Thesis writing guide

T. Ferber,[1] **B. von Krosigk**[2]

[1]*Institute for Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany*

[2]*Kirchhoff-Institute for Physics (KIP), Heidelberg University, 69120 Heidelberg, Germany*

E-mail: torben.ferber@kit.edu, bkrosigk@kip.uni-heidelberg.de

Contents

In the past, BSc, MSc, and PhD students have asked for clear guidance on figures, units, bibliography, and many more things. This document explains best practices and some rules for a thesis for students in our groups in experimental particle physics and astroparticle physics. Most of our recommendations also apply to talks or journal publications and will prepare you for your future academic career. Some university rules or conferences/journals will necessarily override some of our preferences. We recognize the fact that other fields may have recommendations that differ from ours. If in doubt, please talk to your supervisor.

We based our best practices on our own experience, feedback from our teams[1] and on some commonly accepted best practices in our research fields. Following them will make

---

[1] Thanks to (in alphabetical order of first names): A. Heidelbach, G. Heine, L. Damer, L. Frank, J. Kahn, M. Schnepf, N. Faltermann, and S. Dreyer.

discussions much smoother and faster, and generally lead to fewer misunderstandings. We will use the words "recommended" or "should" if these are just recommendations, but not rules.

However, some items, like citations, are mandatory rules. We will use a bold "**must**" in the following text whenever something is not a recommendation but a rule. Please note that some of the more strict rules are already covered by the respective "Prüfungsordnungen" and just spelled out a bit more detailed here.

While we hold all our students to the same high scientific standards, we hope that students feel welcome to disclose any challenges they face to us. We will do our best to support and help every student based on their individual needs.

1   LaTeX

Use LaTeX to write your thesis. We recommend that you use `Overleaf` via the institutional (e.g. KIT) login. You can give *view permissions* to your supervisor(s) who can then comment directly on your LaTeX document within `Overleaf` [1]. You should generally not give *write permissions* to anyone.

2   Citations

It is part of every scientific work to put your work into context of other work in the field, or to use the results of other work. If you do so, you **must** cite the proper source(s) (see Sec. 3). Failure in doing so is considered plagiarism and can lead to a rejection of the thesis or to a retraction of the title if only discovered after graduation.

However: You **must** never copy text 1:1 from other work. If you want to copy single sentences (which is very rare in physics or technical fields but rather common in other fields), you **must** put that sentence in quotation marks and give the proper reference directly after it. You **must not** change even a single word in that sentence if you put it in quotation marks. However, if you want to use a larger part of a source and not just a single, iconic sentence, you **must** use your own words to paraphrase and give the proper reference. As before: Failure in doing so is considered plagiarism and will lead to a rejection of the thesis or to a retraction of the respective title. If in doubt, talk to your supervisor.

More information about good scientific practice can be found in [2, 3].

3   Bibliography

We recommend to use BIBTEX to organize your bibliography [4]. The recommended BIBTEX style is `unsrt`. References should be labeled as numbers in the text which is done automatically when using the `unsrt` BIBTEX style.

Following the recommendations of the Physical Review and Physical Review Letters: "The names of all authors should be given in the references. If the number of authors exceeds ten, then the first ten author names may be listed and then et al. An exception should be made for alphabetic author lists and collaborations. In these cases only the name of the first author needs to precede the et al., and for the collaborations the collaboration name should then follow."

If you cite internal reports or internal notes of your collaboration, or items that are not publicly available for other reasons, or unpublished work, you **must** provide enough information to enable the reader of your thesis to obtain a copy of that item. If possible, avoid citing non-public material. Add e-print archive numbers (like arXiv identifiers) if possible, but do not use them as the primary reference. You may add the collaboration in addition to the authors.

If you have to cite preliminary work that has been submitted to a journal but is not yet published (and hence also not yet peer-reviewed) indicate this in the bibliography by adding "submitted to <JOURNAL>" (replace <JOURNAL> with the actual journal name) if you know which journal it has been submitted to, and give the e-print archive number. If you do not know the journal or if you are not sure if it has actually been submitted, just give the e-print identifier. If possible, avoid citing non peer-reviewed work and be aware that "submitted" and "published" are two very different stages of a publication.

