NEIL CHUE HONG

THE FOUNDATIONS OF DIGITAL RESEARCH

The Foundations of Digital Research

www.software.ac.uk

Software Sustainability Institute
N.ChueHong@software.ac.uk
Neil Chue Hong (@npch)

slideshare.net/npch/the-foundations-of-digital-research

EPSRC

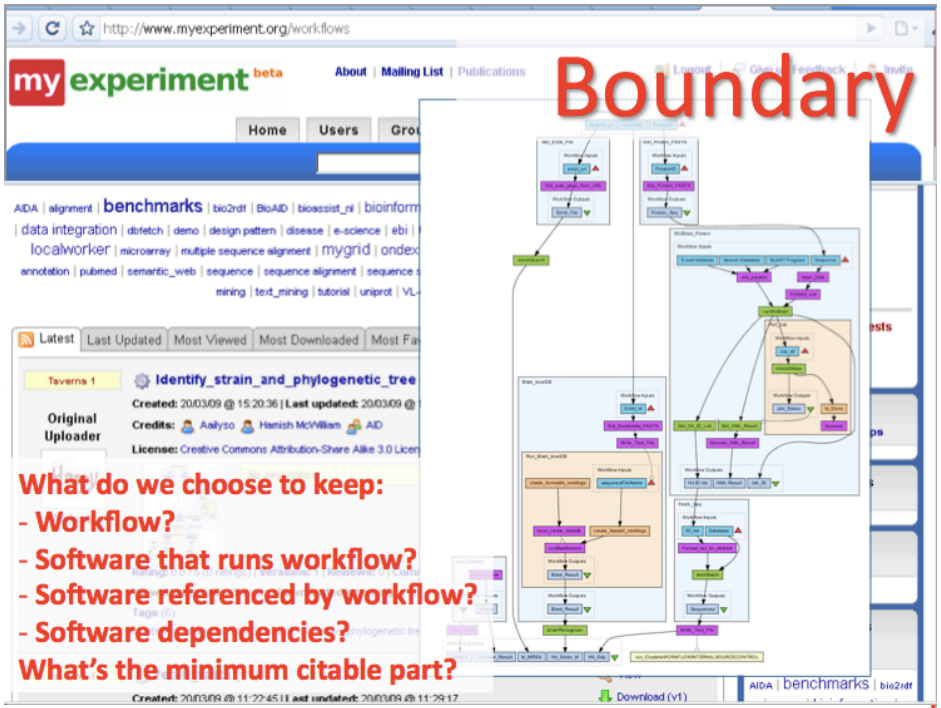
www.software.ac.uk

Isn't software just data?

http://beyond-impact.org/?p=175

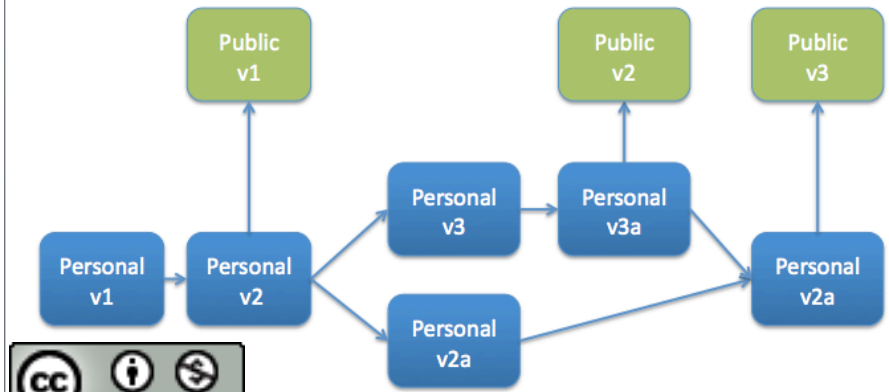
- Journal of Open Research Software
  - http://openresearchsoftware.metajnl.com/
- Role of Repositories
  - http://www.era.lib.ed.ac.uk/handle/1842/5905
- Publication for Discovery
  - JISC-funded Software Hub project

Boundary



Versioning

Why do we version?
- To indicate a change
- To allow sharing
- To confer special status



