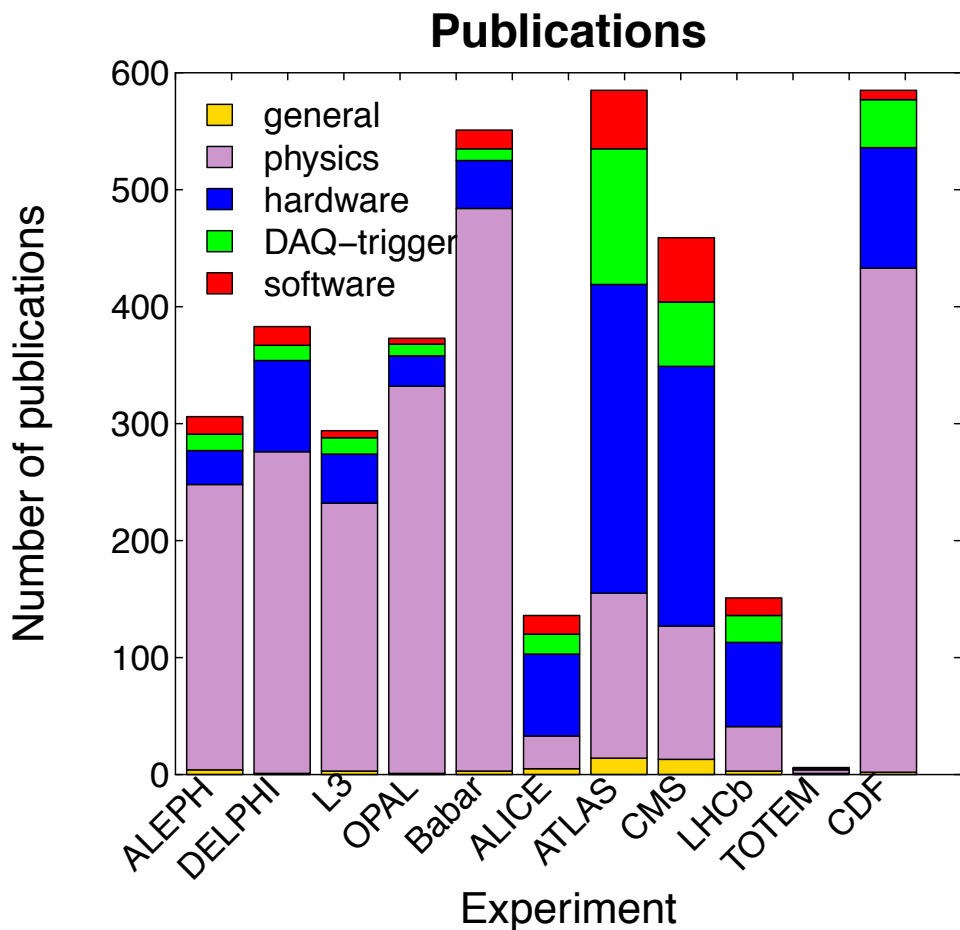
Publications/member vs. year



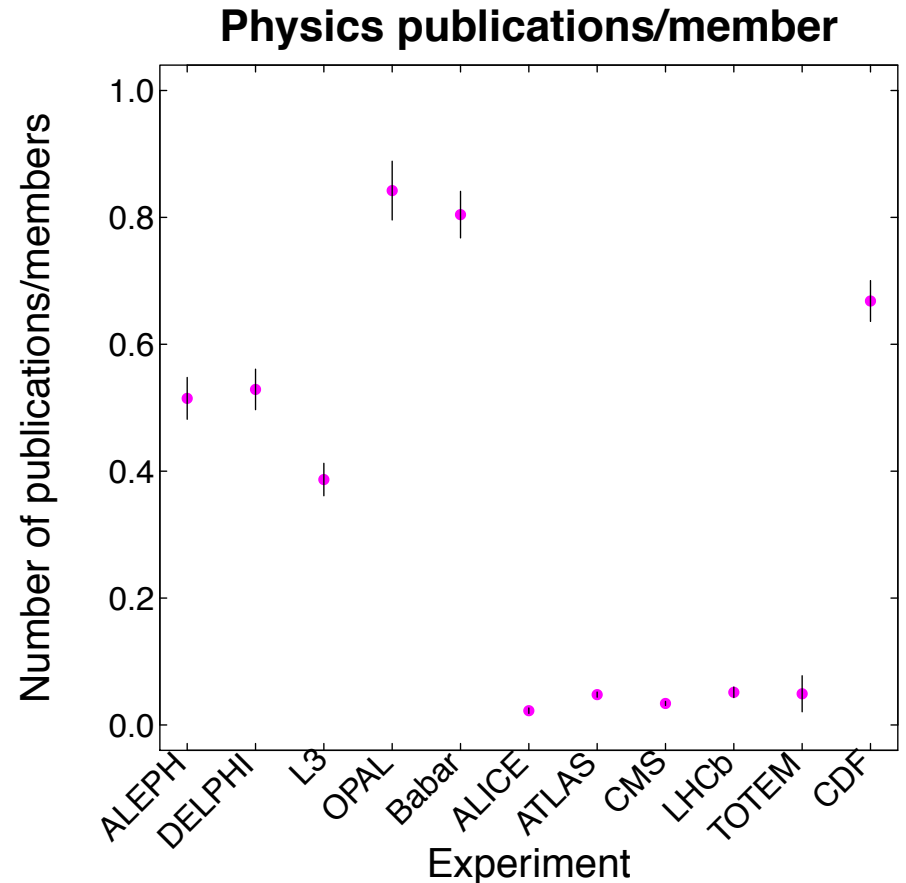
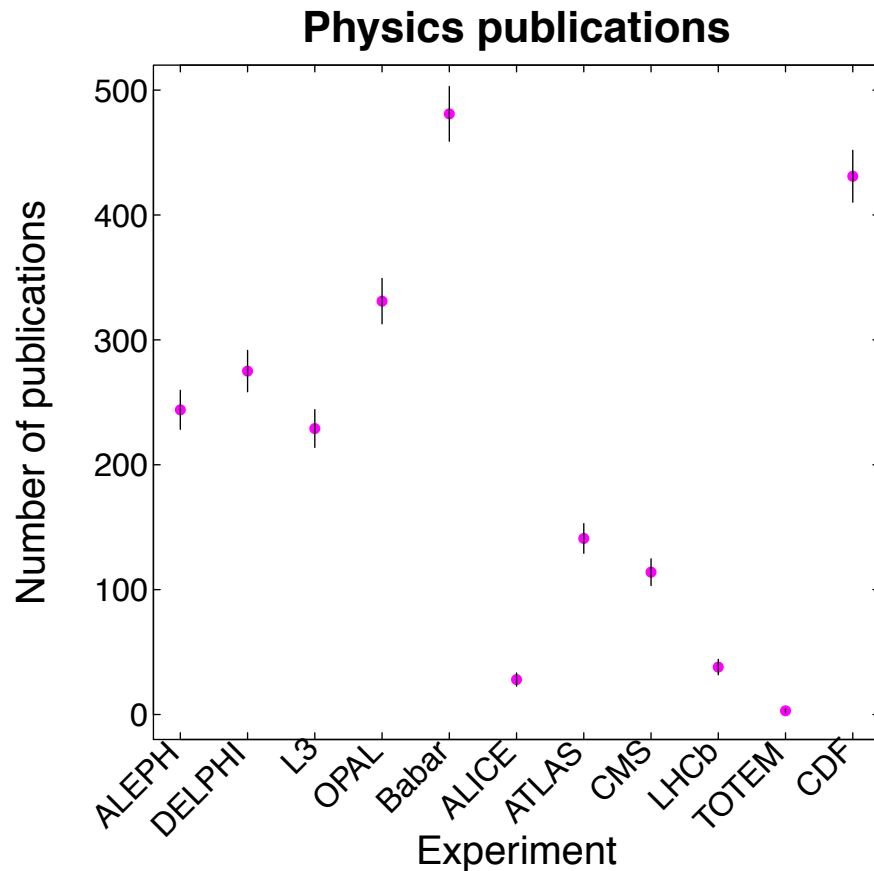
Same as previous slide, rescaled by the number of experiment members