You **must** make sure that the bibliographic information in the references is complete. If reference citations are incorrect or incomplete (e.g. missing author names or wrong titles), this is considered bad scientific practice.

You can use MENDELEY [5] or ZOTERO [6] to organise your literature and export it to BIBTEX in the end.

Always check if the LaTeX code in titles (if any) is displayed correctly.

Example: *F. Abudinen et al. (Belle II Collaboration), "Search for Axion-Like Particles produced in e^+e^- collisions at Belle II", Phys. Rev. Lett. 125 no. 16, (2020) 161806, arXiv:2007.13071 [hep-ex].* There are more than 10 authors and this is a collaboration publication. Only the first author is given, the collaboration name follows in brackets. The paper is published in a journal and the full reference is given, the e-print identifier is given in addition to the journal reference).

4 Natural units

If you decide to (not) use natural units, always and everywhere (never and nowhere) use them which includes axis labels on your plots. We recommend to not use natural units and thus write e.g. MeV/c^2, as opposed to MeV, for masses.

5   References

We recommend to never reference sections, plots, equations, or tables manually, but always use `\ref{}` (or `\cref{}`) to call the reference, and `\label{}` to label the reference.

Example LaTeX: `see Fig.\,\ref{MYFIGURE1}`

Using `\cref{}` (from the `cleveref` package) can help getting the correct prefix (Fig., Tab., ...) automatically.

Prefixes that appear at the beginning of a sentence should not be abbreviated, i.e. write `Figure\,\ref{MYFIGURE1}` as opposed to `Fig.\,\ref{MYFIGURE1}` in these cases (same for Table, Equation, Section).

6   Equations

Non-trivial equation **must** be put in a math environment `\begin{equation}\end{equation}` and given a label that an be referenced in the text (see Sec. 5). You **must** treat those equations like a noun in a sentence, i.e. use regular grammar and punctuation, and if the equation is the end of a sentence, put a period a the end.

Trivial equations can be used as part of the text by putting them in inline math mode `$ $`. You **must not** use two-line equations, like a fraction, in inline math-mode.

Every variable or constant in an equation, no matter how trivial, **must** be defined when first used.

7   Figures

Figures are a core part of your thesis and almost every scientific paper. Take some time to set your code up so that you can include all the following recommendations.

We recommend you use MATPLOTLIB for your plots.

We recommend that you save all plots as vector graphics (e.g. `pdf` or `eps`). For 2D plots with a large number of points, the filesize may become too large for this and you should use `png` with at least 72 dpi resolution.

The following so called style guide follows closely the Belle II style guide [7] but is applicable to SuperCDMS, DELight, DARWIN, or any other experimental setting you are working with.

Plotting style:

- Show data using black markers whenever possible. Never show simulation using black markers.

- Error bars should not be capped unless caps are necessary to distinguish different contributions to the overall uncertainty displayed.

- Error bars in the x-direction should only be shown in cases in which the x-uncertainty is meaningful – that is, not to show bin size except when bins are not of consistent width, or to show an actual uncertainty as in a two-dimensional scatter plot.

- Font style (bold vs not bold, serif vs sans serif) shall not be mixed – either bold or not bold should be used. The particular font chosen should be uniform (no mixing of Arial / Helvetica, even though each is sans serif).

- Where units are needed on axis labels, they should be enclosed in round brackets[2], e.g. (GeV/c).

- Logarithmic variables x on axis labels **must** be dimensionless, for example: "$\log(x/(\text{GeV}/c))$" and not "$\log(\text{x})$ (GeV/c)".

- Bin widths **must** be indicated on axes where relevant, and should be enclosed in rounded brackets, e.g. Entries / (0.500 GeV/c^2), not square brackets.

- Colors used for plots are governed by these expectations:

  - You should not use color when unnecessary.
  - Colour palette used should be color-blind safe. A helpful site with links to sources for such palettes and resources for checking choices you make is found at [8].
  - Palette used should be appropriate to the task – sequential palettes for intensity, yield, etc., and divergent palettes for the case of variables showing differences. We recommend the `viridis_r` colormap for sequential plots with small values in light colors and large values in dark colors..
  - Palette used **must** reproduce sufficiently well when printed in color (some color choices do well on screens but are insufficiently distinct when printed or shown with projectors). They also **must** be chosen such that the purpose of colors is still satisfied in grayscale. We recommend to try this out early enough by printing a page of your thesis!
  - Related to the above, functions plotted over data (or histograms shown without filling in color) need not only to be distinguishable from each other by color, if color is used, but by style – e.g. solid lines, dotted, dashed. There should be no case where distinction of functions plotted is not readily possible on a grayscale version of the plot.