Granularity

And we haven't talked about authorship yet…

5 Stars of Software?

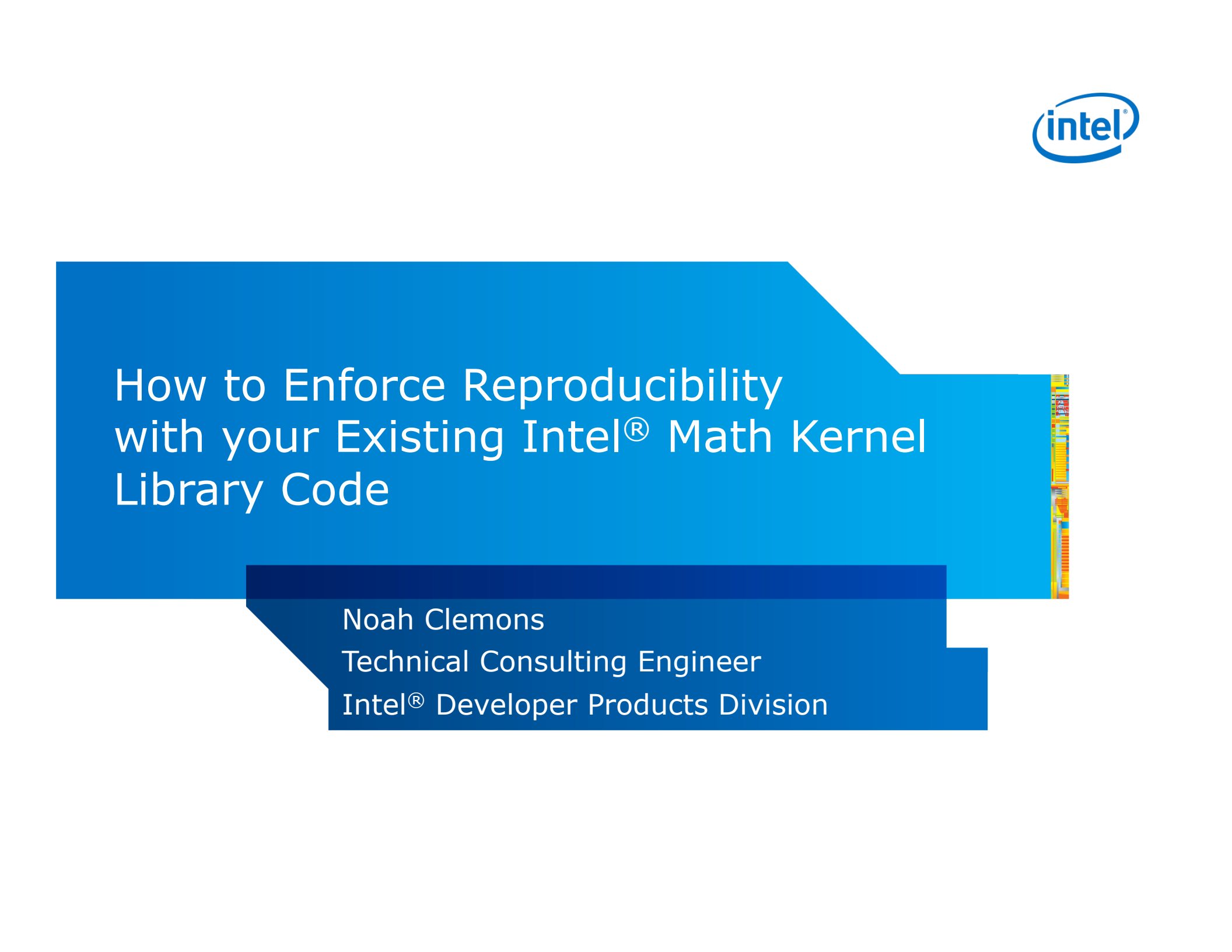
www.software.ac.uk

- Do we need a 5 stars for software?
  - Existence – there is accurate metadata that defines the software
  - Availability – you can access and run the software
  - Openness – the software has an open permissible license
  - Linked – the related data, dependencies and papers are indicated
  - Assured – the software provides ways of assuring its "correctness"

c.f.
5 Stars of Linked Data (Berners-Lee)
5 Stars of Online Journals (Shotton)

NOAH CLEMONS

HOW TO ENFORCE REPRODUCIBILITY WITH YOUR EXISTING MKL CODE



How to Enforce Reproducibility with your Existing Intel® Math Kernel Library Code

Noah Clemons
Technical Consulting Engineer
Intel® Developer Products Division

3 Types of Non-Reproducibility in Intel® Math Kernel Library

- Run to Run – same processor
- Runs between different Intel processors
- Runs between different IA-compabible processors

Maximum Compatiblity ↕ Maximum Performance	For consistent results …	Function Call mkl_cbwr_set(…)	Environment Variable MKL_CBWR=
	on Intel® or Intel®-compatible CPUs supporting SSE2 instructions or later	MKL_CBWR_COMPATIBLE	COMPATIBLE
	on Intel® processors supporting SSE2 instructions or later	MKL_CBWR_SSE2	SSE2
	on Intel processors supporting SSE4.2 instructions or later	MKL_CBWR_SSE4_2	SSE4_2
	on Intel processors supporting Intel® AVX or later	MKL_CBWR_AVX	AVX
	from run to run (but not processor-to-processor)	MKL_CBWR_AUTO	AUTO

Example – Find out the best performing option from a pool of processors

For the best option given a pool of computing resources in a grid setting, you may launch a simple program as follows

```
#include <mkl.h>
int main(void) {
  int my_cbwr_branch;
  /* Find the available MKL_CBWR_BRANCH */
  my_cbwr_branch = mkl_cbwr_get_auto_branch();
  if (!mkl_cbwr_set(my_cbwr_branch)) {
    printf("Error in setting branch. Aborting…\n");
    return;}
  return my_cbwr_branch;
}
```

The full list of options:

COMPATIBLE	3
SSE2	4
SSE3	5
SSSE3	6
SSE4_1	7
SSE4_2	8
AVX	9
AVX2	10

Examine all results and use `mkl_cbwr_set(`*<minimum_result>*`)`

DAVID KETCHESON

REPRODUCING -- AND EASILY MODIFYING -- CLAWPACK SIMULATIONS IN THE CLOUD WITH WAKARI

SEBASTIEN LI-THIAO-TE

LEPTON: LITERATE EXECUTABLE PAPERS

Lepton : Literate Executable Papers

Lepton is a tool to do research as opposed to publishing reproducible research results. It deals with :

- **everyday tasks** such as programming and writing technical reports
- **reviewing** the methods and results by collaborators and in the long term
- **re-using** source code, input data, research results

Further references :

- Website `http://www.math.univ-paris13.fr/~lithiao/ResearchLepton/Lepton.html`
  with program for download, manual, faq, examples
- 2 conference papers
  - Sébastien Li-Thiao-Té. Literate program execution for reproducible research and executable papers. *Procedia Computer Science*, 9(0):439 – 448, 2012. ICCS 2012.
  - Sébastien Li-Thiao-Té. Literate program execution for teaching computational science. *Procedia Computer Science*, 9(0):1723 – 1732, 2012. ICCS 2012.
- A poster + demonstration at ICERM

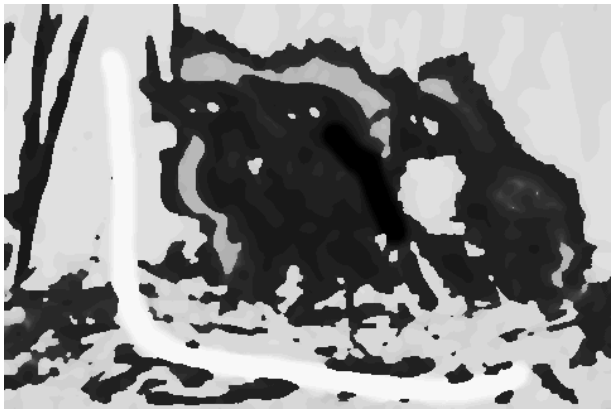
Lepton : Reproducible Research

Lepton provides :

