PHY1600
Some of the Basic Essentials for Scientific Writing

Pekka K. Sinervo, FRSC
Department of Physics

"What Do You Mean I Can't Write?"

Physics
UNIVERSITY OF TORONTO

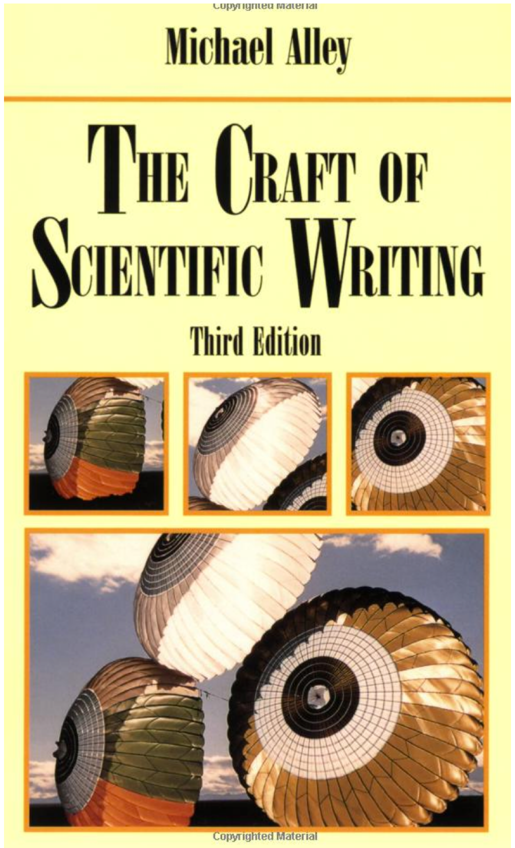
Everyone SHOULD know the basics of grammar and style, but mastery only comes from practice

- We come from various backgrounds, so have different experiences
- For some of us, English is a second language
- Correct grammar and style wasn't a particular emphasis in your UG experience
- Haven't had much practice in writing
- The good news is that it's never too late!
- The better news is that it is not that difficult to be a better writer
- Essential for anyone with an advanced degree

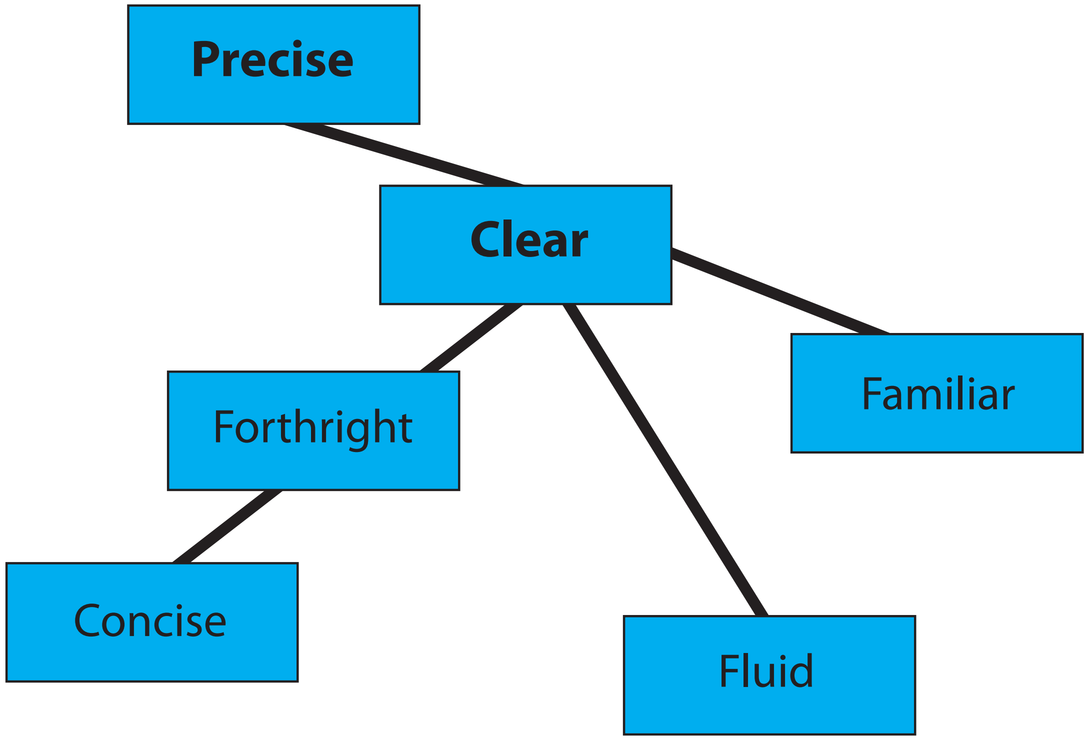
Becoming a good scientific writer – taking practice, time and attention to detail – is one of your goals