---

[2]The Belle II style guide recommends to enclose units in square brackets. This is common practice in particle physics but violates the industry standard ISO 80000-1 and should be avoided.

- Colors **must** be identifiable by simple descriptions (e.g. red or blue) everyone knows. Do not use colors that require additional qualifiers (e.g. ultramarine or skyblue for two different shades of blue).

- Stacked histograms **must** only be made with the same solid fill style (and generally different fill colors) and should not, in general, include other fill styles. We recommend to separate stacked contributions with thin black lines and show the overall outline as a thicker black line.

- Overlapping distributions **must** have no fill style at all.

- We recommend that axis labels are aligned to the right (x) and top (y).

- When using MATPLOTLIB, you **must** set the plot range to show the full variable range (e.g. `[plt.xlim(1,2)]` to show exactly the range from 1 to 2). The default MATPLOTLIB behaviour is to add a bit of white at both ends of the axis which is indistinguishable from empty bins in histograms or missing datapoints.

- Figures **must not** have titles. If additional information is really needed, it should be text within the figure.

- Figures **must** have a legend that explains every line or symbol in the plot. An alternative is to label each line directly in the figure which is often used when showing exclusion limits but is uncommon in other contexts.

- If you are showing 2D histograms, you **must** show the z axis (usually a colormap) with proper label, and remember that bin widths **must** be indicated on axes where relevant. Be thoughtful about the color choice for bins that are below or above your chosen z range, and indicate in the plot if you show values that are out of range.

- Be consistent!

  - Use the same font, font-size, style (and everything that goes with it including bracket-type for units and the position of labels) for all plots.

  - If you decide to use words or parameters for your axis labels, always use the exact same words or parameters) for your axis labels.

  - If you decide to not follow a recommendation but use your own style, do that consistently.

Figure implementation:

- All figures **must** be referenced in the text and are part of the text at the position at which they are referenced.

- The figure should not appear on a page before the page on which it is being referred to. Use `[!h]` to focus placement in LaTeX, or use `\FloatBarrier` as last resort to force floating objects (like figures) to be placed.

- Figure captions **must** appear below the figure.

- The figure should be, together with the caption, stand-alone as much as possible.

- Most abbreviations in the figure should be "de-abbreviated" in the caption (exceptions exist: e.g. a sketch of a PC with "PC" written below it; but if in doubt play it safe and spell the abbreviation in full within the caption)

- Describe, as much as possible and reasonable, what is shown in the figure including the axis labels with as few words as possible (e.g. "Shown is the rate of events over recoil energy. The rate of neutrino induced events (blue dashed) is compared with the rate of cosmogenic background (red dotted) as well as the sum of all rates (black solid). The gray vertical line indicates the analysis threshold.").

- It is generally acceptable to show a plot from someone else or with data provided by someone else (e.g. from our working group or the collaboration). However, you **must** add the respective credits (e.g. "Courtesy of Jane Doe") and you **must** add a citation (see Sec. 2).

- If you include a list of figures using LaTeX (\listoffigures) provide in each caption an additional short version of the caption:
  \caption[Short version for LoF]{Long version to appear below figure}.
  LaTeX will use the short version for the list of figures.

8 Tables

- All tables **must** be referenced in the text and are part of the text at the position they are referenced.

- The table should not appear on a page before the page on which it is being referred to.

- Table captions **must** appear above the table.

- The table should be, together with the caption, stand-alone as much as possible.

- Most abbreviations in the table should be "de-abbreviated" in the caption.

9 Language, spelling and grammar

Several excellent books about scientific writing, language usage guides, and style guides exist. The following list does not replace a writing guide (e.g. [9]) but only provides recommendations on a few common topics.

- Favor the active voice over a passive voice and use subject-verb-object order whenever possible (example passive: "Theoretically, several narrow K^* resonances were predicted several years before the LHCb data." vs active: "Theorists predicted narrow K^* resonances years before LHCb published its observations.")