- provenance information
  - generated documents contain all the information required to reproduce the results
- executable papers
  - a Lepton file is a program and can be executed on the local machine
- coherence and correctness guarantees
  - Lepton executes commands and automatically embeds their output
  - no copy-and-paste
- literate programming features
  - everything in the same bundle,
  - related items placed in close proximity
  - meaningful chunks
- generated, up-to-date documentation
  - run benchmarks with scripts in any language
  - format the results into tables

Lepton : In practice

Image analysis



Writing source code

Code chunk 1: «header»

bp_typedef

```
#include <stdio.h>
#include <stdlib.h>
<<bp_typedef>>
```

Code chunk 2: «bp_typedef»

```
struct bp_node
{
  gsl_vector * belief;
  gsl_vector * evidence;
  gsl_vector * m_left;
  gsl_vector * m_right;
  gsl_vector * m_up;
  gsl_vector * m_down;
};
```

Teaching statistics

Code chunk 3: «R»

```
x = round(runif(5,0,10),1)
cat(x,"\n")
```

Interpret with R

```
1.7 6.2 4.5 7.3 0.8
```

Define the sample mean :

$$\bar{x} = \frac{\sum x_i}{n}$$

Example :

$$\bar{x} = \frac{(1.7 + 6.2 + 4.5 + 7.3 + 0.8)}{5}$$

NICOLAS LIMARE

MY CHRISTMAS LIST FOR REPRODUCIBILITY

BENJAMIN SEIBOLD

STARMAP - A SECOND ORDER STAGGERED GRID METHOD FOR SPHERICAL HARMONICS MOMENT EQUATIONS OF RADIATIVE TRANSFER

MATTHIAS TROYER

PUBLISHING EXECUTABLE PAPERS

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DPHYS
Department of Physics
Institute for Theoretical Physics

Publishing executable papers

Matthias Troyer and Jan Gukelberger (ETH Zurich)
Michael H. Freedman (Microsoft)

with help from the VisTrails team,
especially David Koop, Emanuele Santos, and Juliana Freire

PHYSICAL REVIEW B **85**, 045414 (2012)

Galois conjugates of topological phases

M. H. Freedman,[1] J. Gukelberger,[2] M. B. Hastings,[1] S. Trebst,[1] M. Troyer,[2] and Z. Wang[1]
[1]*Microsoft Research, Station Q, University of California, Santa Barbara, California 93106, USA*
[2]*Theoretische Physik, ETH Zurich, CH-8093 Zurich, Switzerland*

Numerical experiments + theorem and proof

- Can we build quantum computers based on non-unitary conformal field theories?
- First reproducible numerical experiment, then theorem and proof.

PHYSICAL REVIEW B **85**, 045414 (2012)

Galois conjugates of topological phases

M. H. Freedman,[1] J. Gukelberger,[2] M. B. Hastings,[1] S. Trebst,[1] M. Troyer,[2] and Z. Wang[1]
[1]*Microsoft Research, Station Q, University of California, Santa Barbara, California 93106, USA*
[2]*Theoretische Physik, ETH Zurich, CH-8093 Zurich, Switzerland*