This presentation is intended to give you some of the tools to be able to write better

- First, some resources:
  - Michael Alley, *The Craft of Scientific Writing, 3rd edition* (Springer, New York, 1996)
  - Michael Alley, *The Craft of Scientific Presentations, 2nd edition* (Springer, New York, 2013)
  - Strunk and White, *The Elements of Style* (Longman, 1999)
  - APS Style Guide, http://www.apsstylemanual.org/
  - Lynne Truss, *Eats, Shoots and Leaves* (Gotham, New York, 2006)
  - Josh Bernoff, *Writing Without Bullshit* (HarperCollins, New York, 2016)



There are two primary goals of good scientific writing, clarity and precision



Taken from Fig. 1-1, Pg 12, *The Craft of Scientific Writing*

There are three organizational issues that you have to consider up-front before you start writing:

Constraints

Structure

Style

They are related to each other and need to be considered before you start.

Four constraints shape the goal of a writing project

Audience

- Who is going to read it?
- What do they know?
- Why will they be reading?
- How will they be reading?

Format

- "Format" is how the type is arranged on the page
- Often defined by a journal or publisher

Mechanics

- What style has to be used?
  - Eg. British vs American Spelling?
- Many rules in English, not all logical

Politics

- What impact will your writing possibly have?
- Will your audience be biased?

The structure of your writing plays a key role in how effective it is

Elements of Structure

- Title
- Abstract or Executive Summary
- Main Sections
- Illustrations
- Tables
- Conclusion
- Back Matter

Title should be as specific as possible and engage the audience

- Reducing the Hazards of Operation — Weak. What "Operation"?
- Interaction Free Measurements — Weak. What "Measurement"?
- Interaction Free Quantum Measurements of an Object in a Light Beam — Pretty Good. But what "Object"?
- Studies on the Electrodeposition of Lead on Copper — Not great. "Studies" doesn't say anything. Content-free
- Use of an IR FPA in Determining the Temperature of a Grating — OK, but "IR FPA"? Really? Avoid abbreviations and jargon
- Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC — Pretty good. And 5,000 citations and counting.

There are two kinds of abstracts, each of them with their own specific functions

Abstract (or descriptive summary)

- Tells the reader what the article is going to be about, and maybe summarizes key results
- Short – eg., PRL limit is 600 characters

Informative (or Executive) Summary

- Summarizes the key findings and recommendations
- Also short, often 1 page

We report two measurements of the top quark mass M_{top} using the CDF II detector at the Fermilab Tevatron in a 318 pb^{-1} data sample using ttbar events in the lepton + jets final state. One method uses an event-based likelihood technique resulting in Mtop = 173.2 ± 4.1 GeV/c^2. The second method reconstructs a top-quark mass in each event using the measured invariant mass of the hadronically decaying W boson to constrain the jet energy scale to obtain a value for M_{top} of 173.5 ± 3.9 GeV/c^2. We take the latter, which is more precise, as our result.

Sections of paper are important to organize what is presented to the reader

- Introduction
  - What is the work?
  - Why is it important?
  - What is the context?
  - How will it be presented?

- Main section and subsections
  - Describes what was done.
  - Should have a strategy
  - Often has subsections to help navigate when longer
    - Use "strong" headings!

- Conclusions
  - Restates what was done
  - Summarizes key conclusions
  - Implications for future work

A PRL article shows how the elements come together – Introduction first

The top quark is the heaviest known elementary particle with a mass 40 times that of the next-heaviest quark or lepton. Because of this comparatively large mass, top quark studies provide insight into our understanding of mass in general, and test theories that explain the large range of quark and lepton masses.