- You can use present tense or past tense to describe actions. Once you choose a tense for this purpose, use it consistently.

- Correct spelling is important in science to avoid distracting from the key message. Plan time for careful language-checks and use a good spell checker in your LaTeX editor.

- Be consistent!

  - If you decide to use a hyphen (e.g. in data-set) always use a hyphen (don't switch around between "data-set", "dataset", "data set" but use the one that some dictionary or trusted native speaker tells you is the most correct).

  - If you decide to capitalize words, always do so (e.g. don't write about Dark Matter in one chapter and about dark matter in another).

- Words we colloquially use when talking that should not be used in written, non-colloquial text

  - Cut(s): use "selection criterion" or "criteria" etc.

  - ERDM or NRDM (this is SuperCDMS specific): neither ERDM nor NRDM specify an actual Dark Matter candidate or interaction - use the proper term (e.g. instead of ERDM use "LDM" when talking about thermal Light Dark Matter that induces an electron recoil or "LDM-electron scattering" when talking about the interaction itself).

- Constants and parameters **must** always be italic (i.e. use math mode within normal text). This is also true for axis labels!

- Units and abbreviations **must** always be upright (i.e. use the mathrm environment within mathmode text - e.g. `$m_{\rm DM}$` results in $m_{\rm DM}$ where "m" being italic and "DM" being upright). This is also true for axis labels!

- Only introduce abbreviations once: the first time you use it. Do it as "... for which field-programmable gate arrays (FPGAs) are used...". You can use LaTeX package `acronym` to help with that.

- Do not use abbreviations in the abstract.

- In scientific writing, the prevailing style is to spell out numbers smaller than ten (e.g. "five detectors" instead of "5 detectors"). You can choose to spell out numbers up to 100, as long as you are consistent.

- In the thesis title and in all section and subsection headings, either use Title Case (capitalize all principal words, including nouns, verbs, adjectives, and adverbs) or use sentence case (capitalize only the first word and proper nouns), and apply the chosen style consistently throughout the document.

10 Structure

Your thesis is a scientific document and not a chronological report of all the things you worked on. It is important to realize that not all things are worth reporting, that the time spend on a certain topic during research is not always proportional to the number of pages in a thesis, and that the thesis should be written using the following structure in any research section[3]:

- start with a brief motivation describing why the following analysis (or measurement) is presented

- explain the procedure and tools used (e.g. data selection, data processing, ...)

- describe the observation(s),

- describe the consequences taken, if any,

- draw the conclusion(s).

You will very likely have taken detours during your work. They should not move focus from the main story-line of the thesis. Consider removing all material that is interesting but not needed to follow the main arguments, to the appendix. This potentially includes details of lengthy calculations, or sets of of very similar figures were only one figure should be shown and described in the main text.

11 Preface

Many analyses in (astro) particle physics are the result of large teams. We recommend to add a preface chapter that includes a statement indicating the student's contribution to the work presented in the thesis. This preface chapter should indicate if other people contributed significant parts of the work, but also highlight areas where the student contribution is significant. The following list shows a few examples of such work:

- Usage of data taken by a collaboration, or simulation produced by a collaboration.

- Usage of reconstruction software, or analysis software from the collaboration or other projects.

- Help with the electronics design, hardware design or prototype construction.

- Usage of systematic uncertainties provided centrally by the collaboration.

- Usage of event selection criteria developed by a colleague.

- Usage of generative AI (see Sec.13).

---

[3]Some sections like the introduction, the summary, or descriptions of the detectors will have slightly different format.

- ...

Note that this preface chapter does not replace the signed statement of good scientific practice required by most faculties in Germany. It also does not replace the need for correct citations in the thesis text or figures.

12 Additional Tips

- Add a "`\,`" between a number and its unit, e.g. `15\,cm` or `0.3\,GeV/cm$^3$`. This way they are tied to each other (no line break can separate them) and they only have a small space in between: $15\,\mathrm{cm}$ and $0.3\,\mathrm{GeV/cm^3}$.

- Always store backups of your thesis, we recommend that you use git for this, for example by creating a private git repository on github (https://github.com/) or by using the github servers provided by your institute (if any). If you use `Overleaf` the Git-Bridge and GitHub feature [10] can be used to facilitate synchronization between your `Overleaf` project and your git repository.