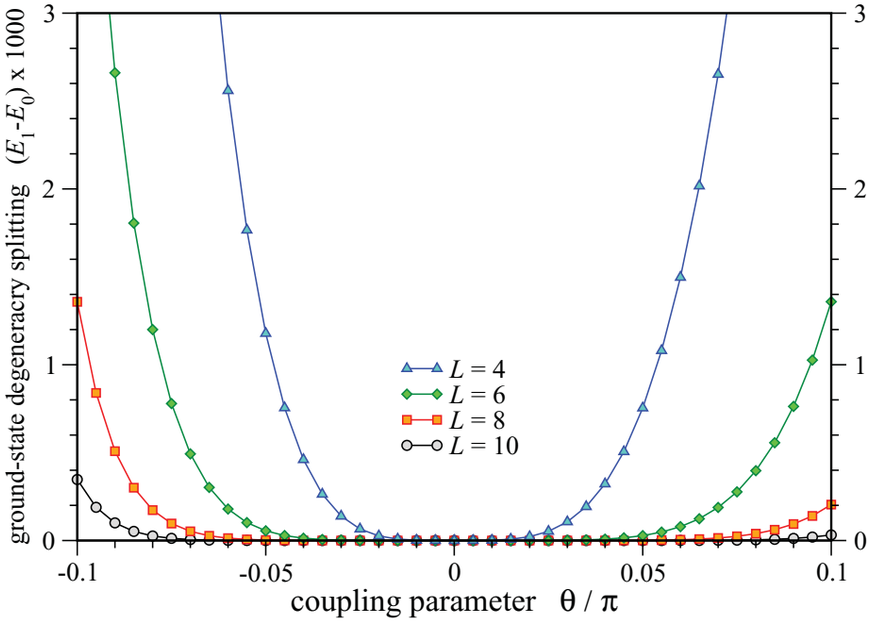
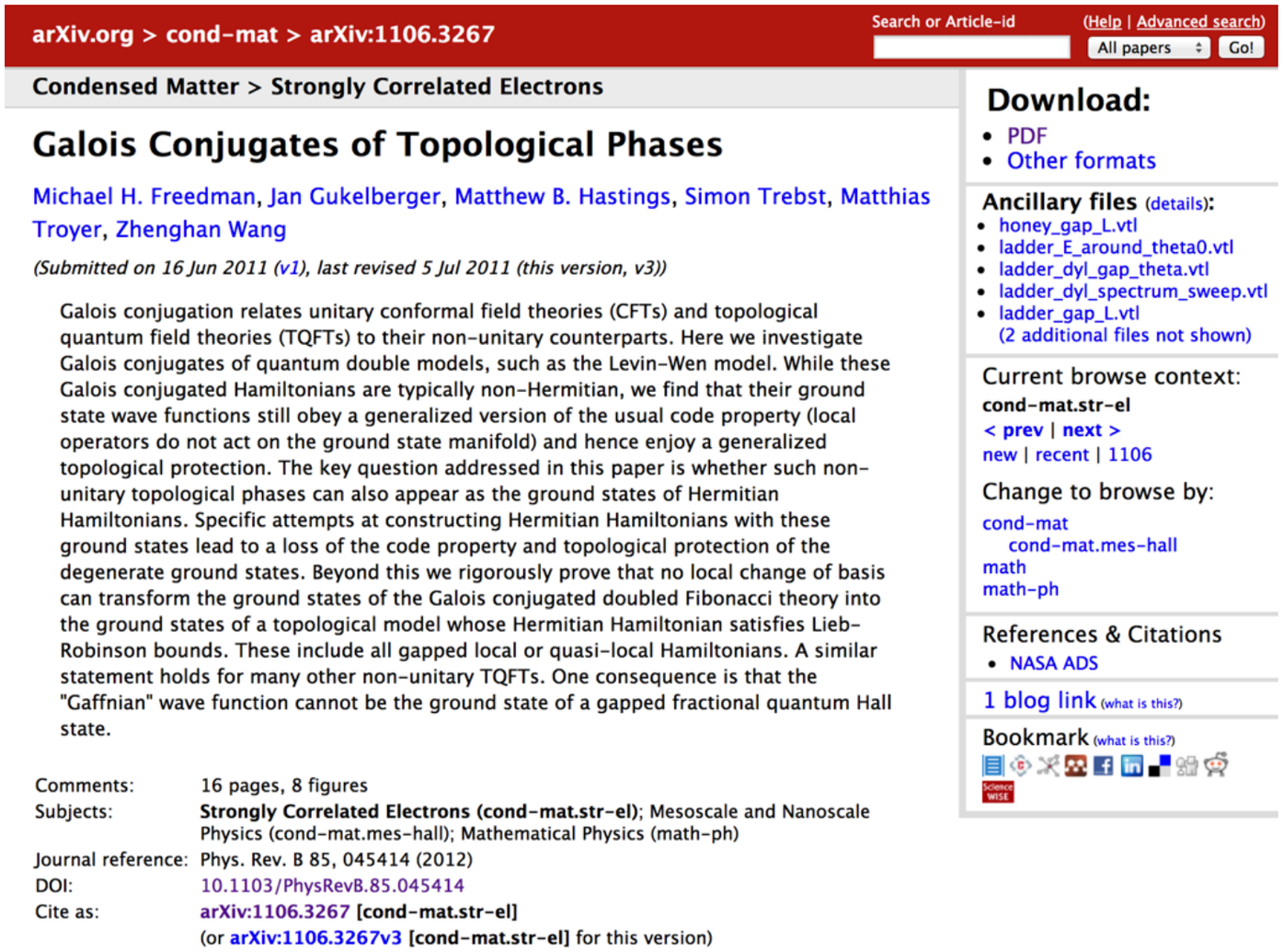


FIG. 6. (Color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$). This figure can be reproduced using the VisTrails[33] workflow Fig. 6 included in the Supplementary Material.[37]

Theorem IV.5. Fixing the number $n \geqslant 5$ and particle type $\tau \otimes \tau$ of DFib anyons on S^2 and any vertex normalization f, there can be no continuous uniform Γ family of (g.s. weakly) local normalizer operators O_Γ: $\mathcal{H} \to \mathcal{H}$, so that $\mathcal{O}_\Gamma G^{\mathcal{G}}_{n,\Gamma,f}$ is, for all anyon positions Γ, the ground-state manifold of a uniformly Lieb-Robinson and uniformly gapped family of Hermitian Hamiltonians $H(\Gamma)$ defining a topological phase [see Eq. (1)].

An executable paper: see laptop demo tonight

- The arXiv version has all data and workflow
- Clicking on the figure downloads the VisTrails workflow that reproduces the figure.



arXiv.org > cond-mat > arXiv:1106.3267

Search or Article-id (Help | Advanced search) All papers Go!

Condensed Matter > Strongly Correlated Electrons

Galois Conjugates of Topological Phases

Michael H. Freedman, Jan Gukelberger, Matthew B. Hastings, Simon Trebst, Matthias Troyer, Zhenghan Wang

(Submitted on 16 Jun 2011 (v1), last revised 5 Jul 2011 (this version, v3))

Galois conjugation relates unitary conformal field theories (CFTs) and topological quantum field theories (TQFTs) to their non-unitary counterparts. Here we investigate Galois conjugates of quantum double models, such as the Levin-Wen model. While these Galois conjugated Hamiltonians are typically non-Hermitian, we find that their ground state wave functions still obey a generalized version of the usual code property (local operators do not act on the ground state manifold) and hence enjoy a generalized topological protection. The key question addressed in this paper is whether such non-unitary topological phases can also appear as the ground states of Hermitian Hamiltonians. Specific attempts at constructing Hermitian Hamiltonians with these ground states lead to a loss of the code property and topological protection of the degenerate ground states. Beyond this we rigorously prove that no local change of basis can transform the ground states of the Galois conjugated doubled Fibonacci theory into the ground states of a topological model whose Hermitian Hamiltonian satisfies Lieb-Robinson bounds. These include all gapped local or quasi-local Hamiltonians. A similar statement holds for many other non-unitary TQFTs. One consequence is that the "Gaffnian" wave function cannot be the ground state of a gapped fractional quantum Hall state.

