

Open-source, knowledge sharing and scientific collaboration

Tord Riemann Königs Wusterhausen, Germany

<http://indico.if.us.edu.pl/conferenceDisplay.py?confId=0>

Matter To The Deepest, Recent Developments In Physics Of Fundamental Interactions XXXVII International Conference of Univ. Katowice, Ustron, Poland

Intro

The copyright statement for this talk:

©2013 tordriemann@googlemail.com, <http://zfitter.com>

The document is not foreseen for distribution outside the webpages of Ustron 2013.

→ Please do not download.

Please erase copies of it in your system, if any.

→ Please do not read the contents.

→ The author feels not responsible for the contents.

HEP = High Energy Physics became more and more a kind of industry.
To some extent even in theoretical physics.

The contacts of cooperating scientists get more anonymously.
This trend is supported by the internet.
At the same time, the competition became more complex.

Knowledge sharing in early times:

→ You gave a Fortran code to somebody or you did not.

→ Or alternatively: The CERNLIB model; its software was nearly free.

Knowledge sharing today:

→ many ways of distribution

→ more anonymously

Aggressiveness ...

There are folks around distributing rumors like
. . . those found in a diploma thesis in 2008 . . .
we quote from there [we highlight] :

Several fit programs exist to extract ... electroweak precision measurements and lots of results have been published in the past.

The most prominent fitting packages are **ZFitter** [1, 2] and **TOPAZ0** [3, 4].

However,

- the present situation is **unsatisfactory**. Most programs are . . .
- relatively **old**,
- **coded in Fortran** and
- **no longer maintained**.

This makes it **dangerous to rely on them in the LHC and later ILC times** when they are still needed...

The author of the corresponding diploma thesis was awarded for his work the Otto-Stern-Preis of “Freunde und Förderer der Physik” of Hamburg University

Disagreement of XXfitt C++ program and XXfitt article

Example: $127 \neq 1270$

We have the α_s^4 terms of Baikov et al.

Do we control the α_s^3 terms and all that sufficiently?

Let us look into some actually promoted code from a competitor:

$$A_3 \sim m_{CA3} = -4544045/864.0 + 1340\zeta(2) + 118915/36.0\zeta(3) - 1270\zeta(5) \\ + (71621/162.0 - 209/2.0\zeta(2) - 216\zeta(3) + 5\zeta(4) + 55\zeta(5))m_{nf1} \\ + (-13171/1944.0 + 16/9.0\zeta(2) + 26/9.0\zeta(3))m_{nf1}^2;$$

In the **description of the competitor** [in fact a diploma thesis] we find:

$$A_3 \sim C_{23}^A = -\frac{4544045}{864} + 1340 \zeta(2) + \frac{118915}{36} \zeta(3) - 127 \zeta(5) \\ + \left[\frac{71621}{162} - \frac{209}{\zeta} (2) - 216 \zeta(3) + 5 \zeta(4) + 55 \zeta(5) \right] m_{nf1} \\ + \left[-\frac{13171}{1944} + \frac{16}{9} \zeta(2) + \frac{26}{9} \zeta(3) \right] m_{nf1}^2.$$

These NNLO terms do not agree.

Look into the **original reference given**:

"QCD corrections to the e+e cross-section and the Z boson decay rate"

Chetyrkin, Kühn, Kwiatkowski, Dec 1994, 87 pp.

In: "Reports of the working group on precision calculations for the Z resonance", pp. 175-263, e-Print:

hep-ph/9503396

There is no formula of this kind at all . . .

Agreement of XXfitt C++ program and ZFITTER v.6.42 Fortran program

Then we look into:

Bardin, Christova, Jack, Kalinovskaya, Olchevski, Riemann, Riemann

ZFITTER, CPC133 (2001): [1, Bardin:1999yd]

$$\begin{aligned}
 A_3 \sim C_{23}^A &= \text{COEFA3} = -\frac{4544045}{864} + 1340\zeta(2) + \frac{118915}{36}\zeta(3) - 127\zeta(5) \\
 &+ \left[\frac{71621}{162} - \frac{209}{\zeta}(2) - 216\zeta(3) + 5\zeta(4) + 55\zeta(5) \right] m_{nf1} \\
 &+ \left[-\frac{13171}{1944} + \frac{16}{9}\zeta(2) + \frac{26}{9}\zeta(3) \right] m_{nf1}^2;
 \end{aligned}$$

Here we see the $-127\zeta(5)$ of the description and not the $-1270\zeta(5)$ found in the C++ code.

