Common Gotchas in HEP Thesis Writing (And How to Avoid Them)

Andy Buckley, University of Glasgow

May 24, 2019

This document is a collection of common things that go wrong when writing and typesetting a PhD thesis, based on my 20 years experience of writing one and reading/examining many theses, and of writing umpteen LaTeX papers, macros, styles and classes. While the only way to really learn things is often to make the mistake yourself, I hope this short guide will help you to avoid some of the more foreseeable pitfalls, and produce a more readable thesis with less swearing at your computer. It's not intended to be read from cover-to-cover, but to be dipped into and gradually absorbed: I recommend first browsing all the item titles, dig deeper where something grabs your attention, and periodically revisit.

1 Content

Know your audience: The absolute number-one rule in *all* communication is to know who you are communicating to. This applies to presentations, papers, and even emails as much as it does to a PhD thesis. What do they need to know? What's most important, and what can be left out? Will they be looking for particular angles that you can anticipate and satisfy?

With a PhD thesis you have a dual audience: other researchers, and your PhD examiners. Of these, the examiner is the clear & present danger: they are guaranteed to read it all, in detail, and to specifically look for areas where they can grill you. In their defence, the grilling is not sadism – the job is to identify a) that you know what you're talking about, and b) that you actually did the work described. Make life easy by pro-actively giving them what they need to answer these questions.

This context means that you writing something a bit different from a physics paper; aping all the tropes of papers is hence not the best plan. This is a positive: often paper-writing, particularly in an experimental collaboration, is heavily constrained by conventions, journal rules, and internal review processes. By contrast, your thesis is yours and you get to say what you want, in the way you like to say it: enjoy it while you can! On the other hand, this means you *need* to actually think about what to say and how best to structure it.

In particular, you will have to say more than usual about the nitty-gritty of what *you* actually did: this includes stuff that might be "below threshold" for a paper, but there the *raison d'être* is to explain the science whereas here the examiner's prime directive is to understand *your* understanding and contributions. This means that you may wish to write more in the first person than is normal (i.e. you're allowed/encouraged to use "I", "my", etc.) and to write about the thought process rather than giving a passive, anonymised account of a polished final result.

Tell a story: Being non-fiction doesn't mean you don't need a narrative. Stories are how we make writing compelling. This is particularly important in long work like a thesis: don't dump plots, data samples, etc. into the pages without clear purpose: the reader will be wondering "where is this going?" and may even stop in exasperation. They should always know how they got to what they're currently reading, and how it connects to both the overall thesis and the aims of the current chapter.

Signpost: At the same time, this is a non-fiction technical document, and it's easier to get lost in those than in a novel. So make use of headings, intro sections, etc. to explicitly *tell* the reader where they are, and where they're going. Tie potentially confusing sections together with little "so that's where we've got to and why it matters, and it naturally takes us on to the next task…" paragraphs, to preserve the narrative momentum.

Think about the poor examiner! Believe it or not, reading 200-page technical documents isn't every academic's idea of a wild time. They *want* to see the great work that you did – and to push you a little to test that you really understand rather than just regurgitating stuff you heard along the way. Also, they are not necessarily a specialist in your area, and you'd be surprised how specific to an experiment or sub-topic some jargon can be: prefer to spell things out than to assume fluency in specialist jargon. If you make your thesis easy and compelling to read (remember: *know your audience*, and *tell a story*), they will thank you for it.

Lastly, it doesn't need to be 400+ pages! Length and quality are not the same thing. So don't put stuff in just because you can: put it in because it's important and demonstrates relevant command of the topic. Don't include anything that you can't defend or explain in an oral viva: this is asking for trouble. If it's *still* 400 pages…well done.

2 Style

> *"vi. Break any of these rules sooner than say anything outright barbarous."*
>
> — *George Orwell, in "Politics and the English Language"*
>
> `http://www.orwell.ru/library/essays/politics/english/e_polit` (read it)

Avoid cargo-cult academic writing: Much academic writing is painful stuff, driven by the need to show each other how very smart we are. Don't tell me it doesn't work this way! Actually, it takes a long time to build the level of self-confidence where you can write clearly again, free of the insecurities that encourage opaque tech-waffle and verbal gymnastics. See if you can tell it straight, without worrying that you're not using enough jargon: your readers (and examiner) are far more likely to gripe about the existence of impenetrable jargon and assumptions than about their absence. If you're not sure why a particular strange word is being used, question it and maybe find another: it is not uncommon in experiments to find a totally nonsensical word being used for things because of a mishearing or mistranslation far in the past.