Within the context of the Standard Model of particle physics, the top quark mass is related to the masses of the W boson and the Higgs boson, the latter object being the key to our understanding of the origin of mass [1]. Precision measurements of the top quark and W boson masses test the consistency of the Standard Model, and in particular the Higgs mechanism. An improved measurement of the top quark mass is therefore a main goal of the experiments at the Fermilab Tevatron collider.

In this Letter we present two measurements of the top quark mass in the lepton + jets decay channel……

The body consists of 17 paragraphs, 3 figures and 2 tables

The DLM technique, described in detail in [8], defines a likelihood for each event based on the differential cross section per unit phase space volume of the final state partons, $d\sigma_{t\bar{t}}/d\Phi$, as a function of $M_{\rm top}$. Detector resolution effects are accounted for using $t\bar{t}$ events generated by the HERWIG Monte Carlo program [9] and full detector simulation to derive a transfer function (TF). The TF relates the transverse energies of the final state quarks denoted by $\mathbf{x}$, and the observed jets. For a given event, a Monte Carlo integration is performed over the possible $t\bar{t}$ final state kinematics in the following way: we first generate a random value for the virtual mass squared of the W boson in the leptonic channel, s_W, according to the Breit-Wigner form. We identify the momentum of the electron or muon daughter with the measured value, and the neutrino transverse momentum with the measured missing transverse energy. We then generate random values for the momenta of final state quarks according to the TF probabilities. We determine the z component of the neutrino momentum, with a twofold ambiguity, using s_W as a constraint. Thus, for a given set of $\mathbf{x}$ and s_W, we fully determine the event kinematics, and the event likelihood as a function of $M_{\rm top}$ is given by

$$L(M_{\rm top}) = N \sum_{I_j} \sum_{I_\nu} \frac{d\sigma_{t\bar{t}}}{d\Phi}(M_{\rm top}; \mathbf{x}, s_W), \quad (1)$$

where the normalization factor N is independent of $M_{\rm top}$ for a given event, and the indices I_j and I_ν run over the parton-jet assignments and the two neutrino solutions, respectively. The event likelihood is obtained by numerically integrating over $\mathbf{x}$ given by the TF and s_W given by the Breit-Wigner distribution.

Figure 1 shows the distribution of the top-quark mass value at the point of maximum likelihood in each event compared with the expectation from simulated events. An

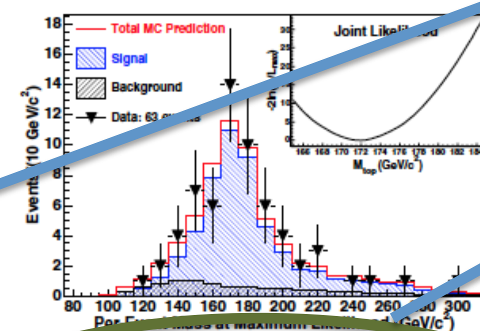


FIG. 1 (color online). The value of the top-quark mass at the maximum of the DLM likelihood is plotted for each event. Data events (points) are compared to an expected distribution (histogram) comprising simulated $t\bar{t}$ ($M_{\rm top} = 172.5$ GeV$/c^2$) and background events. The last bin includes events with masses $>$ 305 GeV$/c^2$. The inset shows the joint log-likelihood for the 63 events, before accounting for the presence of background.

of assigning the jets to the four quarks is resolved by selecting the assignment with the lowest χ^2, taking into account the b-tagging information. We construct a histogram of $m_t^{\rm reco}$ for each subsample, discarding events with $\chi^2 > 9$, corresponding to poorly reconstructed or background events.

The parameter $\Delta_{\rm JES}$ is determined within this event sample by removing the W boson mass constraints, and identifying for each event all pairs of jets that would be consistent with the W boson final state. We form histograms of the invariant masses of these jet pairs for each of the four event subsamples and compare these with what we expect given the precisely known W boson mass [11].

Unusually long paragraph – OK but be careful

Title with informative caption. Give thought to these!

Displayed equation numbered! Recommend numbering all equations!