The CPC133 is the only place where to find the expression ...

And finally ($D5 = \zeta(5)$):

In the Fortran program ZFITTER/dizet6_42.f:

$$\begin{aligned}
 \text{COEFA3} &= -4544045D0/864 + 1340 * D2 + 118915D0/36 * D3 - 1270D0 * D5 \\
 &+ (71621D0/162 - 209D0/2 * D2 - 216D0 * D3 + 5D0 * D4 + 55D0 * D5)ANF \\
 &+ (-13171D0/1944 + 16D0/9 * D2 + 26D0/9 * D3)ANF^2*
 \end{aligned}$$

This number agrees with the C++ code of the competitor of 2009 ..., but not with its description ...

Explanation: private communications, copy-paste, wrong and incomplete referencing and all that.

Another lesson – When copying try to be correct ...

In ZFITTER file `bkqcdI5_14.f` we find the Fortran function `XRMQCD`, which is authored by ZFITTER:

```

FUNCTION XRMQCD (AMZ2, AMW2, AMT2, S)
...
XRMQCD=1D0 / (12D0*SW2)
&      * (3D0/4D0/CW2 * (1D0+VB2)
&      +AMT2/4D0/AMW2 * (VT2 * (XDVFTZ/ALTZ-XPVFTZ)
&      +XDAFTZ/ALTZ-XPAFTZ)
&      -AMT2/AMW2 * (3D0*D2+105D0/8D0) )
ELSE
XRMQCD=1D0/4D0/SW2/CW2
&      * (AMT2/AMZ2 * ((1D0-4D0*QTM*SW2)**2*XV1r+XA1r)
&      +AMT2 / (AMZ2-S) * ((1-4*QTM*SW2)**2 * (XV1rs-XV1r) +XA1rs-XA1r)
&      +2D0*AMT2/AMZ2 * (-23D0/8D0+D2+3D0*D3)
&      -1D0/4D0 * (1D0 + (1D0-4D0*QBM*SW2)**2) * S / (AMZ2-S) * LOG(S/AMZ2) )

```

In a C++ code of a competitor we see similar but not identical coding:

```

rmqcd = 1.0 / (4.0*SW2*CW2) * ( mt2/MZ2*(VT2*XV1r + XA1r)
                                + mt2 / (MZ2-S) * (VT2 * (XV1rs - XV1r) + XA1rs - XA1r)
                                + 2.0*mt2/MZ2 * (-23.0/8.0*D2+3.0*D3)
                                - 1.0/4.0 * (1.0 + QBM) * S / (MZ2-S) * TMath::Log(S/MZ2) );

```

One may observe two mistakes, resulting from copy-paste with a loss, in C++/rmqcd compared to ZFITTER, where we have:

$$\begin{aligned}
 \dots &= \\
 &+2D0 * AMT2/AMZ2 * (-23D0/8D0 + D2 + 3D0 * D3) \\
 &-1D0/4D0 * (1D0 + (1D0 - 4D0 * QBM * SW2) ** 2) * S / (AMZ2 - S) * LOG(S/AMZ2)
 \end{aligned} \tag{1}$$

In both cases: the $D3 = \zeta(3)$ comes from a QCD 2-loop Formula (Kniehl 1990).