Comments:	16 pages, 8 figures
Subjects:	**Strongly Correlated Electrons (cond-mat.str-el)**; Mesoscale and Nanoscale Physics (cond-mat.mes-hall); Mathematical Physics (math-ph)
Journal reference:	Phys. Rev. B 85, 045414 (2012)
DOI:	10.1103/PhysRevB.85.045414
Cite as:	**arXiv:1106.3267 [cond-mat.str-el]** (or **arXiv:1106.3267v3 [cond-mat.str-el]** for this version)

Download:
- PDF
- Other formats

Ancillary files (details):
- honey_gap_L.vtl
- ladder_E_around_theta0.vtl
- ladder_dyl_gap_theta.vtl
- ladder_dyl_spectrum_sweep.vtl
- ladder_gap_L.vtl

(2 additional files not shown)

Current browse context:
cond-mat.str-el
< prev | next >
new | recent | 1106

Change to browse by:
cond-mat
cond-mat.mes-hall
math
math-ph

References & Citations
- NASA ADS

1 blog link (what is this?)

Bookmark (what is this?)

An executable paper: see laptop demo tonight

- The arXiv version has all data and workflow
- Clicking on the figure downloads the VisTrails workflow that reproduces the figure.

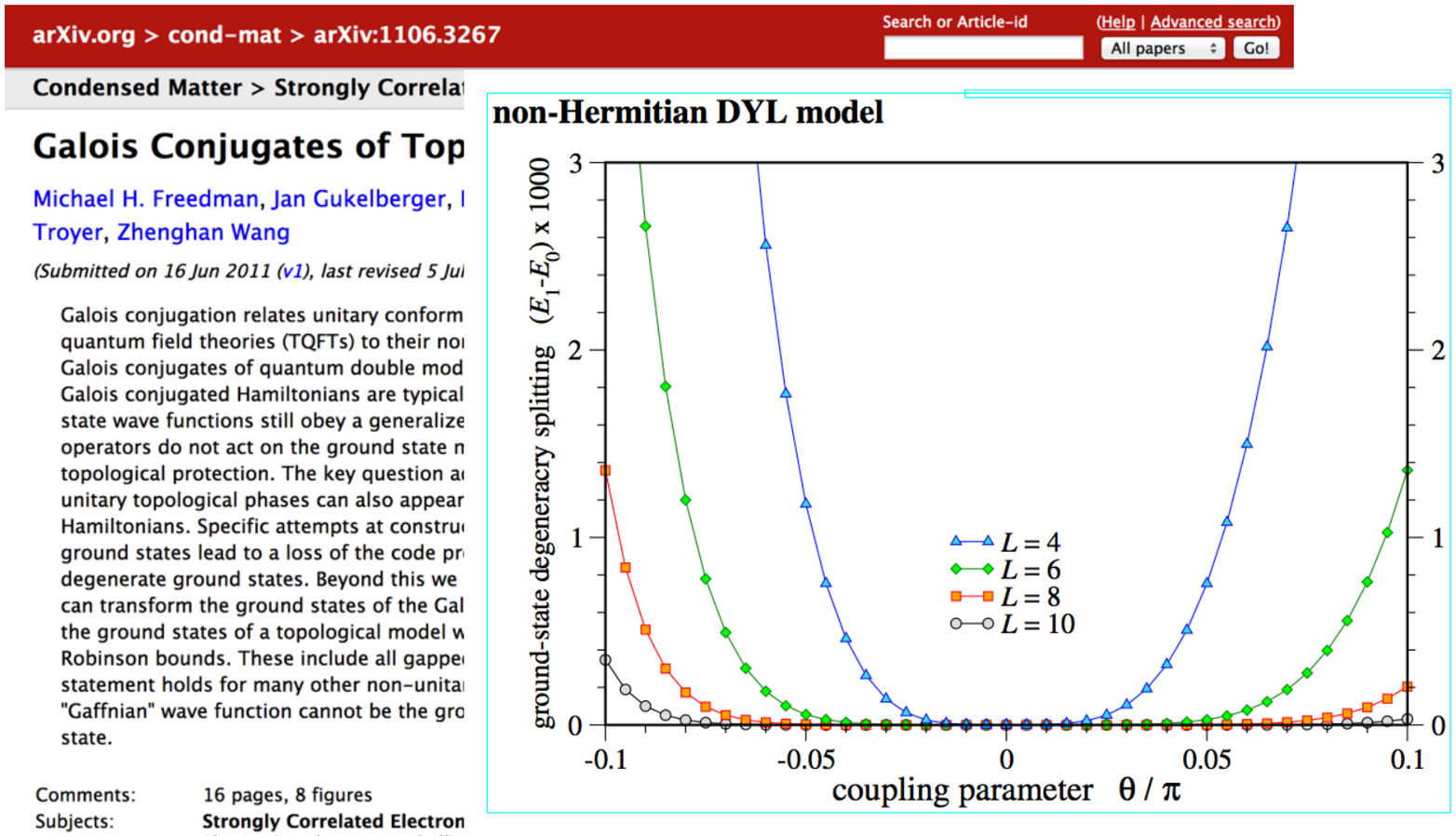
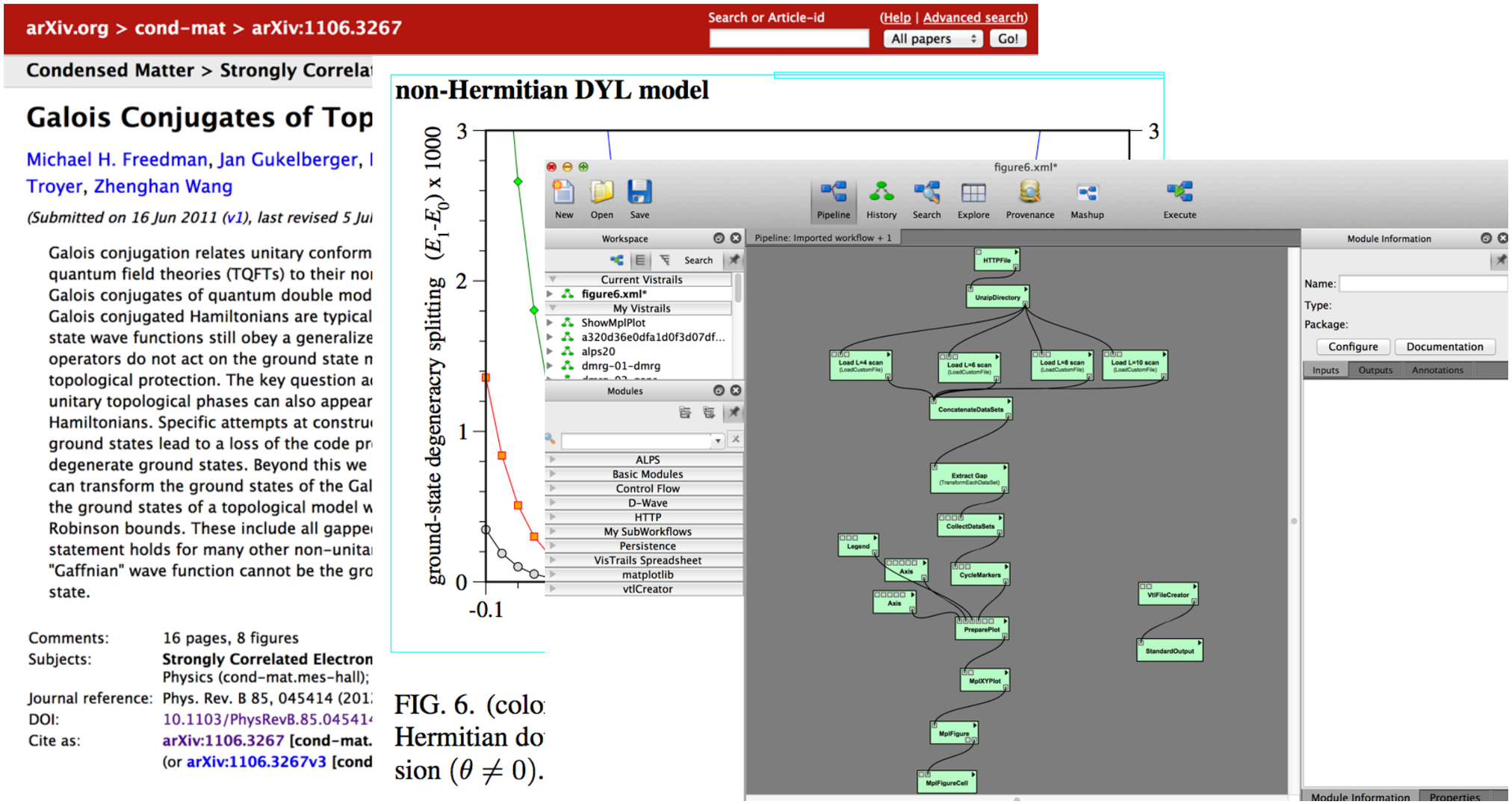