The conclusions MUST be given plenty of consideration

In summary, we have presented two new measurements of the top quark mass. An analysis using the DLM method results in M_{top} = 173.2 ± 4.1 GeV/c^2; the analysis using the template technique results in M_{top} = 173.5 ± 3.9 GeV/c^2. There is a large statistical correlation between these measurements given the common data sample so that we quote as a result only the more accurate measurement, the template method result of M_{top} = 173.5 ± 3.9 GeV/$c^{2.}$ This provides the most precise single constraint on this important physical parameter, exceeding the accuracy of the previous world average for the top quark mass [14].

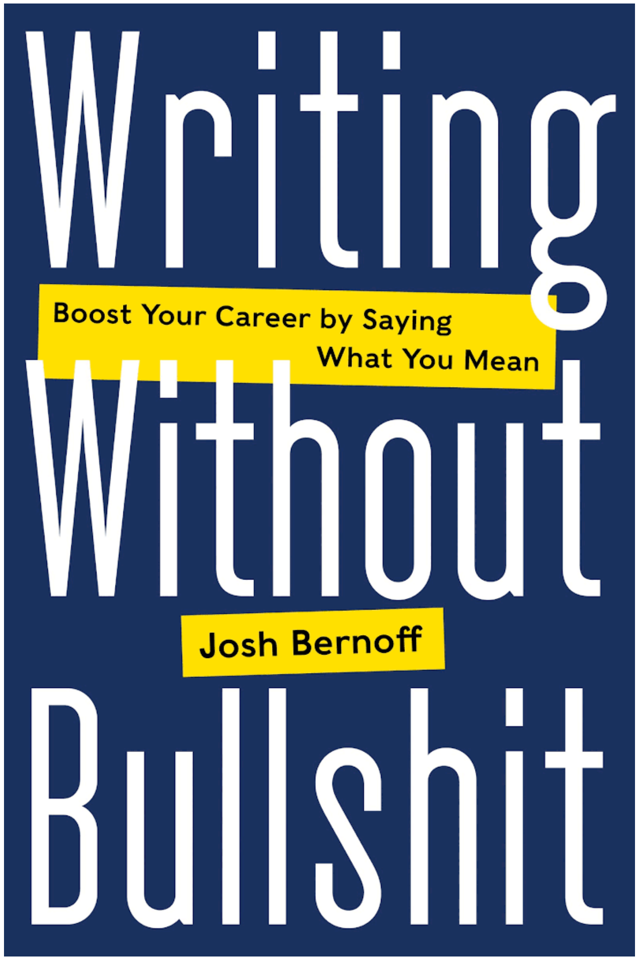
Acknowledgements

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; te National Science Council of the Republic of China; the Swiss National Science…

The style of the writing depends on the context, but there are some good guidelines to follow

Clarity	KISS, avoid complexity, avoid ambiguity
Conciseness	Omit unnecessary words, phrases, sentences
Precision	Use the correct word
	Avoid excessive use of synonyms
Familiar	Avoid jargon; remember your audience
Forthright	Avoid pretentious words, clichés
	Use concrete nouns rather than abstract nouns
	Avoid passive verbs
	Be suspicious of any sentences starting with "it"
	Be specific rather than general
Fluidity	Vary rythm
	Vary length of sentences and paragraphs
	Vary openings of sentences

Writing Without Bullshit is a screed against "poorly written crap."



"The tide of bullshit is rising.", first sentence of WWOBS.

Bernoff makes a strong case:

- More and more written material (thank you, WWW)
- More and more uneditted stuff
- Less discipline

He invokes an **Iron Imperative**:

Treat the reader's time as more valuable than your own.

Bernoff even has a way of measuring what he calls "bullshit" writing:

Innovalon is a leading technology company that combines advance cloud-based data analytics and data-driven intervention platforms to achieve meaningful insight and impact in clinical and quality outcomes, utilization and financial performance across the healthcare landscape. Inovalons's unique achievement of value is delivered through the effective progression of Turning Data into Insight, and Insight into Action. Large proprietary datasets, advanced integration technologies, sophisticated predictive analytics, data-driven intervention platforms, and deep subject matter expertise deliver a seamless end-to-end capability that brings the benefits of big data and large-scale analytics to the point of care.