The conflict of ZFITTER and XXfitt lasts since March 2011

Memorandum on ZFITTER/Gfitter

Prof. A. A. Akhundov, Prof. P. Christova, Dr. S. Riemann, Dr. T. Riemann et al.

contact: <http://zfitter.com>, tordriemann@googlemail.com

4 November 2012 v2.9.11.12 - v2.1.10.11.12 - v2.2.12.11.12 - v2.3.16.11.12 - v2.4.19.11.12

Since 1 March 2011, members of the ZFITTER group A. Akhundov, S. Riemann, T. Riemann, aided by further scientists, have proven substantial violations of the commonly accepted Rules of Good Scientific Practice by the Gfitter collaboration when using scientific results of the ZFITTER group.

The Gfitter collaboration has not admitted this publicly, and the necessary consequences were not drawn.

The only sanction so far is a disciplinary measure taken against the ZFITTER spokesperson on 30 September 2012.

What does ZFITTER propose in the present situation?

The Gfitter group should consider to ...

- ... Accept the validity of the CPC licence hold by ZFITTER authors.
- ... Accept that ZFITTER authors have personal copyrights with the ZFITTER project.
- ... Admit the deviations from Good Scientific Practice when using software and text of ZFITTER authors from Summer 2006 till Summer 2011.
- ... Understand that the text of the so-called erratum to EPJC60(2009)543 and the corresponding phrases in arXiv:1107.0975v1 are deliberately misleading and incorrect. A code is introduced which does not exist, with authors who do not agree, in order to heal the conflict with ZFITTER. Ignoring the expressed positions of the ZFITTER authors.
- ... Write the necessary errata to publications. We mention three main publications – a diploma thesis, the article EPJC60(2009)543 and arXiv:1107.0975v1 – plus further journal articles and hep-ph submissions and talks linked at the Gfitter webpages.
- ... Change the Gfitter webpages at DESY and CERN appropriately.
- ... Start a dialogue with ZFITTER on how to deal with the Gfitter/GSM code. Note that Gfitter cannot publish the Gfitter/GSM code without written permission from the ZFITTER authors because it is derived from ZFITTER.

DESY is asked to consider to ...

- ... Respect the administrative regulations of GO/BO (GO for Geschäftsordnung) and the recommendations of DEG related to issues of scientific misconduct when treating matters related to

A collection of statements – authors

A1 an author says:

When I create software, I want to get cited for its use.

I a way which I define.

Sometimes by applying a GPL-type licence, sometimes by references in articles, etc.

A2 an author or an institution says:

Our software has not to be touched by the user.

Because I guarantee for its high standards.

Because it is a standard candle [etalon] for others, and it was created in order to be so.

So, please link my software to yours, or refuse from usage.

Or, please write your interface to my package as a whole.

A collection of statements – users

U1 a user or an institution says:

I need open-source software.

And I can tell you what that is.

Because I have to adapt and develop your software for my purposes, in favor of the scientific progress.

If you do not want misuse, you cannot publish your software.

This is the only safe way. I go this way.

U2 - a user or an institution says:

Any software in the internet with anonymous download is open-source software.

U3 - a user or an institution says:

There are no commonly accepted rules for the use and citation of open-source software.

Of course, I need not cite the software when using it.

And I can do what I like, even if the authors claim to have a licence.

A bit of discussion

Academic research

Let us discuss exclusively about academic research.

And only about academic software.

Just to be definite.

This would exclude e.g. any reference to commercial software.

Internationality

We live in an international community. As a consequence, national law, national licences, institutional regulations are not valid automatically. See:

http://en.wikipedia.org/wiki/Berne_Convention_for_the_Protection_of_Literary_and_Artistic_Works

Long-term projects with many authors

Many of us work on long-term projects, often in teams, sometimes in huge teams with changing compositions.

Since the Renaissance epoche, the tradition of citation of the work of others [works = creations] became more and more an essential part of scientific ethics in basic research.

This is beyond commercial arguings, but not beyond of material interests.

Equilibrium of competition and cooperation

Academic research is based on an equilibrium of competition and cooperation
Academic researchers depend in many respects on the **recognition of their contributions to scientific progress**.

We have to fight for ...

- project money
- money for PhD and postdoc positions
- permanent positions
- resources, e.g. clusters of comuters for large-scale calculations
- professorships, better professorships, Nobel prizes
- etc. etc. etc.