Publications

Share of hardware, software and DAQ-trigger publications



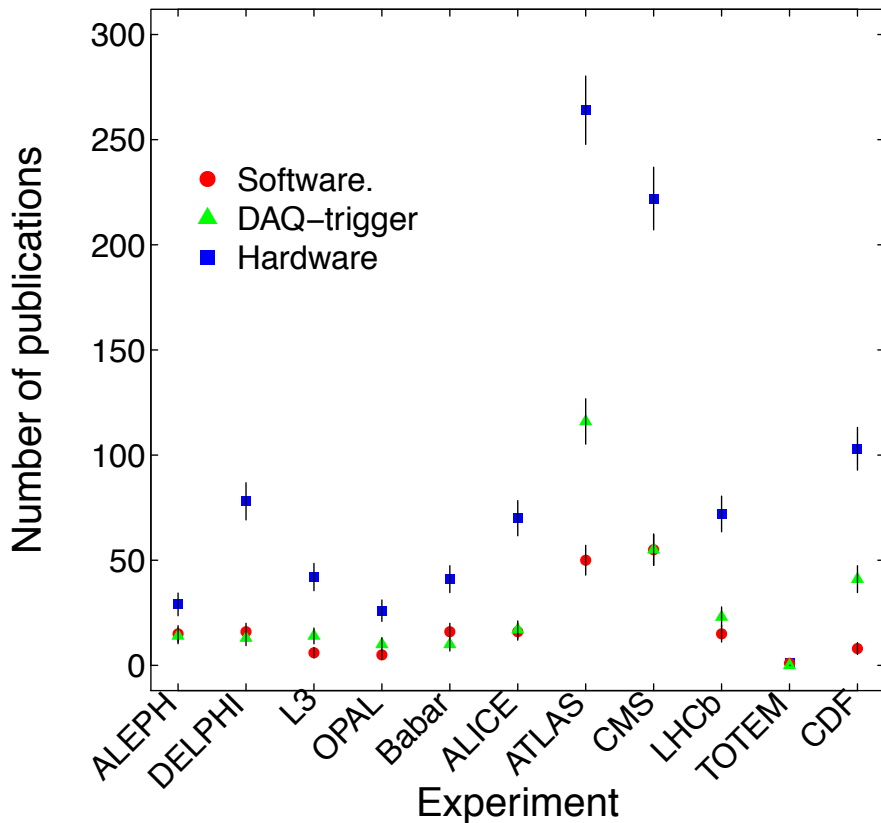
Physics publications



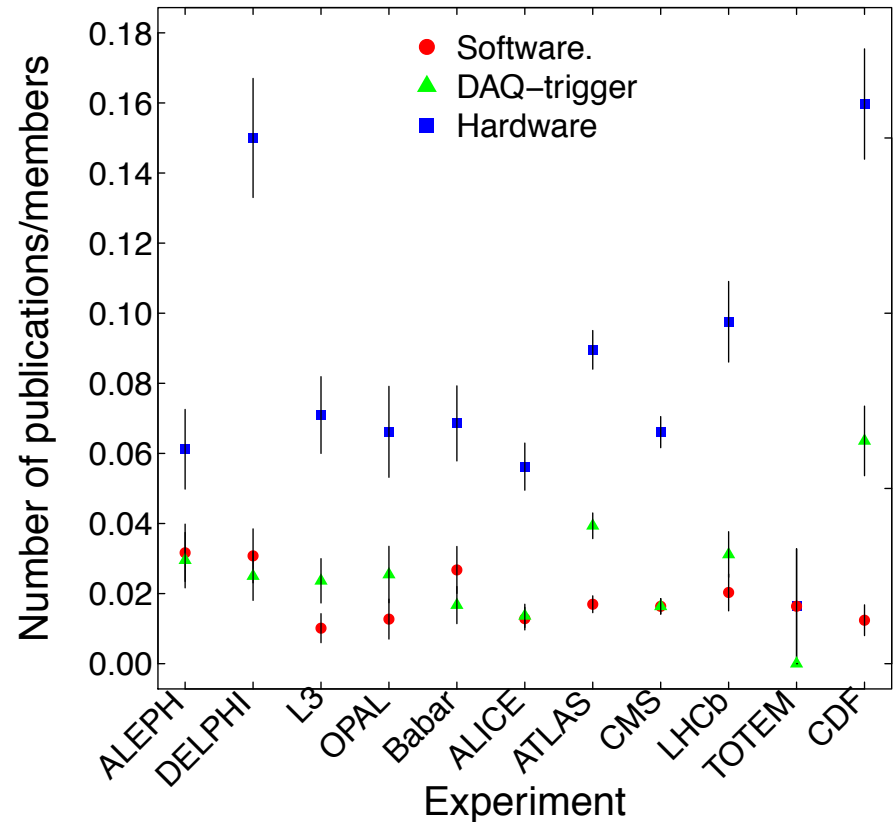
LEP experiments completed their life-cycle
LHC experiments: at an early stage of their physics production