- It helps to define words or equations that you re-use very often first and just use this definition later: `\newcommand{\supercdms} {SuperCDMS\xspace}` or `\newcommand{\mb}{\ensuremath{\hat{M}_B^2}\xspace}`, where the `xspace` ensure a correct whitespace after the expression in both text and math mode. In your text you just use `\supercdms` which will show SuperCDMS, or `\mb` which will show $\hat{M}_B^2$.

- Using `\SI{<NUMBER>}{<UNITS>}` (from the `siunitx` package) can help getting the units consistent, and the spacing between numbers and units correctly (example: `\SI{2.5}{\giga\electronvolt}` will show $2.5\,\mathrm{GeV}$.)

- Avoid "can be" statements and present facts instead. Example:
  "The current signal can be converted into a voltage signal using a SQUID (...) ."
  Now does this mean a SQUID is used or not? To make it a clear statement about a fact write instead "The current signal is converted into a voltage signal using a SQUID (...) ."

13 Usage of generative artificial intelligence tools

Generative artificial intelligence (AI) refers to algorithms that generate content such as text, translations, images, or software code.

The use of AI for familiarization with new topics, literature research, or structuring content is permitted. The use of AI for grammatical or stylistic revision is permitted. However, the tools used must be named explicitly and the extent of their use must be described. This information should be provided in the preface chapter (see Sec.,11). Responsibility for the content, its accuracy, and the correct citation of primary sources lies solely with the author and cannot be delegated to AI systems. Text generated by AI systems often fails to meet scientific standards in terms of precision, depth, and source reliability and may therefore negatively affect the evaluation of the thesis.

In software intensive theses, the core scientific algorithms and analysis code must be designed and implemented by the author. Generative AI must not generate the central scientific components of the work. The author remains fully responsible for the correctness, validation, documentation, and legal compliance of all submitted code, including adherence to licensing and copyright requirements.

The university rules for good scientific practice apply without restriction. Unauthorized use of AI tools or failure to document their use constitutes an attempt to cheat. If supervisors, faculty or examination regulations impose stricter rules, those rules take precedence.

14 Experiment specific rules and recommendations

14.1 Belle II

- Measured invariant masses **must** be labelled with a capital M, model parameter masses (e.g. input values to an event-generator) **must** be labelled m. This rule does not apply to other variables.

- Follow the Belle II style-guide when producing figures [7].

- All figures that show data or simulation that is not officially approved **must** be labelled with "own work". Work that is approved but not published **must** be labelled "preliminary". You **must not** show "own work" plots in public (usage in a thesis and thesis defense is accepted, but not in a public talk).

- Experiment name, date, and integrated luminosity information should appear in the upper left corner, and the legend (if any) appear in the upper right corner of plots as standard practice. For posters, you should replace the text label for the experiment name with the logo.

References

[1] Track changes and comments in LaTeX. https://de.overleaf.com/track-changes-and-comments-in-latex. Accessed: 2021-10-29.

[2] Deutsche Forschungsgemeinschaft. Guidelines for Safeguarding Good Research Practice. Code of Conduct. https://doi.org/10.5281/zenodo.3923602, September 2019.

[3] Scientific Integrity at KIT. https://www.kit.edu/research/scientific-integrity.php. Accessed: 2021-11-25.

[4] BibTex. https://ctan.org/pkg/bibtex?lang=en. Accessed: 2021-10-29.

[5] Mendeley. https://www.mendeley.com/. Accessed: 2021-11-19.

[6] Zotero. https://www.zotero.org/. Accessed: 2021-11-19.

[7] Belle II styleguide. https://stash.desy.de/projects/B2D/repos/belle2style/browse. Accessed: 2021-11-15.

[8] Three tools to help you make colorblind-friendly graphics .
https://knightlab.northwestern.edu/2016/07/18/
three-tools-to-help-you-make-colorblind-friendly-graphics/. Accessed:
2021-11-15.

[9] Anne E. Greene. *Writing Science in Plain English*. University of Chicago Press, 2013.

[10] Using Git and GitHub.
https://www.overleaf.com/learn/how-to/Using_Git_and_GitHub. Accessed: 2021-11-17.