Attribution

The equilibrium of Competition and Cooperation gets disbalanced when researchers use the work of others without quotation [more general: attribution]

Conditions of use, licences and all that I

In practice, there are additional expectations of the creators of scientific work to their users. Accepted by society, as for creators of work in general.

Remember the regulations in the general internet for photographs, videos, music etc. etc.

One has to care about the conditions of use when downloading anything.

Maybe in form of licences = standardized conditions of use.

For software, there may be certain very specific regulations. E.g. in German law this applies.

But there are not so many persons who assume that software has no genuine scientific content, i.e. is not a result of creative work.

Although, in scientific practice it may happen that software is considered of minor relevance when accounting scientific ingenuity and scientific progress.

Because national law often cannot be applied in practice, it is of importance that researchers feel an ethical need to respect the “conditions of use”, formulated by the authors of software.

Conditions of use, licences and all that II

Part of the “conditions of use” are licences.

Examples [please search for details by yourself]:

- **Gnu public licence = GPL** and the derivatives, like e.g. the lesserGPL. Is often used and/or recommended.
But: seems not appropriate for academic software:
GPL does not expect proper citation.
- **Creative Commons Licence = CC** with derivatives. → See below.
Seems to be OK, although often not recommended for software.
- **Computer Physics Communications software deposit licence** → See below.
Was in use for decades by e.g. GEANT, MINUIT, ZFITTER, FF, etc.
Now the users of the CPC software deposit have a choice of licence.

Conditions of use, licences and all that III

- The decision on a licence is taken by the authors of the software (and maybe others)
– but not by the users.
- It is obvious that there is not the one appropriate licence model for all academic software projects and their purposes.

Have in mind:

There may be [several] employers [in different countries].

Sometimes a project is supported by some organization (e.g. in Germany DFG Deutsche Forschungsgemeinschaft) and undergoes demands from that, e.g. to make the results publicly available in appropriate form.

Sometimes a project is made in a collaboration like ATLAS or CMS; then there might be stringent reasons NOT to make it public due to Competition.

Sometimes software is made in an institution which does not mind at all.

In the view of some experts of law, these licences are so-called licences, i.e. no true licences.

This might be true because national law might have certain opinions on what is a licence.

But: Not to respect these licences is not a good concept.

Because the licences in use in basic research are, in the best case, a commonly accepted frame of use of work of others.

They are, in practice, the rules of the game.

If somebody recommends not to respect licences formulated for scientific, non-commercial, academic software, he/she should then explain what else is the basis of an agreed use of the software.

Some researchers find a discussion of the rules of using software not necessary, because the lawyers of their institutes care about that.

Counter argument:

Legality of use of software by third parties is not a case for lawyers.

By no means: We are poor. We have no time. We are not mighty.

Certainly, there happen quite different situations.

Software created in a huge experimental collaboration like ATLAS or CMS often underlies agreements of the parties.

So, ideally there are well-defined partners and they certainly respect the rules.

Published academic theory software.

If it is made available for others often it is not -, often the “others” are not pre-defined. Then one has to hope as author, that the users respect the interests of the authors. In the sense defined above.

Unpublished academic theory software. Like BlackHat. Here is no problem. But do we want this? Remember: Cooperation...

If researchers use software of others, but do not publish their own software, do not quote the use of the software of others do we accept this as a honest model of academic research?

Is the distribution of software in form of executables to selected circles of users preferable? Or is it ethically forbidden?

Academic research and institutions I

Researchers in academic research are paid by society.
There are expectations to be fulfilled

- Seeking for the truth of Nature
- Honesty against the public and among ourselves
- Making achievements publicly available

Sometimes the research institutions develop the idea that they are the true holders of the scientific achievements and can - more or less free of any boundary condition - decide what has to happen with a work, e.g. with a certain software. Evidently, there are arguments PRO.

But there are also arguments CONTRA:

→ Assume that a software is created by a postdoc with a scientific carrier in mind, hired for a short term. Would he/she like to have a certain amount of control? Should he/she have this? In a collaboration? Or in general?