Technological publications

Technological publications



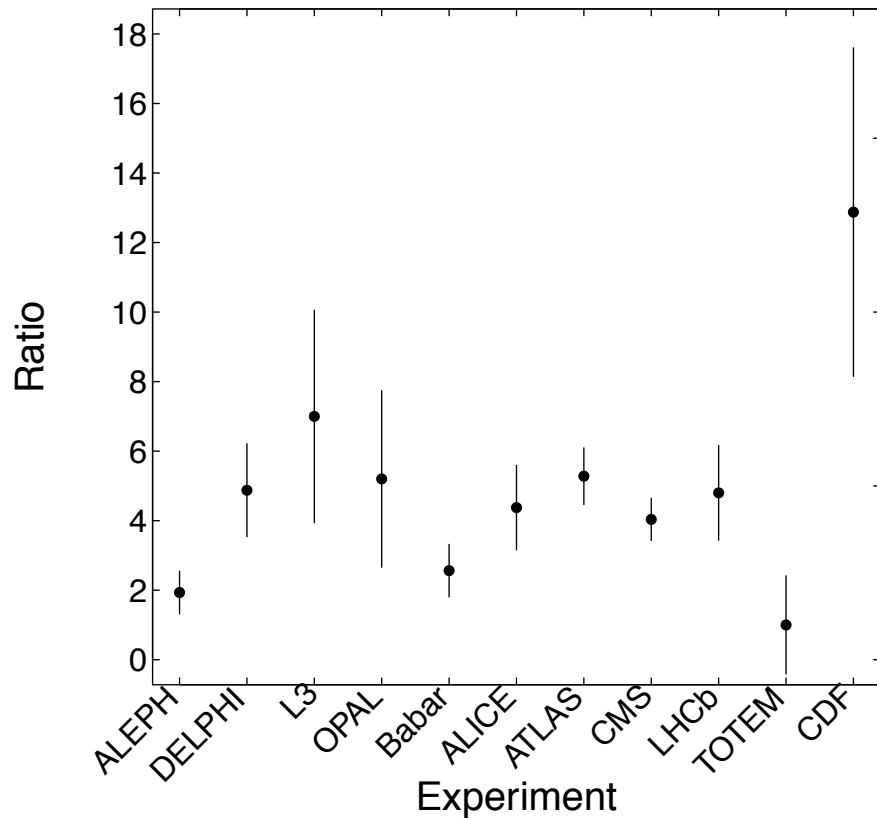
Technological publications/member



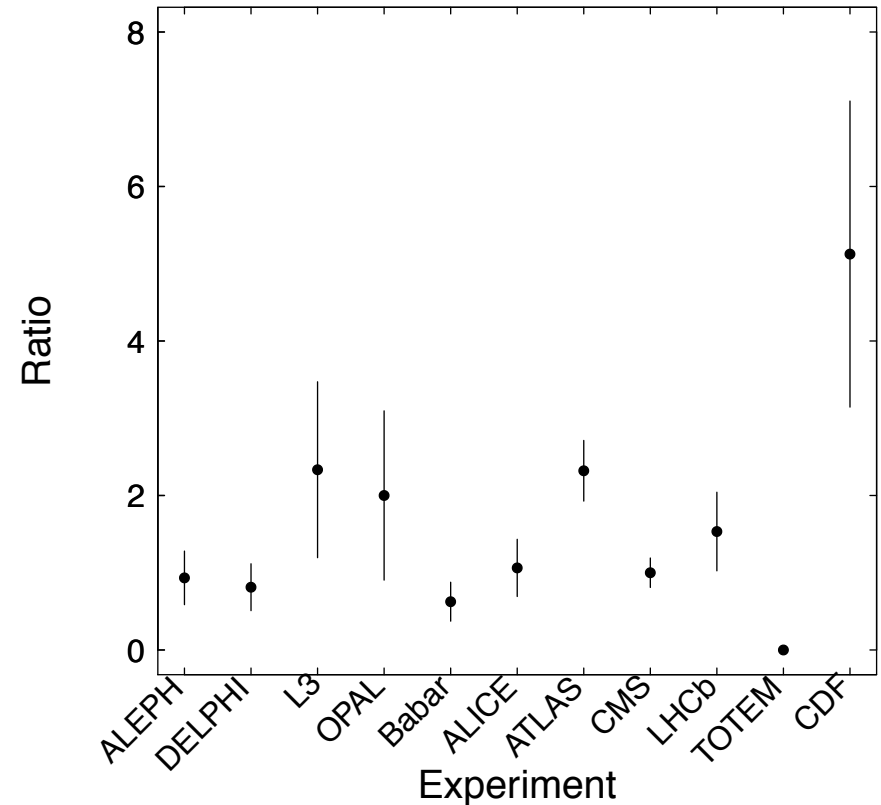
Roughly constant trends, once the number of publications is normalized to the number of collaborators