There are some funny ideas around that "science writing is always in present tense", or "science writing has to be in passive voice": neither is true, and you'll do your readers a much better service by writing clearly than by trying to follow odd conventions in other theses and papers. Chances are that their oddities are not "because that's how it's done", but because their authors were also guessing, slavishly copying, and trying to cover insecurities. Break the cycle.

Consistency and attention to detail: There is a mix of stylistic and typographical advice in this guide – not because the typography really, really matters in some deep way, but because it's distracting for a reader to keep spotting presentational defects or inconsistencies. You'd rather they were thinking about your ideas! Good typography and writing is also important because sloppiness and lack of attention to detail are not qualities an examiner wants to find in a PhD candidate: set them up to expect high-quality physics by showing you can handle this easy stuff with words and symbols.

Clear, coherent English: The primary role of a thesis, or any academic writing, is to convey information. In your case the information is to explain what you did for the last few years, and hence show that you have the understanding and the research record required for a PhD. So make sure that your writing gets that across. Don't hide behind jargon, and if a concept took *you* a long time to get to grips with, then spend a bit of time on it in your writing, too. If need be, go back over bits of explanation text several times until it's honed: there are sure to be some.

You should also make sure that your style of English is consistent, e.g. that you don't flit between active and passive voice, or between tenses: these will feel like several people are writing, at different times. This is almost the case, but part of your task is to disguise that and make it flow.

Finally: are you using European or American English spelling? This most obviously shows up in the -ise or -ize endings of words like "renormalise/renormalize"[1]. For papers, this normally depends on the target journal – but honestly, when writing a paper let the journal's typesetting team deal with enforcing their irrelevant conventions. As an international field, HEP often gravitates to the American spelling, but for a thesis it's up to you: just be consistent. As you approach the end, or each time you send a chapter for review, run a text search or `grep` check for the "wrong" version: spell checkers won't flag up correctly-spelled inconsistencies.

Ubiquitous present tense? One of the great mysteries of HEP paper writing is the common use of the "historical present tense" – using present tense for everything, including things that definitely happened in the past: "the jets are being calibrated" or "the MC samples are generated". Readers can adjust to it, but why make them do so? Let's be clear: there is absolutely no rule that says "science is written in present tense", neither in general nor for journals. So why not write the natural and intuitive way?

This does raise the question of when to be in the present tense, and when to write in the past. You're writing a thesis, about your work of the last several years, so definitively most or all of the things that you did should be described with the past tense. But not everything in a thesis falls in that category: some statements are timelessly true forever, e.g. explanations of how a theory implies something, and presentations/discussions of your data (it's final, so observations about this particular plot will be true forever even if it is superseded by future experiments).

As ever, though, consistency is the key. Just make sure you're always clear about what tense you are writing in, and stick to it. It is quite disturbing as a reader to have events alternating between "are" and "were". My feeling is that this goes wrong because as a writer you are reflecting on past events and so your *natural* voice is the "were" and "have been" form. The way to minimise errors is to write in the way that is natural in your head, rather than have to keep contorting your thoughts to fit an unnatural style. Either way, do a search of your text to find and fix inconsistencies before submitting, either for informal review or as the finished article.

[1] Actually, -ize is valid everywhere, but -ise is certainly more common in the UK.

Section-title capitalisation: In section, chapter and probably whole-document titling, the field convention is to just capitalise the first word and any proper nouns. Capitalising Every Word is for high school, and our documents are already hard enough to read without more capital letters spoiling the readability.

"Respectively": A horrible HEP trope is to "parallelise" discussions about multiple quantities by use of parentheses, e.g. "10 (8) sensors in the r (ϕ) direction". The big problems with this are that it takes two passes to understand, and that it cannot be sensibly read aloud. Sometimes it's really the best option, but English does have a standard, elegant, and speakable version: "10 or 8 sensors in the r and ϕ directions respectively".

Painful (and hyphenated) compound adjectives: In speech, we often use nouns as adjectives, e.g. "the calorimeter calibration", "the track parameters". Unfortunately as our concepts become more complicated, we get compound nouns like "charged primary particle" and then twist them into an adjective role e.g. "charged-primary-particle p_T". This works fine in speech but horribly on the page. Notice that I also had to hyphenate here to avoid nominal ambiguities like the p_T being somehow charged: people very often get this wrong.

The best plan is to rephrase – making the written word depart from informal speech – to avoid such horrors, e.g. a flip of the sentence ordering like "the p_T of each charged primary particle" to avoid the monstrous compound. Yes, it's more work. Yes, it's worth the pain.

You can define acronyms, too, but please keep it to a minimum: it's easy to end up with text which is free of ambiguities and hyphenated compound-adjective chains, but is still hard to parse because the reader doesn't have enough intuition for your non-standard abbreviations.