→ Assume a software is created by researchers of **N1 Countries**, **N2 institutions**, over **N3 years**, financed by **N4 funding agencies**, published in **N5 journals/archives/etc.**, in changing compositions, etc. What is here the role of institutions? It might not be distinguished.

Academic research and institutions II

→ Assume the software has to be supported over a longer period.

By whom? An institution??? This may happen.

Independent of the copyright of institutions, however this might be defined, there is the question:

Do software authors have any copyright at all?

E.g. the right to be quoted?

It is evident that any regulation, which is hindering the arbitrary use of software, might also hinder scientific progress.

But, on a longer term, would the absence of copyright regulations destroy the equilibrium of Competition and Cooperation?

And thus, on a longer term, scientific progress?

Dialogue and Respect – but who are the “players”?

We did not introduce so far the terms: Dialogue and Respect.

It is fatal if there is a lack of the two.

Dialogue and Respect help to establish the ethical stability of basic, academic research.

Players in the world of rules of using the creations of others – here of software:

- **Countries** with their national ethical and law systems
- **Universities and research labs**
- **Experimental, but also theoretical collaborations**
- Last, but not least: **Single researchers who create knowledge**
The latter sometimes are not considered as players because they are usually hired.
But research relies on the acceptance of responsibility by the creators of works.
Responsibility in any respect, e.g. for correctness, user support and further developments, but also the care about the fate of the creation.

You need special “conditions of use”? → Ask the authors!

Different kinds of “rights”

There are different categories of authors rights, among them

- The right to get **attribution** – e.g. by citations
- The right to formulate **conditions of use** – e.g. in form of licences
- Both are part of the **copyright** – with national differences

And there are the different violations of them ...

- ... including the category of “**Plagiarism**” = use without attribution

Suspicion of plagiarism should lead to a three-step-procedure

- Investigation of facts
if leading to accusations, then also ...
- Estimation in some ethical or legal frame, potentially leading to ...
- Sanctions

Concluding remarks

We did not mention: **Role of scientific journals**

Many of them have documents like

Springer Publisher's "Policy_on_Publishing_Integrity.pdf" (2010)

We did not mention:

open-source software

versus

source-open software

or

public software – might be executables or n-tuples or source-open software

We did not mention:

Importance of reproducibility:

Physics is a natural science and relies on reproducibility.

- **Copied software** is not an independent scientific tool and should not claim to be.
- **Confidential software**, when used for discoveries like the Higgs, makes physics a non-scientific adventure.

What is important?

- Equilibrium of **Competition and Cooperation**
- **Dialogue and Respect**
- We authors are players in the game.
- Conditions of use (licences) are formulated by authors, not by users. Make them explicit!
- **Publish ASAP** = As Source-open As Possible.
- **Institutions should have compliance statements ...**
see “*CERN Code of Conduct*” (July 2010) or “*Leitbild of DESY*” (August 2013)
... and not only definitions of plagiarism and lists of sanctions.

Creative Commons licence I

English:

<http://creativecommons.org/licenses/by-sa/3.0/>

Chinese:

http://creativecommons.org/licenses/by-sa/3.0/deed.zh_TW

The Licenses:

Attribution [CC BY]

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

Attribution-ShareAlike [CC BY-SA]

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical

Creative Commons licence II

terms. This license is often compared to copyleft free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

Attribution-NoDerivs [CC BY-ND]

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

Attribution-NonCommercial [CC BY-NC]

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

Attribution-NonCommercial-ShareAlike [CC BY-NC-SA]

Creative Commons licence III

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

Attribution-NonCommercial-NoDerivs [CC BY-NC-ND]

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they cant change them in any way or use them commercially.