Software vs. hardware

Hardware/software publications

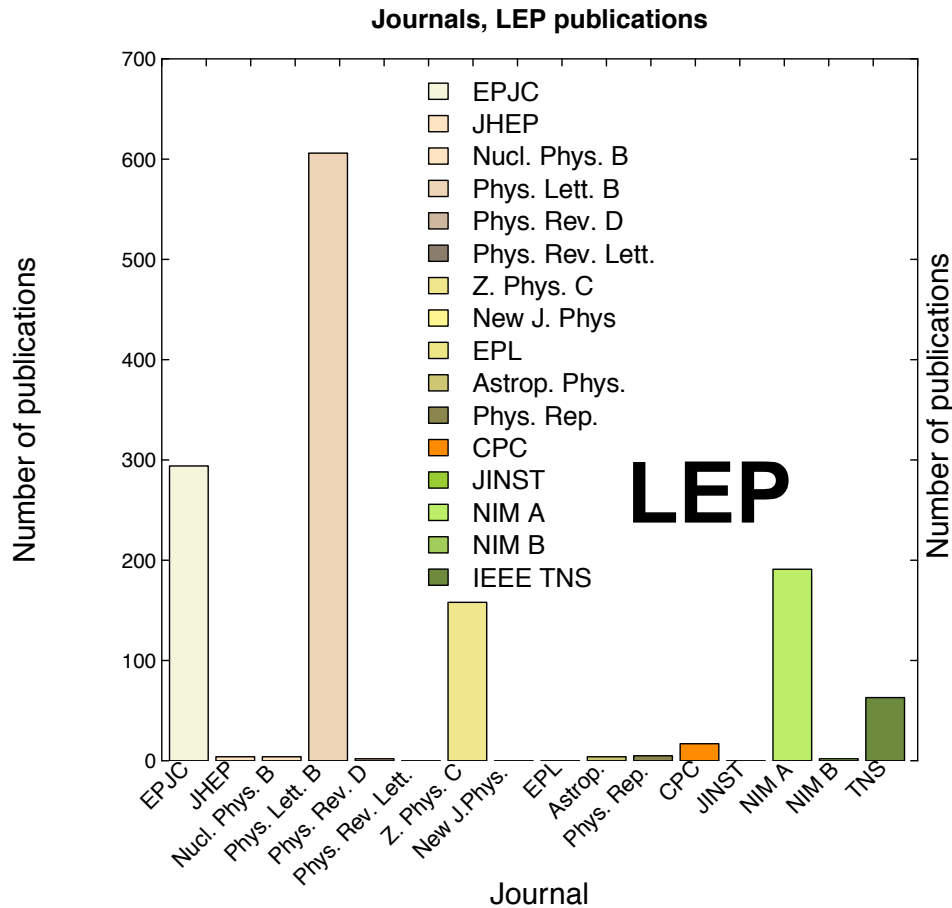


DAQ-trigger/software publications

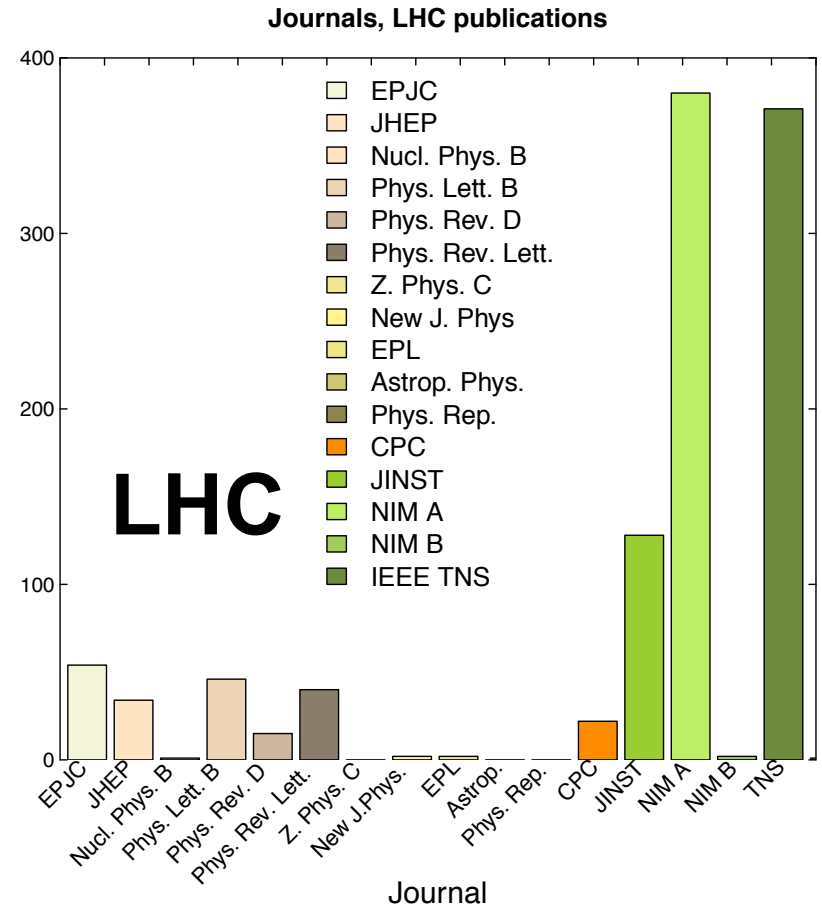


Hardware publications: approximately 4 times more than software
DAQ-trigger publications: approximately 1.3 times more than software

Journals: LEP and LHC



Dominated by physics publications



Still dominated by technological publications

Citations

The most cited papers are often the general reference papers about the detector published by each experiment

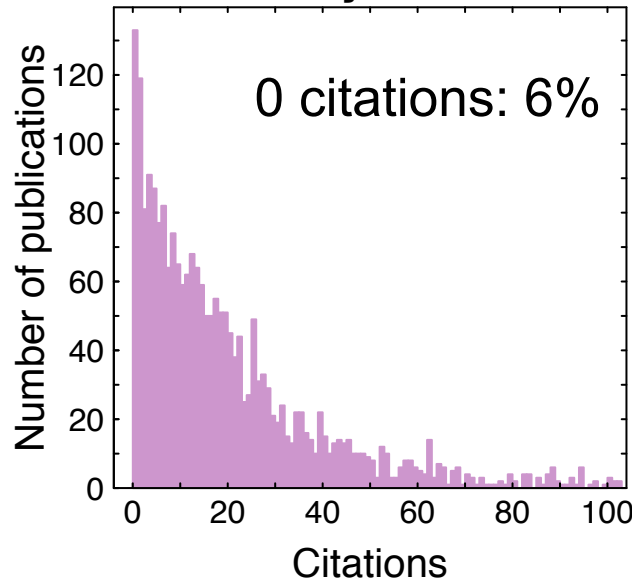
Citations of the most cited paper

| | |
|----------------|------------|
| ALEPH: | 340 |
| DELPHI: | 309 |
| L3: | 509 |
| OPAL: | 473 |
| BaBar: | 859 |
| ALICE: | 116 |
| CMS: | 129 |
| LHCb: | 101 |
| TOTEM: | 35 |

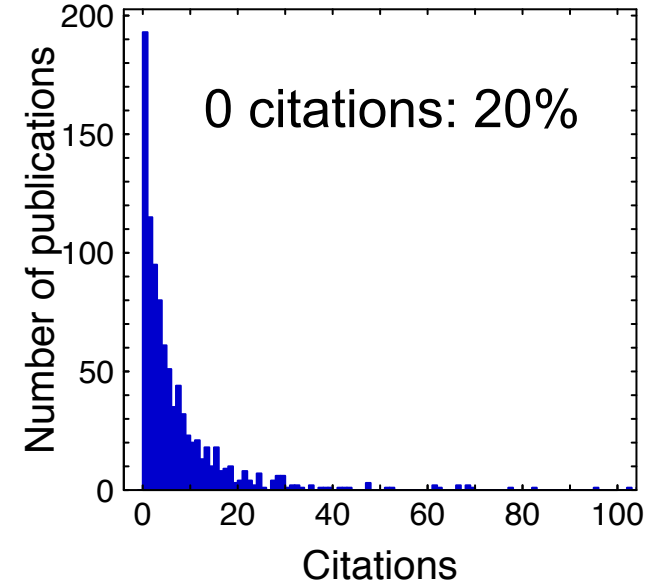
ATLAS: ATLAS pixel detector electronics and sensors: **185**

