A Global Digital Library for High-Energy Physics

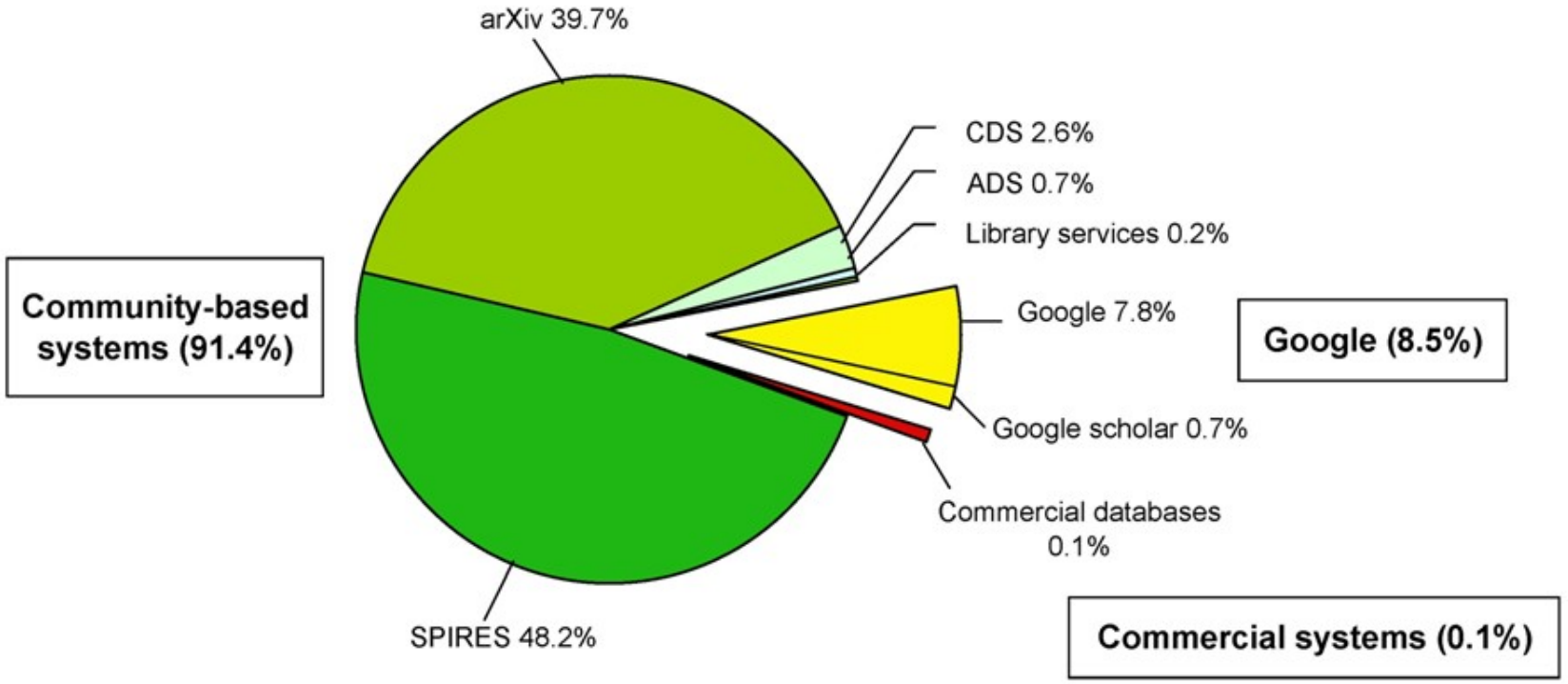
Annette Holtkamp

CERN-UNESCO School on Digital Libraries – Rabat, Nov 2010

HEP community

- closely-knit community
  - 20-30k active researchers publishing 10k articles
  - large collaborations (up to 5000 members)
  - very international (even small author groups)
  - authors = readers
- rapid information exchange essential
  - mailing of preprints since the 60's
  - long OA tradition
  - >90% of HEP journal articles on arXiv
- dominance of community based information systems
  - arXiv
  - SPIRES

Dominance of community services



From 2007 survey of 2,000 physicists. Gentil-Beccot et al, *Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course*.
J.Am.Soc.Inf.Sci.60:150-160,2009 arXiv:0804.2701

SPIRES (1974-)

- network of databases
  - HEP literature, conferences, institutions, experiments, hepnames, jobs
- SLAC – DESY – Fermilab Collaboration
- SPIRES-HEP
  - Metadata for 850k objects, ~800 new records per week
  - Preprints, journal articles, conference contributions, books, grey literature
  - since 1974, web server since 1991
  - 100k searches/day
- high data quality, manually curated, comprehensive coverage
- high acceptance, user involvement

But:

- outdated technology from the 70‘s

http://inspirebeta.net

run by

Fermilab SLAC

Bibliographic Content

- SPIRES content (plus part of CDS):
  journal articles, conference proceedings, preprints, experimental notes, theses
- going beyond SPIRES:
  conference slides, multimedia, software, high-level research data…
- going back before 1974
- more material from neighboring disciplines
  astrophysics, nuclear physics, mathematics…
  cited by core HEP articles

“Fulltext” repository

- all freely accessible articles
  - esp. “endangered” material
- access restricted articles
  - “hidden archive”
  - agreements with Springer and APS
- historical material
  - scanning of old preprint series
- beyond articles
  - slides, multimedia, software, wikis…

Search

- Google-like freetext search
- fulltext search
- Complex second-order searches

Example:

Find the most influential HEP core papers that cite the Hitchin article „Generalized Calabi-Yau manifolds“ but don‘t cite any papers by Polchinski

```
refersto:reportnumber:math/0209099 collection:core
  cited:100->9999 NOT refersto:author:Polchinski
```

Fulltext search - snippets

HEP 11 records found Search took 0.02 seconds.

1. **Spectral Methods in Quantum Field Theory.**
Noah Graham (Middlebury Coll.), Markus Quandt (Tubingen U.), Herbert Weigel (Stellenbosch U.). Apr 2009. 191 pp.
Published in **Lect.Notes Phys. 777 (2009) 1**

References | BibTeX | LaTeX(US) | LaTeX(EU) | EndNote

Snippets courtesy of Springer
... . 4.3 Bosons, Fermions, **Supersymmetry**, Central Charge, and the BPS Bound . . . . . 4.3.2 **Supersymmetry** and Central Charge . . . .
... from a Majorana fermion is M/(2π). 4.3 Bosons, Fermions, **Supersymmetry**, Central Charge, and the BPS Bound Here we will consider the prime denotes a derivative with respect to the argument. **Supersymmetry** of the combined boson–fermion system is established by relating the the fermion and boson masses are identical, as required by **supersymmetry**. 4.3 Bosons, Fermions, **Supersymmetry**, Central Charge, and the BPS Bound...

Detailed record - Similar records - Cited by 1 record

2. **Detecting dark matter WIMPs in the Draco dwarf: A multi-wavelength perspective.**
Sergio Colafrancesco (Rome Observ.), S. Profumo (Caltech), P. Ullio (SISSA, Trieste & INFN, Trieste). Jul 2006. 25 pp.
Published in **Phys.Rev. D75 (2007) 023513**
e-Print: **astro-ph/0607073**

References | BibTeX | LaTeX(US) | LaTeX(EU) | EndNote

... [53]. We adjusted here the values of the universal soft **supersymmetry** breaking scalar mass m0 given in [75] in order to relic abundance: B′ lies in the bulk region of small **supersymmetry** breaking masses, and gives a dominant b − ¯ final the funnel region where neutralinos rapidly annihilate through s-channel heavy **Higgs**es exchanges, dominantly producing b − b pairs as outcome of...

Detailed record - Similar records - Cited by 40 records

Detailed record page

- abstract
- keywords
- publication info
- thumbnails of figures
- various export formats
- tabs for
  - references
  - citations
  - fulltext
  - full-sized plots with captions

Detailed record with plots

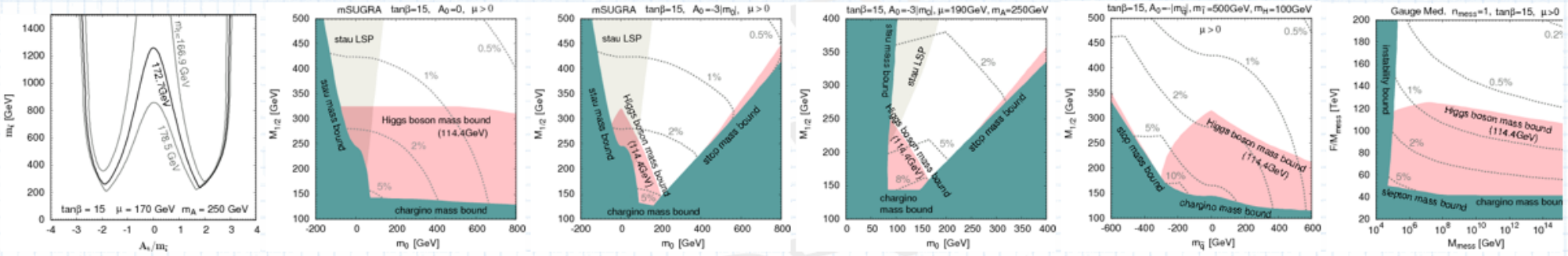
Information | References (127) | Citations (60) | Fulltext | Plots

Supersymmetry, naturalness, and signatures at the LHC.