Bernoff even has a way of measuring what he calls "bullshit" writing:

Innovalon is a **leading** technology company that combines **advance cloud-based data analytics** and **data-driven intervention platforms** to achieve **meaningful** insight and impact in clinical and quality outcomes, **utilization** and financial performance **across the healthcare landscape**. Inovalon's **unique achievement** of value is delivered through the **effective** progression of Turning Data into Insight, and Insight into Action. **Large proprietary datasets**, **advanced integration technologies**, **sophisticated predictive analytics**, **data-driven intervention platforms**, and **deep** subject matter expertise deliver a **seamless end-to-end capability** that brings the benefits of **big data** and **large-scale analytics** to the point of care.

$$\text{meaning ratio} = \frac{\text{meaningful words}}{\text{total words}}$$
$$= 59\%$$

Being precise is critical to effective communication of science

Example of an email posting following an analysis procedure distributed to a 3,000 member collaboration:

I have a question on :

"Analyses strongly dependent on the c-jet and light-jet uncertainties should not include them when profiling."

How would you define "strongly dependent on" ?

And what's the rationale behind forbidding people to profile some NP? In principle ….

The rejoinder didn't help much:

About "strongly dependent", it should just mean that a variation of the light and c-jet mistag rate would significantly affect the analysis result.

"NP" -> Nuisance Parameter

Here are some examples of things to avoid – redundant and wordy phrases

Redundant phrases

- “alternative choices”
- “basic fundamentals”
- “close proximity”
- “consensus of opinion”
- “as to whether”
- “clearly demonstrates”
- “for the forseeable future”

Wordy phrases

- “due to the fact that”
- “in order to”
- “in the near future”
- “great deal of”
- “as to whether”
- “made the decision”
- “has the potential to”

General agreement that writing in active voice is a more effective way of communicating

Two categories:

1. Passive verbs and passive phrases

2. 1^{st} or 2^{nd} person

1. "the voltage was displayed on the oscilloscope…"
   better as
   "the oscilloscope displayed the voltage…"

2. "The measurement was done by analyzing data…"
   better as
   "We made the measurement by analyzing data…"

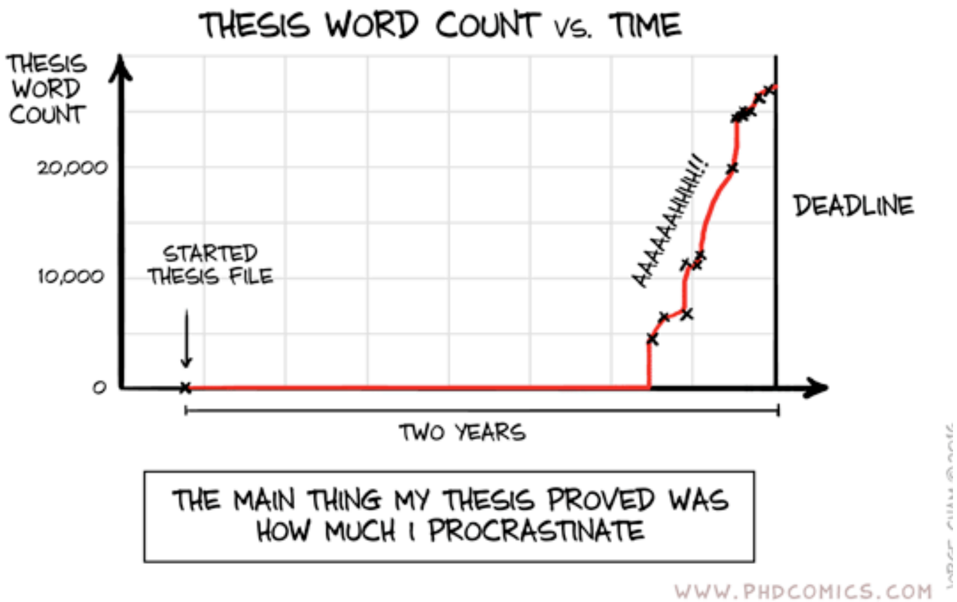
   (with the caveat that use of 1^{st} person should be tempered)

Here are a set of random notes and tips collected* over the years

- Read and study published papers in your field
- Start with a detailed and logically sequenced outline
- Write the abstract last
- Remember that first sentences are particularly important
- One set of ideas to each paragraph
- Short sentences are better – generally – than longer ones
- Redundancy is a needless repetition – but sometimes important content needs to be emphasized!
- Put yourself in the place of the audience, and think about the questions they might have
- Write first, polish after
- Then proof-read again and again, and then proof-read once more

* Credit to Prof. Tony Key, who assembled the initial list!