Maria Grazia Pia, *INFN Genova*

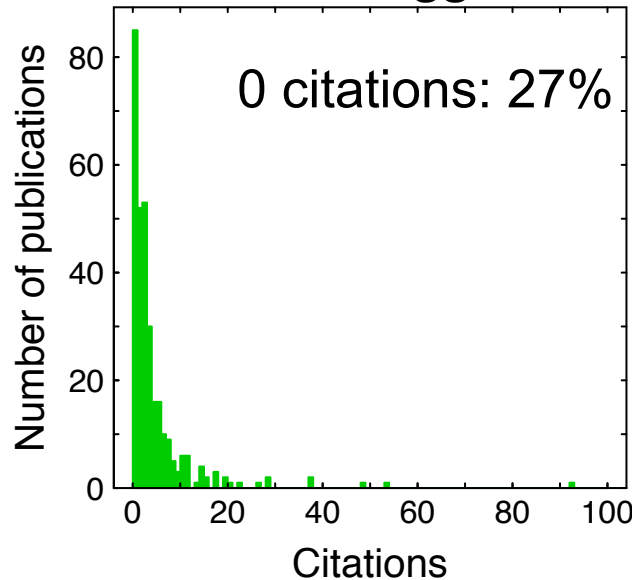
Physics



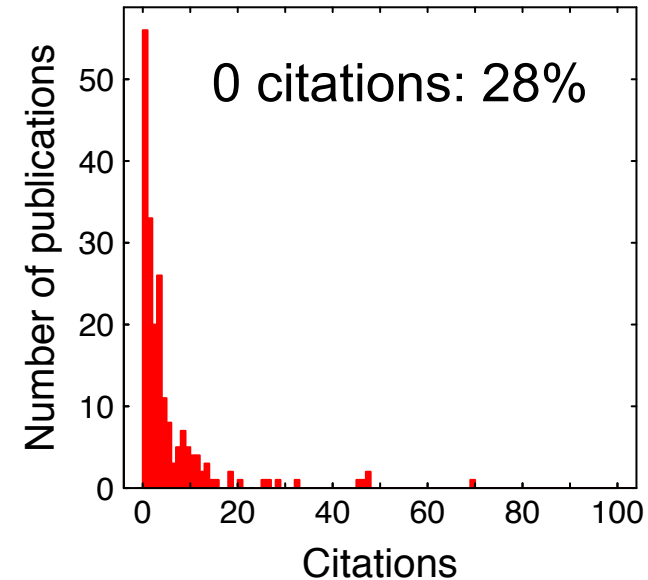
Hardware



DAQ-Trigger



Software



References

Physics papers cite more references than technological papers

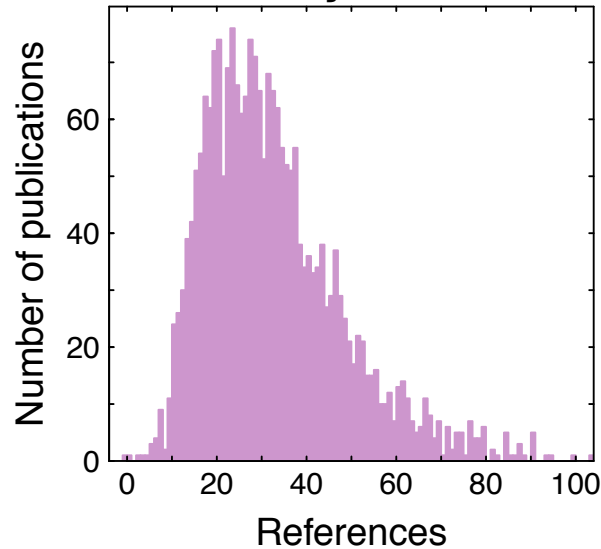
More references



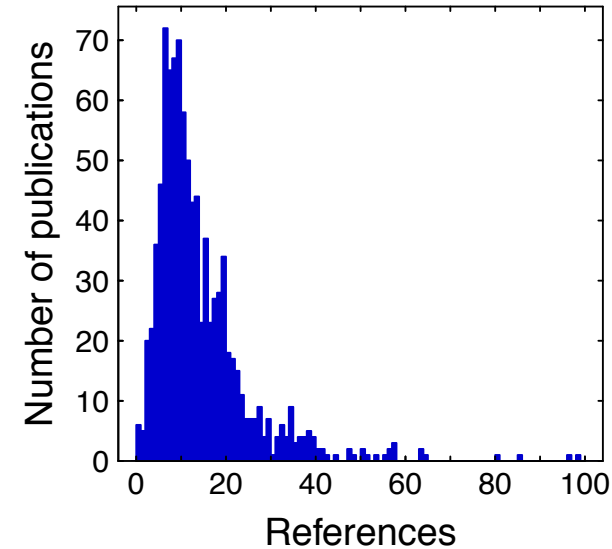
more citations