Ryuichiro Kitano (SLAC), Yasunori Nomura (UC, Berkeley & LBL, Berkeley).
Feb 2006

Phys.Rev. D73 (2006) 095004
e-Print: **hep-ph/0602096**

Abstract: Weak scale supersymmetry is often said to be fine-tuned, especially if the matter content is minimal. This is not true if there is a large A term for the top squarks. We present a systematic study on fine-tuning in minimal supersymmetric theories and identify low energy spectra that do not lead to severe fine-tuning. Characteristic features of these spectra are: a large A term for the top squarks, small top squark masses, moderately large tan\beta, and a small \mu parameter. There are classes of theories leading to these features, which are discussed. In one class, which allows a complete elimination of fine-tuning, the Higgsinos are the lightest among all the superpartners of the standard model particles, leading to three nearly degenerate neutralino/chargino states. This gives interesting signals at the LHC -- the dilepton invariant mass distribution has a very small endpoint and shows a particular shape determined by the Higgsino nature of the two lightest neutralinos. We demonstrate that these signals are indeed useful in realistic analyses by performing Monte Carlo simulations, including detector simulations and background estimations. We also present a method that allows the determination of all the relevant superparticle masses without using input from particular models, despite the limited kinematical information due to short cascades. This allows us to test various possible models, which is demonstrated in the case of a model with mixed moduli-anomaly mediation. We also give a simple derivation of special renormalization group properties associated with moduli mediated supersymmetry breaking, which are relevant in a model without fine-tuning.



Keyword(s): INSPIRE: supersymmetry: symmetry breaking | sparticle: mass spectrum | p p: inclusive reaction | neutralino: hadroproduction | neutralino: decay | dilepton: production | Higgsino: mass | gluino: mass | squark: mass | anomaly: moduli | renormalization group | CERN LHC Coll | numerical calculations: Monte Carlo | bibliography

Plot extraction

- figures extracted from LaTeX sources (arXiv)
- captions searchable

soon to come:

- extraction from pdf
- phrase from fulltext referencing a figure

Citation analysis

- cited by
- co-cited with
- self-citations
- citation history

Citation analysis: Example

Fusing gauge theory tree amplitudes into loop amplitudes - Bern, Zvi *et al* hep-ph/9409265 SLAC-PUB-6563, SACLAY-SPH-T-94-95, UCLA-TEP-94-29, SWAT-94-36

Cited by: 346 records

(254) **Iteration of planar amplitudes in maximally supersymmetric Yang-Mills theory at three loops and beyond** - Bern, Zvi *et al* hep-th/0505205 SLAC-PUB-11210, UCLA-05-TEP-14

(252) **Generalized unitarity and one-loop amplitudes in N=4 super-Yang-Mills** - Britto, Ruth *et al* hep-th/0412103

(252) **New recursion relations for tree amplitudes of gluons** - Britto, Ruth *et al* hep-th/0412308

(232) **Calculating scattering amplitudes efficiently** - Dixon, Lance J. hep-ph/9601359 SLAC-PUB-7106, C95-06-04.1

(232) **One loop amplitudes for e+ e- to four partons** - Bern, Zvi *et al* hep-ph/9708239 SLAC-PUB-7529, SACLAY-SPH-T-97-090, UCLA-97-TEP-10

more

Co-cited with: 4182 records

(316) **One loop n point gauge theory amplitudes, unitarity and collinear limits** - Bern, Zvi *et al* hep-ph/9403226 SLAC-PUB-6415, SACLAY-SPH-T-94-20, UCLA-TEP-94-4, SWAT-94-17

(196) **Generalized unitarity and one-loop amplitudes in N=4 super-Yang-Mills** - Britto, Ruth *et al* hep-th/0412103

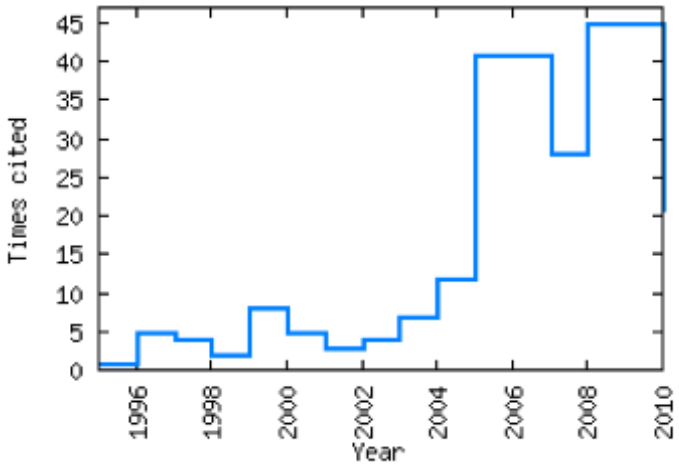
(171) **Perturbative gauge theory as a string theory in twistor space** - Witten, Edward hep-th/0312171