FIG. 6. (color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$).

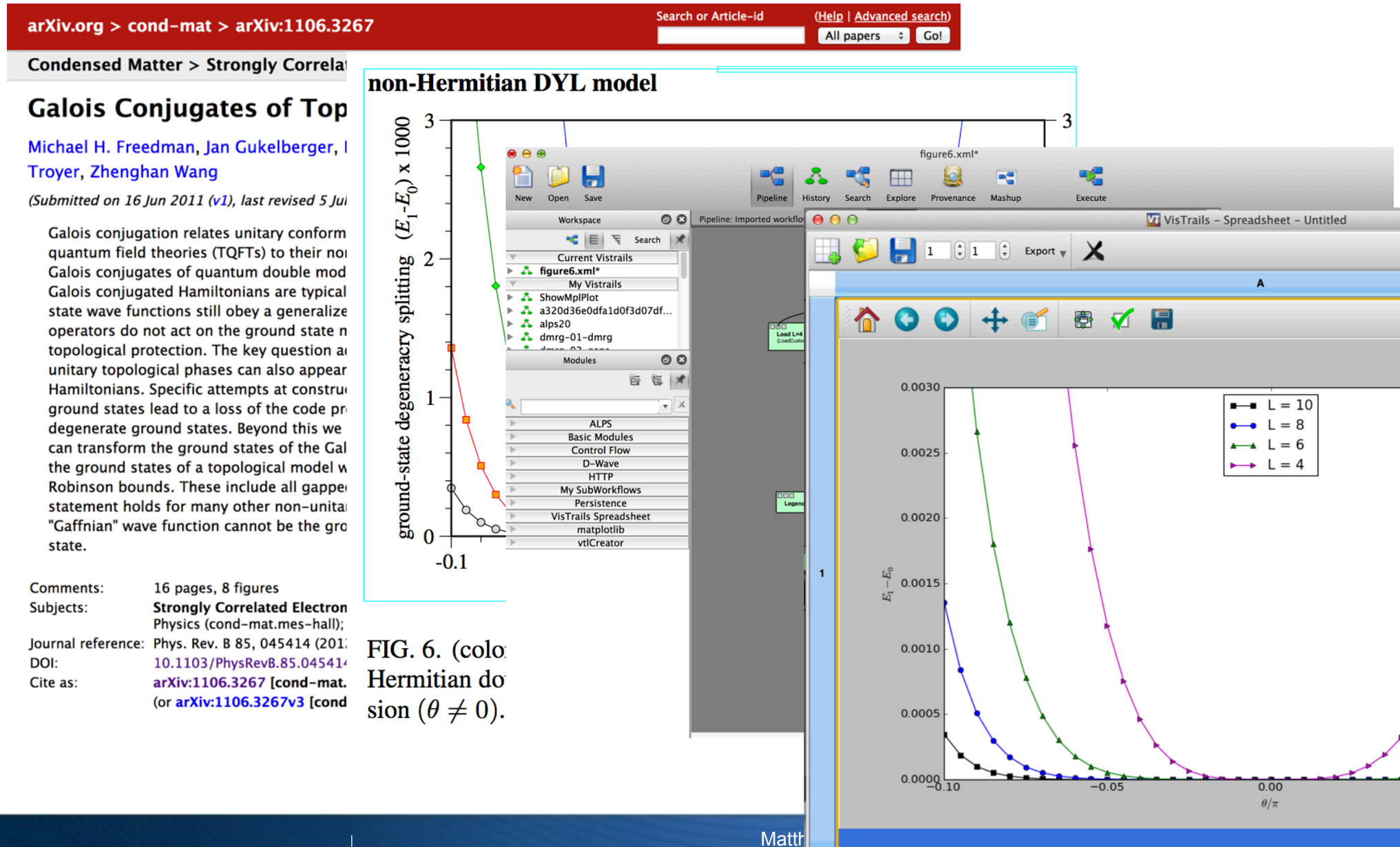
An executable paper: see laptop demo tonight

- The arXiv version has all data and workflow
- Clicking on the figure downloads the VisTrails workflow that reproduces the figure.