Bibliographical entries in software papers are often **web sites**

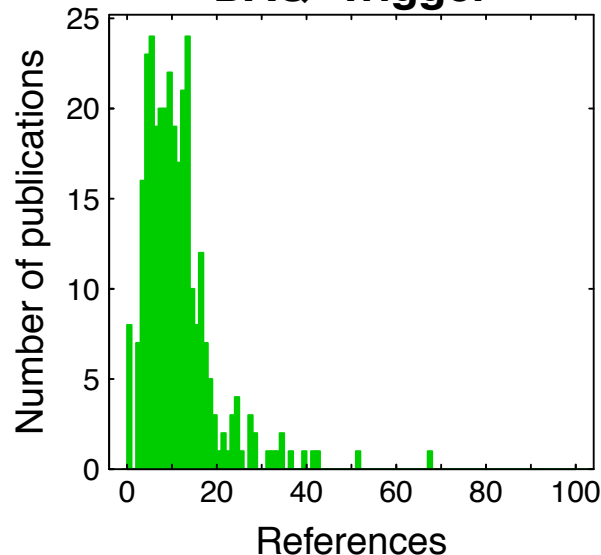
Physics



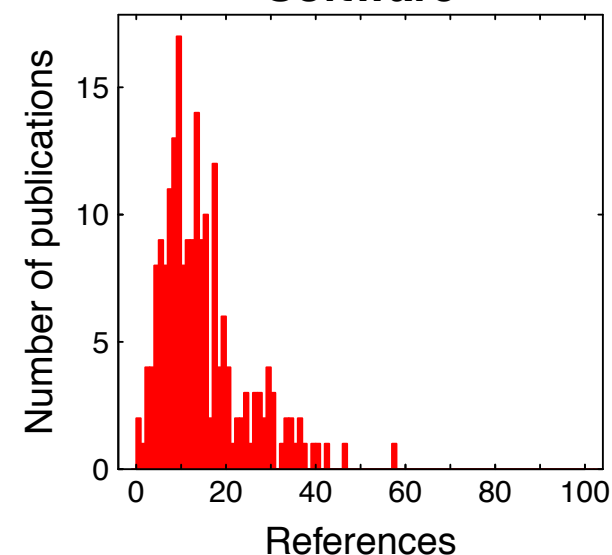
Hardware



DAQ-Trigger

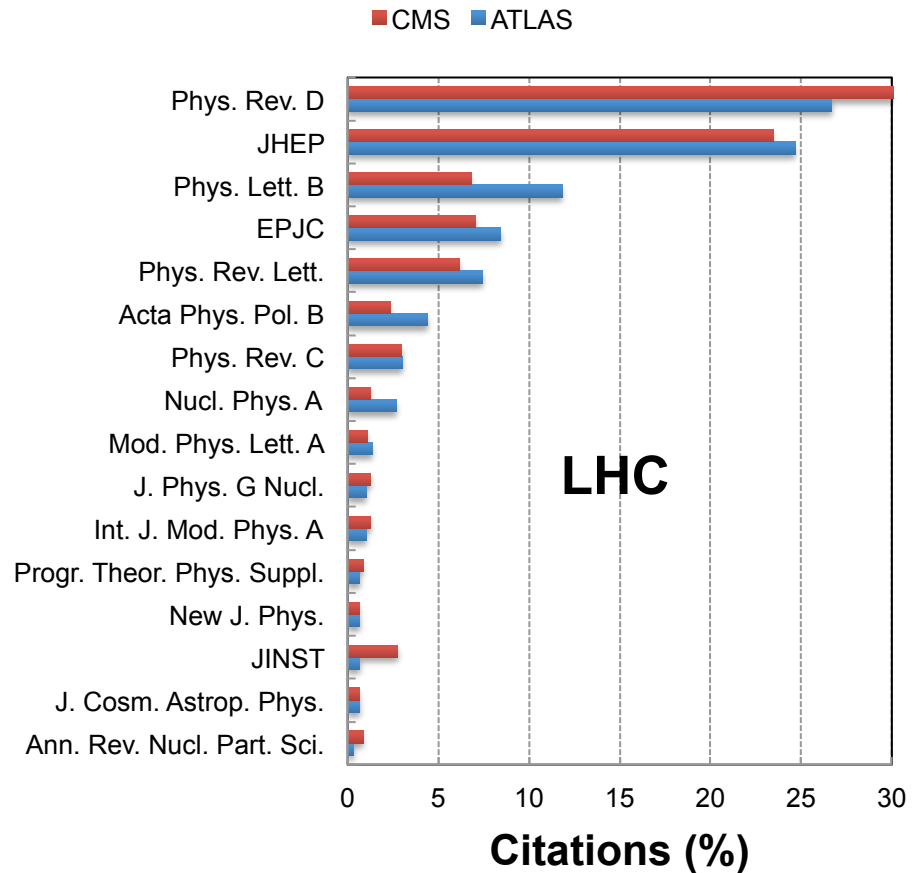
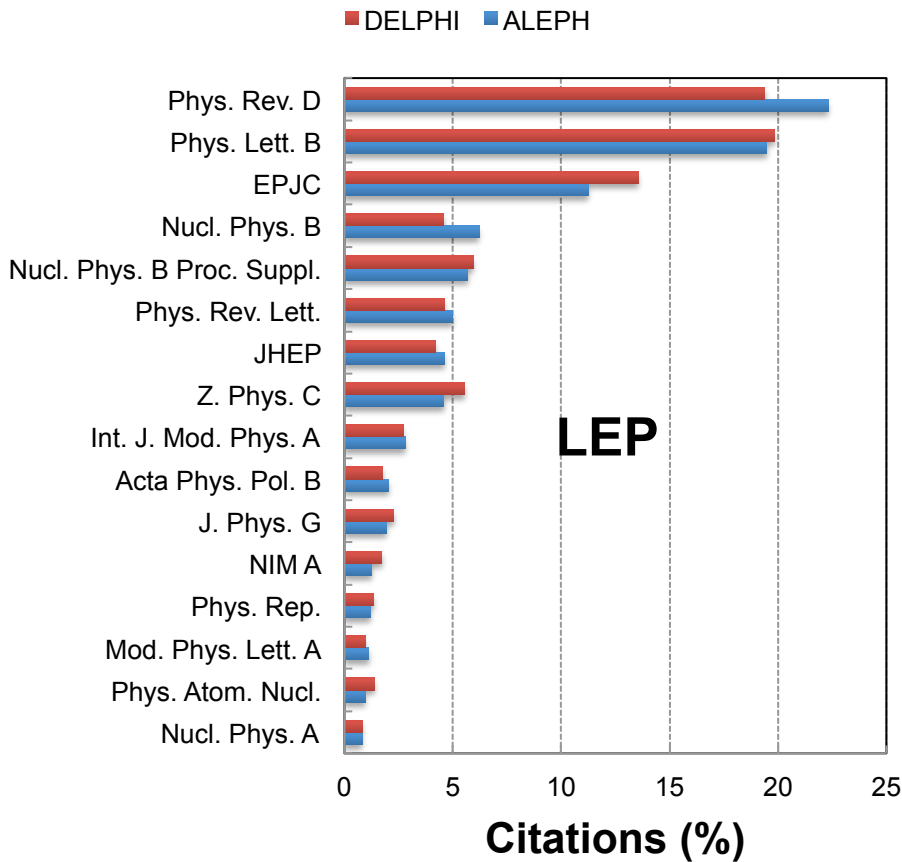


Software



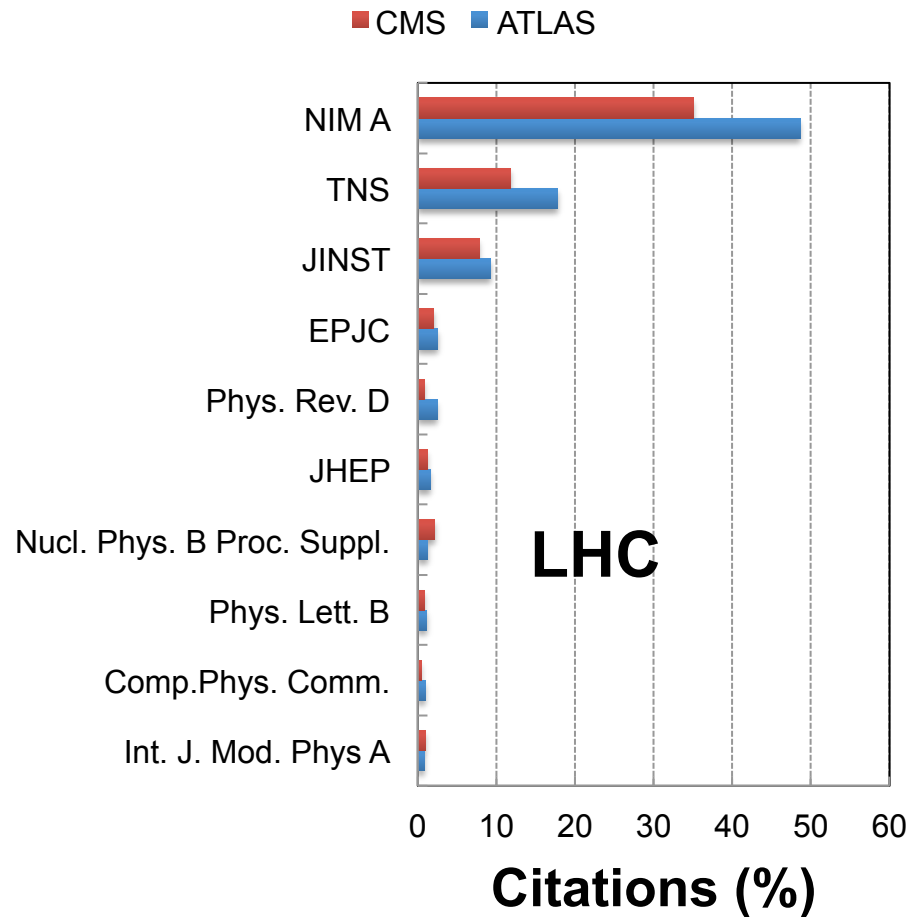
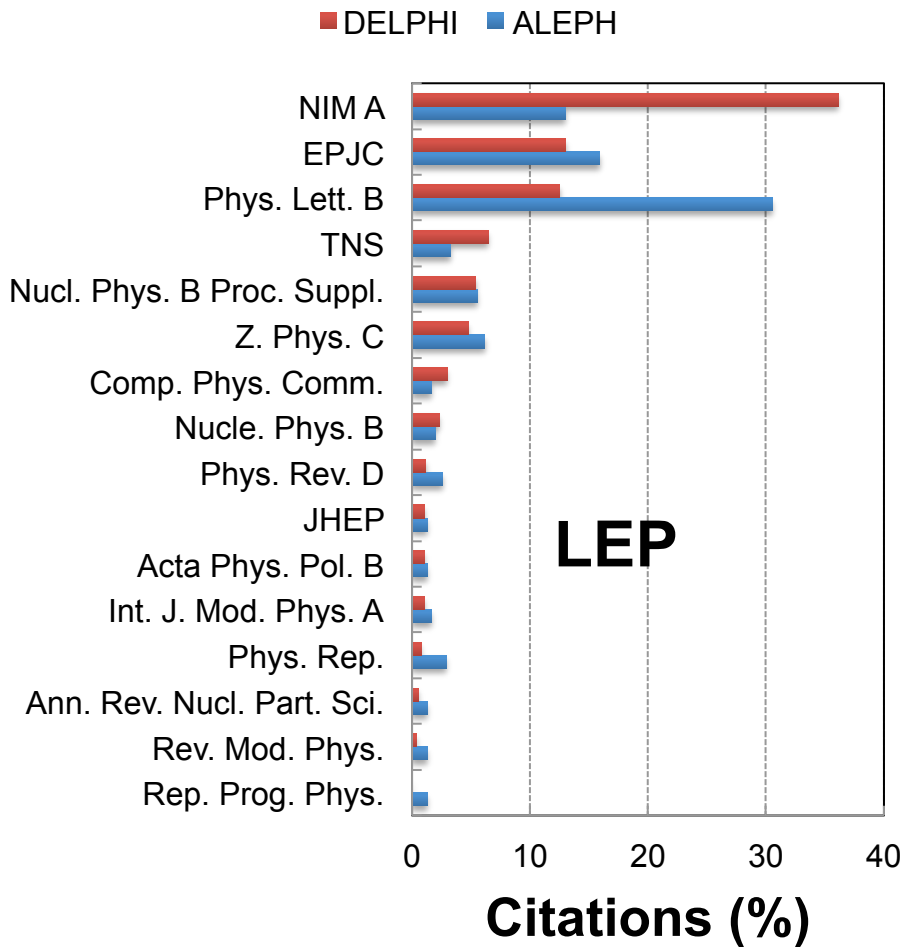
Sources of citations of physics papers