(159) **Direct proof of tree-level recursion relation in Yang-Mills theory** - Britto, Ruth *et al* hep-th/0501052

(153) **MHV vertices and tree amplitudes in gauge theory** - Cachazo, Freddy *et al* hep-th/0403047

more

Citation history:



Author page

- affiliation history
- coauthors
- frequent keywords
- article classification
- citation summary

Randall, Lisa

Papers:	Affiliations:	Frequent co-authors:
All papers (134)	MIT, LNS (65)	Csaki, Csaba (8)
Published (107)	Harvard U., Phys. Dept. (23)	Fitzpatrick, A.Liam (5)
Conference (11)	Harvard U. (16)	Georgi, Howard (5)
Introductory (2)	LBL, Berkeley (13)	Karch, Andreas (5)
Review (2)	Princeton U. (9)	Poppitz, Erich (5)
Thesis (1)	UC, Berkeley (8)	Chivukula, R.Sekhar (4)
		Hall, Lawrence J. (4)
		Sather, Eric (4)
		Schwartz, Matthew D. (4)

Citations:

Citation summary results	All papers	Published only
Total number of citable papers analyzed:	122	107
Total number of citations:	16,715	16,559
Average citations per paper:	137.0	154.8
Breakdown of papers by citations:		
Renowned papers (500+)	3	3
Famous papers (250-499)	9	9
Very well-known papers (100-249)	11	11
Well-known papers (50-99)	21	20
Known papers (10-49)	47	43
Less known papers (1-9)	26	18
Unknown papers (0)	5	3

HEPNAMES

3. **Zheng, Yangheng** (郑阳恒) (Beijing, GUCAS) [PAPERS] [Papers at this affiliation] [arXiv] [GOOGLE] [EXPTS] [STUDENTS] [Similar names]
Ph.D. advisor: Olsen, Stephen L. & Browder, Thomas E.
Ph.D. institution: Hawaii U. (2002)
Email: *Click number by name to see email address and affiliation history.*
Field: hep-ex, hep-ph, nucl-ex
Date verified 10/17/07
Author ID number: INSPIRE-00138380
Experiments: FNAL-E-0830

UPDATE

HEP taxonomy

hierarchical structure of all important

- HEP concepts (dynamical symmetry breaking)

providing

- synonyms (dynamically broken)
- related terms (spontaneous symmetry breaking)
- broader/narrower (symmetry breaking)
- definitions
- subject areas (high-energy physics – theory)

Keyword extraction

arXiv:0903.3933

Author keywords:
quantum cosmology -> quantum cosmology
wheeler-dewitt equation ->
tunneling probability -> tunneling
positive cosmological constant -> cosmological constant

Composite keywords:
10 transformation, canonical [22, 24]
9 potential, symplectic [22, 33]
3 tensor, energy-momentum [3, 3]
2 quantization, canonical [8, 24]
2 symmetry, gauge [4, 2]
2 oscillator, harmonic [2, 2]
1 dimension, 2 [0, 33]
1 fluid, pressure [22, 2]
1 operator, differential [16, 1]
1 inflation, open [4, 1]
1 field theory, scalar [0, 1]

Single keywords:
19 wave function
14 tunneling
13 Wheeler-DeWitt equation
13 cosmological constant
8 zero mode
7 Robertson-Walker
7 quantum cosmology
6 variational
5 Schroedinger equation
4 boundary condition
4 Poisson bracket
4 phase space

Acronyms:
WDW Wheeler-DeWitt equation

Core keywords:
Wheeler-DeWitt equation
quantum cosmology

Taxonomy applications

- fast automatic generation of keywords
  - enabling e.g. prompt alerts
  - manually curated afterwards
- automatic selection of HEP relevant articles
  - no longer time delay in border areas due to manual selection
- improved search algorithm (planned)
  - A search for „SUSY“ will also find „supersymmetry“
  - narrow/broaden search
- user tagging (planned)
  - improve Inspire generated classification
  - improve taxonomy

Author identification

- INSPIRE author id
  - compatible with other identification schemes
  - active participation in ORCID
- author disambiguation
  - using e.g. lab id's, affiliation history, coauthors and more
  - 22.000 INSPIRE-id's already assigned
- automatic association of papers with authors
  - using info on affiliations, coauthors, research topics, from publishers

    G. Chen: 963 docs, 21 real authors, only 22 docs not assigned, 97.2% success rate
  - INSPIRE-id part of author lists of large collaborations

Future projects

- innovative metrics
- semantic analysis
- content indexing of plots and tables
- recommender systems
  - combining citations, keywords, fulltext, usage pattern data...
- open API for 3rd party tools and searching
- object aggregation (OAI-ORE)
- OAIS standards for long-term document preservation