An executable paper: see laptop demo tonight

- The arXiv version has all data and workflow
- Clicking on the figure downloads the VisTrails workflow that reproduces the figure.



Publishers were excited!

- This is how it should be!
- Start a trial project to see how it can be made to work!

Publishers were excited!

- This is how it should be!
- Start a trial project to see how it can be made to work!

- But they soon gave up ....

Publishing requires compromises (1)

- No stable URL or DOI for supplementary material
- No link from the figure, but only a reference

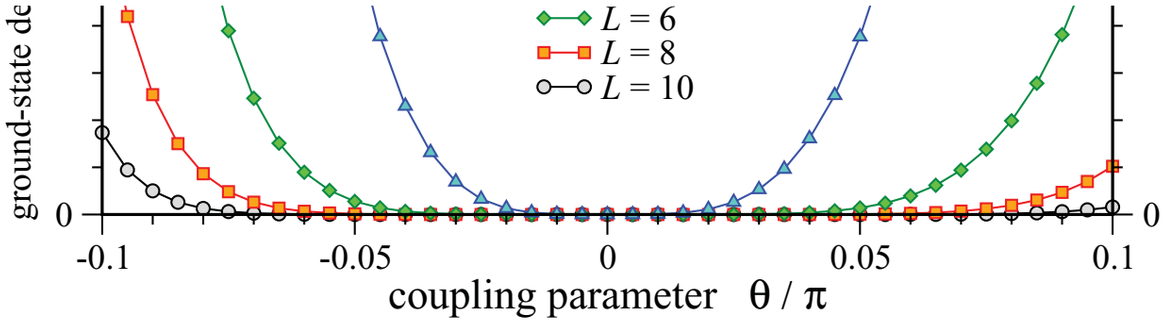


FIG. 6. (Color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$). This figure can be reproduced using the VisTrails[33] workflow Fig. 6 included in the Supplementary Material.[37]

Publishing requires compromises (1)

- No stable URL or DOI for supplementary material
- No link from the figure, but only a reference
  - reference goes to supplementary material page

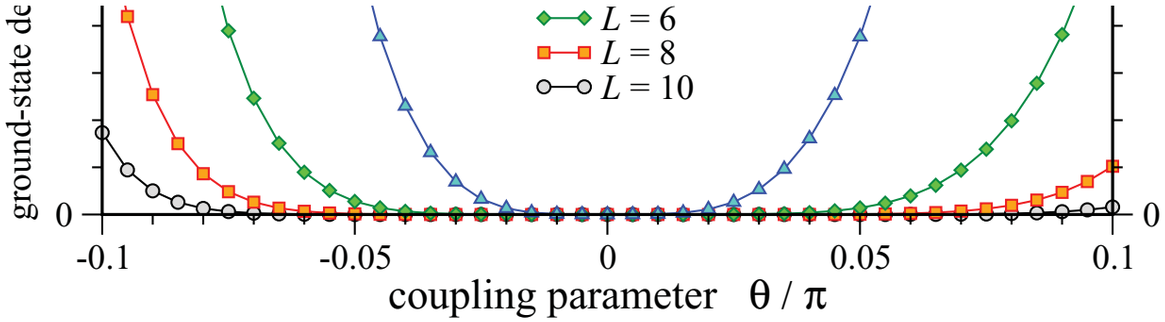


FIG. 6. (Color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$). This figure can be reproduced using the VisTrails[33] workflow Fig. 6 included in the Supplementary Material.[37]

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.85.045414 for full provenance information and workflows to recreate the figures.

Publishing requires compromises (1)

- No stable URL or DOI for supplementary material
- No link from the figure, but only a reference
  - reference goes to supplementary material page
  - find the right workflow, then download the workflow

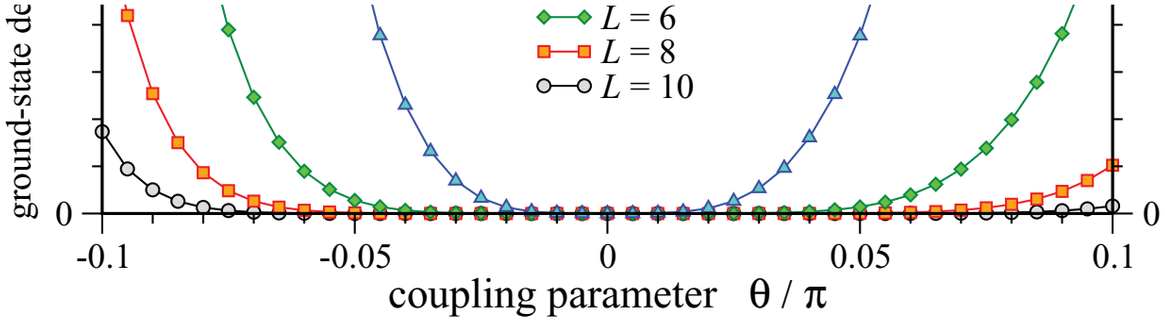


FIG. 6. (Color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$). This figure can be reproduced using the VisTrails[33] workflow Fig. 6 included in the Supplementary Material.[37]

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.85.045414 for full provenance information and workflows to recreate the figures.