Samples in plots account for >90% of citations



Citations to HEP physics papers mostly come from journals specialized in HEP and a few related fields (astroparticle and nuclear physics)

Sources of citations of technological papers

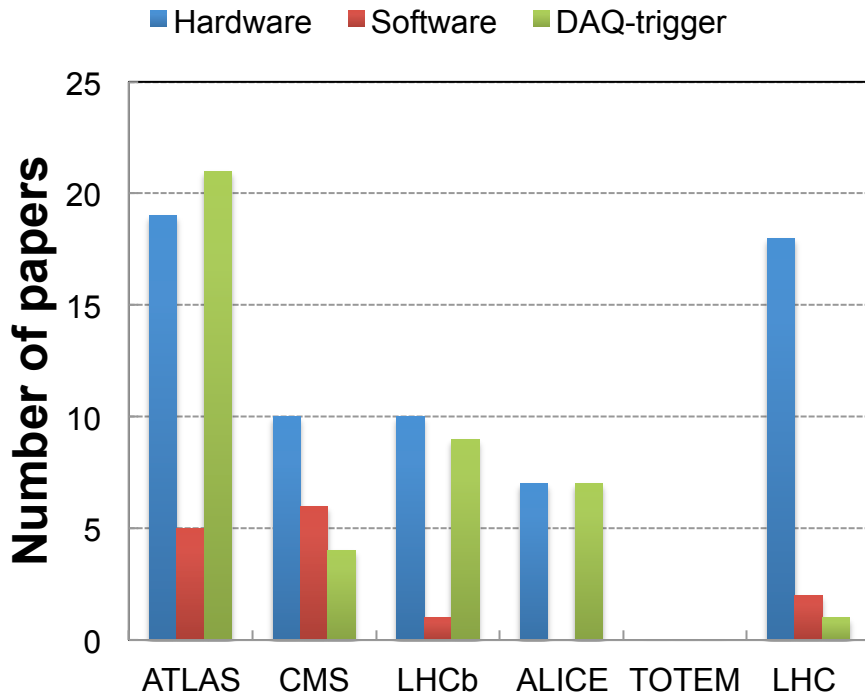


Citations from HEP physics and technology journals

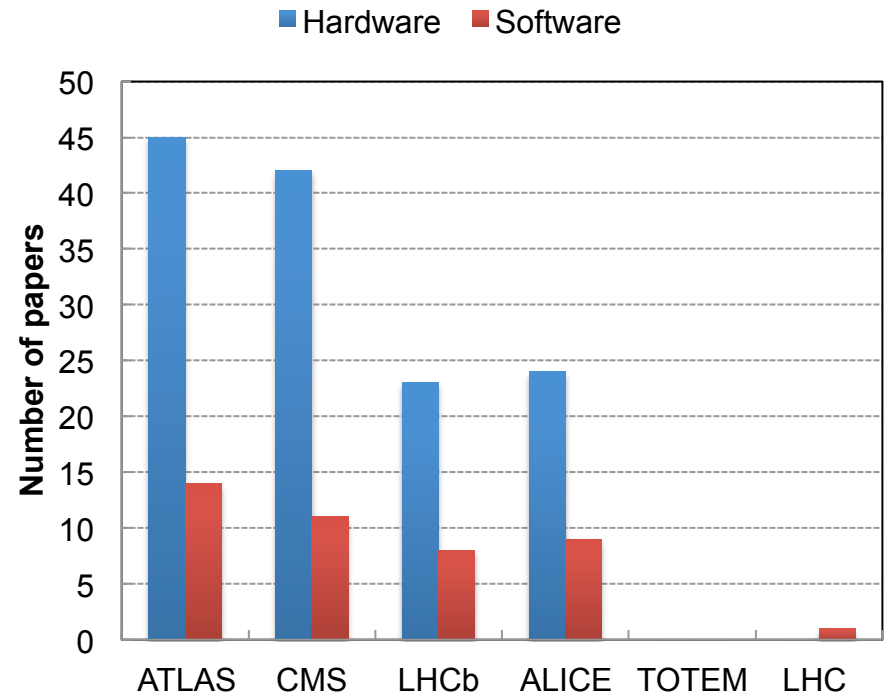
2008-2011

More refined analysis of technological papers published since the start of LHC run