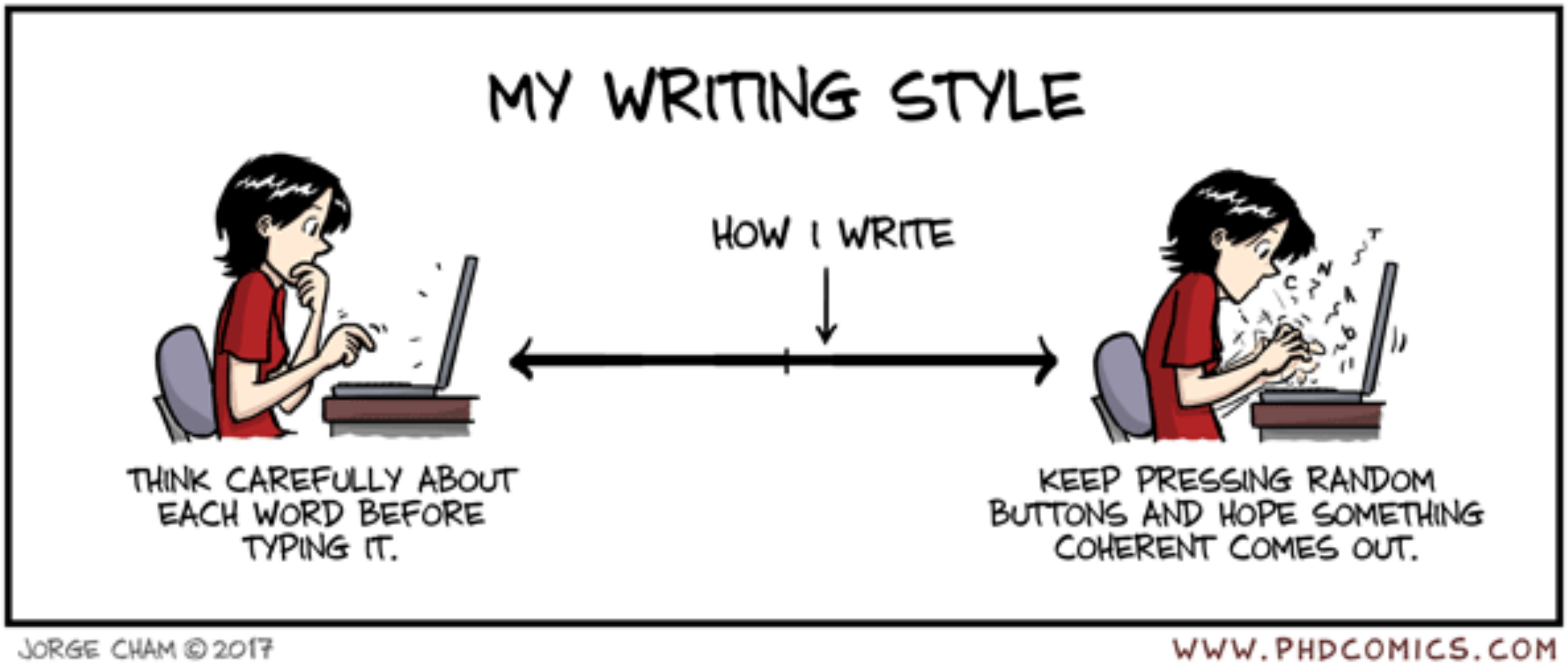
The need to polish and proof-read – and at times to rewrite – requires one to manage time carefully.



Rough rule of thumb:

- Plan to use 25% of your time to write first draft
- Use 75% to review, rewrite, polish and proof-read

Remember this is about "your time", not wall-clock time


MY WRITING STYLE
HOW I WRITE
THINK CAREFULLY ABOUT EACH WORD BEFORE TYPING IT.
KEEP PRESSING RANDOM BUTTONS AND HOPE SOMETHING COHERENT COMES OUT.
JORGE CHAM © 2017
WWW.PHDCOMICS.COM

You can become a good – perhaps excellent – scientific writer – make it one of your personal goals!