Phys. Rev. B 85, 045414 (2012) [15 pages]

Galois conjugates of topological phases

Abstract | References | Citing Articles (2) | **Supplemental Material**

Download: PDF (974 kB) **Export:** BibTeX or EndNote (RIS)

- README.TXT
- figure6.vtl
- figure4b.vtl
- dyl_ladder_gap.zip
- figure7a.vtl
- proj_ladder_gap.zip
- figure7b.vtl
- proj_ladder_thetasweep.zip
- figure8a.vtl
- proj_honey_gap_vs_L.zip
- figure8b.vtl
- figure4a.vtl
- dyl_ladder_thetasweep.zip
- proj_ladder_gap_vs_L.zip

Publishing requires compromises (1)

- No stable URL or DOI for supplementary material
- No link from the figure, but only a reference
  - reference goes to supplementary material page
  - find the right workflow, then download the workflow

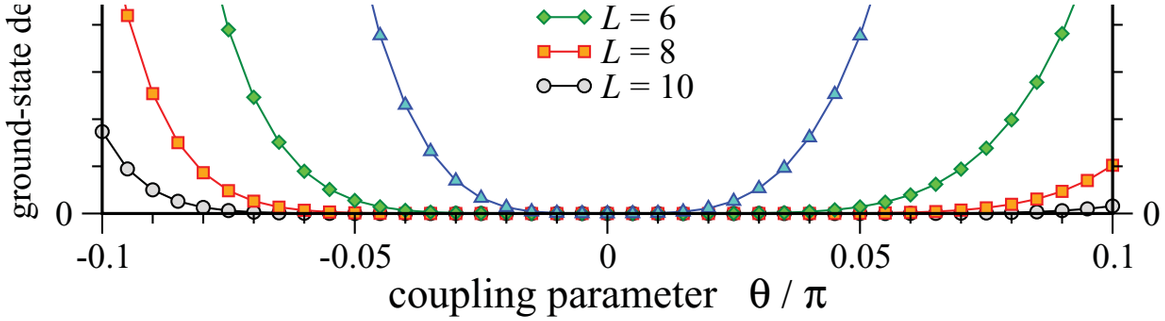


FIG. 6. (Color online) Ground-state degeneracy splitting of the non-Hermitian doubled Yang-Lee model when perturbed by a string tension ($\theta \neq 0$). This figure can be reproduced using the VisTrails[33] workflow Fig. 6 included in the Supplementary Material.[37]

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.85.045414 for full provenance information and workflows to recreate the figures.

Phys. Rev. B 85, 045414 (2012) [15 pages]

Galois conjugates of topological phases

Abstract | References | Citing Articles (2) | **Supplemental Material**

Download: PDF (974 kB) **Export:** BibTeX or EndNote (RIS)

- README.TXT
- figure6.vtl
- figure4b.vtl
- dyl_ladder_gap.zip
- figure7a.vtl
- proj_ladder_gap.zip
- figure7b.vtl
- proj_ladder_thetasweep.zip
- figure8a.vtl
- proj_honey_gap_vs_L.zip
- figure8b.vtl
- figure4a.vtl
- dyl_ladder_thetasweep.zip
- proj_ladder_gap_vs_L.zip

Publishing requires compromises (2)

- The workflow needs to fetch the raw data, but
  - No stable URL or DOI for supplementary material
  - Even unstable URL only know **after** publication
- How did we solve it?

Publishing requires compromises (2)

- The workflow needs to fetch the raw data, but
  - No stable URL or DOI for supplementary material
  - Even unstable URL only know **after** publication
- How did we solve it?
- Journal of Statistical Mechanics (JSTAT), an IOP journal
  - Production editor started publication process before the lunch break and sent us the URL
  - We had an hour to prepare final workflows and sent them back
  - He finished publication process after returning from lunch

Publishing requires compromises (2)

- The workflow needs to fetch the raw data, but
  - No stable URL or DOI for supplementary material
  - Even unstable URL only know **after** publication
- How did we solve it?
- Journal of Statistical Mechanics (JSTAT), an IOP journal
  - Production editor started publication process before the lunch break and sent us the URL
  - We had an hour to prepare final workflows and sent them back
  - He finished publication process after returning from lunch
- Physical Review, an APS journal
  - Editors told us to give up
  - Production manager informed us that we can replace the supplementary material anytime after publication without leaving a trace ....
  - We then just sent the working workflows with the right URLs for data after publication

Our next approach

- Publishers desire reproducible papers but are not yet ready to handle executable papers in the publication process

- Our intermediate solution:
  - Publish raw data and workflows through our institutional library and obtain DOIs
  - Refer to that data from the paper and just include a backup copy with the papers

YIHUE XIE

REPRODUCIBLE RESEARCH WITH KNITR AND R

knitr: Starting From Reproducible Homework

Yihui Xie

Department of Statistics, Iowa State University

ICERM, 12/12/12

$$\text{knitr} = \text{knit} + \text{R}$$

$$\text{Sweave} = \text{S} + \text{weave}$$

my homework & solutions in past three years at Iowa State[1]

[1] e.g. https://github.com/yihui/stat579/downloads

I love LaTeX more than anyone else, but do not tie users to LaTeX

I love LaTeX more than anyone else, but do not tie users to LaTeX
(but do keep them away from Word)

Markdown

Markdown

Markdown

Markdown

- markdown
- markdown

The value of π is `` `r pi` ``, and a Monte Carlo estimate is:

```{r}
est_pi = function(n) {
  x = runif(n, -1, 1)
  y = runif(n, -1, 1)
  4 * mean(x^2 + y^2 <= 1)
}
est_pi(10000)
```

The value of π is 3.1416, and a Monte Carlo estimate is:

```
est_pi = function(n) {
    x = runif(n, -1, 1)
    y = runif(n, -1, 1)
    4 * mean(x^2 + y^2 <= 1)
}
est_pi(5000)

## [1] 3.128
```

reproducible homework (happier students, happier professors)

evidence that we underestimated the power and imagination of students: `http://www.rpubs.com`

written in R, but not for R only (bash scripts, C++, ...)

If reproducible homework comes, can reproducible research be far behind?

IN CODE WE TRUST