TNS 2008-2011

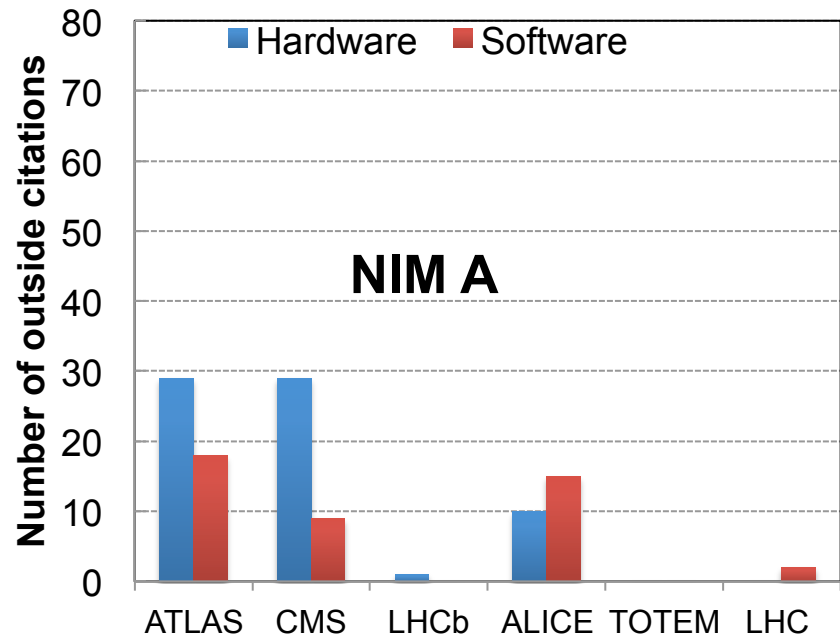
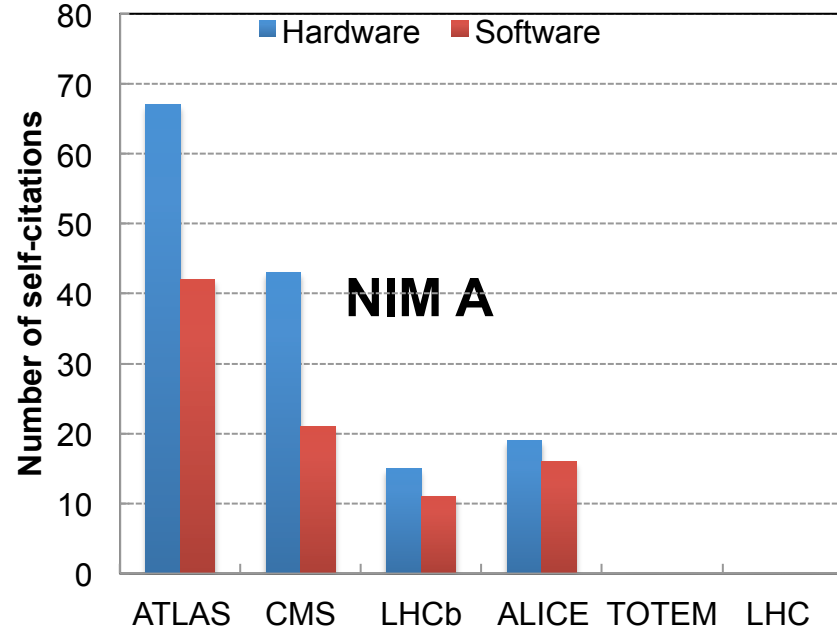
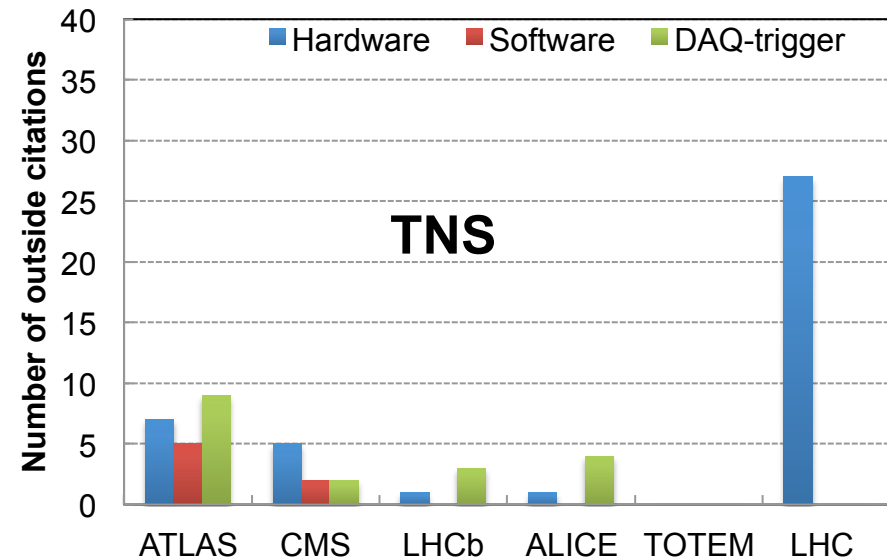
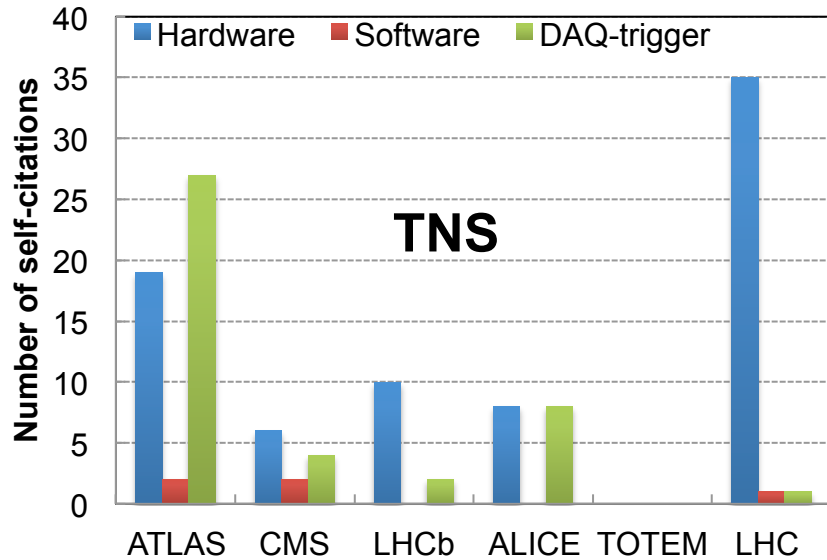


NIM 2008-2011



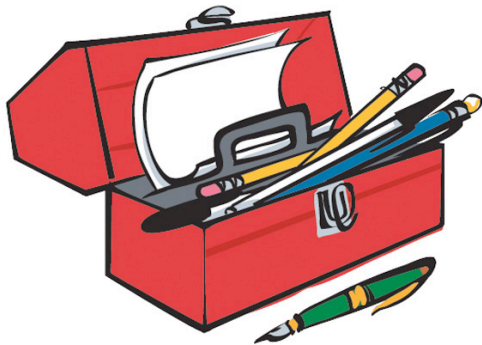
Self-citations

Outside citations



Conclusions

- Software is largely **underrepresented** in HEP scholarly literature w.r.t. hardware
- **Publication patterns** appear similar in the LEP and LHC era
- **Citation patterns** are different for publications by HEP experiments and about general software tools



Publish!

...and don't forget to cite