License chooser:

<http://creativecommons.org/choose/>

Computer Physics Communications – The Program Library I

<http://cpc.cs.qub.ac.uk/>

Computer Physics Communications Program Library

Programs in Physics & Physical Chemistry

Introduction

With the aid of a grant from the UK Science Research Council the Computer Physics Communications International Program Library was established at The Queen's University of Belfast in 1969 by Professor P G Burke CBE FRS. The Program Library was financially self-supporting but non-profit making. In 1996 the CPC Program Library became an integral part of the Elsevier Science journal Computer Physics Communications. The Library's function is the storage and dissemination of refereed computer programs in physics and physical chemistry, whose detailed descriptions have been published in the journal Computer Physics Communications. In addition, the Program Library also provides: access via the Internet to the Library programs, and to the corresponding full-text articles, for all members of institutes with a subscription to Computer Physics Communications; an annual subscription scheme to individual scientists who require programs relevant to their research but who do not have access through a subscription to the Journal; a free Science Direct Alert service;

Computer Physics Communications – The Program Library II

The Library Contents

The Program Library currently holds over 2200 refereed computer programs which have been contributed by scientists from all over the world. As such it represents a major repository of computational knowledge and technique. These programs range in size from under one thousand to tens of thousands of lines of code. The contributed programs have been coded in a variety of programming languages including Algol, Ada, C, C++, Lisp, Mathematica, Maple, Occam and Pascal: however, the vast majority of contributions are programmed in Fortran. Each Library program is concisely described by a Program Summary. These are indexed under the twenty-three headings ranging from Astrophysics to Statistical Physics. An information retrieval system enables the Subject Index and Author Index to be explored and the Program Summaries to be viewed and searched. A list of the most recently published programs is also available. The copyright and all other rights to each program in the Program Library remains with the program's author. By submitting a program to the Program Library the author gives the Library permission to distribute it to all those eligible to receive it under the Program Library's distribution service. All programs in the program library are covered by the Disclaimer and non-profit use Licence Agreement. Where a program is covered by a more restrictive Licence Agreement the conditions will be stated in the corresponding Program Summary printed in the CPC journal and distributed with the program source. The Disclaimer and Licence Agreement should be

Computer Physics Communications – The Program Library III

consulted before acquiring any program distributed by the Program Library. receive it under the Program Library's distribution service. All programs in the program library are covered by the Disclaimer and non-profit use Licence Agreement. Where a program is covered by a more restrictive Licence Agreement the conditions will be stated in the corresponding Program Summary printed in the CPC journal and distributed with the program source. The Disclaimer and Licence Agreement should be consulted before acquiring any program distributed by the Program Library.

The CPC Licence I

CPC Licence

Disclaimer and Non-profit use Licence Agreement

The CPC non-profit use licence agreement is an **agreement between the author(s) of a program** distributed by the CPC Program Library **and the person who acquires it**. By acquiring the program the person is agreeing to be bound by the terms of this agreement.

CPC Licence – (1)

- This licence entitles the licensee (one person) and the licensee's research group to obtain a copy of the source or executable code and to use the acquired program for academic or non-profit use within a research group; or, it entitles the licensee (one company, organisation or computing centre) to install the program and allow access to the executable code to members of the licensee's organisation for academic or non-profit use. **No user or site will re-distribute the source code or executable code to a third party in original or modified form without the written permission of the author.**

CPC Licence – (2)

- Publications which result from using the acquired program will reference the article in Computer Physics Communications which describes the program.

CPC Licence – (3)

- This licence does not permit any commercial (profit-making or proprietary) use or re-licensing or re-distributions. Persons interested in for-profit use should contact the author.

CPC Licence – (4)

- To the extent permissible under applicable laws, no responsibility is assumed and is hereby disclaimed by Elsevier for any injury and/or damage to persons or ...

References I



D. Bardin, M. Bilenky, P. Christova, M. Jack, L. Kalinovskaya, A. Olchevski, et al., ZFITTER v.6.21: A Semianalytical program for fermion pair production in $e^+ e^-$ annihilation, Comput. Phys. Commun. 133 (2001) 229–395.

[arXiv:hep-ph/9908433](https://arxiv.org/abs/hep-ph/9908433), [doi:10.1016/S0010-4655\(00\)00152-1](https://doi.org/10.1016/S0010-4655(00)00152-1).