Long sentences: Don't. It's easy to end up with very long sentences when both striving to eliminate ambiguity (and hence piling up chains of clarifying adjectives) and writing in an iterative fashion over many months. But it is not an unavoidable feature of science writing, and in fact you can make your work clearer and punchier by using short sentences most of the time.

When you proof-read, keep an eye open specifically for over-long sentences: if it's more than two lines long, it's probably bad. These can almost always be split, often by tricks like moving clauses between commas or parentheses to separate sentences, or by a flip of ordering as above. You can of course have too much of a good thing: a document entirely composed of very short sentences will also be painful. Variety is the spice of life, so feel free to add in some long-ish sentences for texture – provided that your meaning remains *clear*.

The English comma: In English, commas are primarily pauses for emphasis. The mental model for writing is hence "how would I read this out loud?". This is quite cool: by placement of punctuation, you are telling people when to (mentally) breathe! Note that this is not like German, where commas have a much stricter logical role. In English we also use commas to separate list items, or little "asides" within a sentence where they almost play the role of parentheses – but they always coincide with places that you want the reader to pause for emphasis or logical separation. Anyone who complains about commas (or their absence) in a non-structural role deserves a stern telling-off for focusing on the wrong thing.

Semicolons and other devices: Semicolons (;) also have a "pausing" role, but are a bit more structural. They are used to separate (by a big breath-pause) two parts of a sentence that are semi-disjoint but not enough for a full stop; they are also useful for delimiting list items which can themselves contain commas[2].

I am also quite fond of long dashes, either the "–" en-dash or the "—" em-dash, as a mechanism for structuring asides and clauses in sentences. To my mind they are more elegant and less disruptive to the reader than parentheses. I don't think it matters which you use in that role, but…be consistent.

All these punctuation tricks should be used sparingly: often it's better to rewrite, splitting the big clause-laden sentence into smaller, unencumbered ones. But they can be very handy, so it's a good idea to know how to use them: the more in your linguistic toolbox, the better you can express yourself.

Spell-check: It seems obvious. It is obvious! Run a spell-checker over your document every time you submit a copy to someone else. They want to check your ideas, which is much harder if misspellings (or typographical gaffes) keep drawing their attention away from the content.

Enjoy expressing yourself: Lastly, *enjoy* this challenge of grappling with language. As scientists, we have often spent our education focusing on hard concepts, mathematics, and lab skills – we didn't think about about language as a poetic, visceral thing in its own right, as well as a conduit for information. Once you start thinking about writing as something to *craft* and revisit until it zings, it becomes something thrilling and addictive. Your readers will enjoy the experience much more, too.

3 Presentation

I focus here on LaTeX as the main tool for writing theses and other long-form material in HEP. The real issues are the presentation effects themselves, not the niceties (or not-so-niceties) or the LaTeX language, but it's useful to be able to give concrete recipes to fix common issues whenever possible, rather than just waffle about typography. Being strategic about using features like macros and packages can also help to minimise inconsistencies, and allow easy reformatting in the late stages of your writing.

Avoid VERY LOUD ACRONYM disease: Particle physics is just *great* for jargon, isn't it? Total acronym soup: I still get regularly messed up on whether I'm in a context where e.g. "VR" means "variable radius" or "validation region". The best plan is to avoid them as much as possible, and actually say what you mean – particularly in a PhD thesis, where your target audience is not part of your research in-group.

But sometimes they are unavoidable, avoidance would be very contrived, or repeating cumbersome phrases would be more injurious to the flow of your argument than deploying the acronyms. Now you risk the horrors of a page stuffed so full of ALL-CAPS words that our hard-learned typographic cues for rapid reading stop working: we are good at parsing letters of lower-case size, so a lot of capital-size lettering turns a nicely written document into a hideous slog. In this situation, judicious use of small-caps can work wonders – or, often, SmallCaps, e.g. for event generators like

[2] See what I did there?

Pythia. (There is little hope for the reader forced to contend with MadGraph5_aMC@NLO, but you can do your best.) If you get very picky, you might like to operate different schemes for acronyms (*pronounceable* first-letter sets like "Unesco") and initialisms (the unpronounceable sort, like "UNHCR"). This is the way taken by The Economist's and The Guardian's style guides. It's up to you.

Non-breaking spaces: Learn to love the non-breaking space `~` (and/or its relatives like `\,`) – use it in units; in references to sections, tables, etc.; in use of `\cite`, after "e.g.", "i.e.", "cf."[3], and so-on. Basically, non-breaking spaces should be used anywhere that you don't want a symbol awkwardly starting the next line, e.g. `compare~\& contrast` (but anyway only use "&" where there's good reason, such as when you need an "and" in a list item).

Line-breaking in displayed titles: Choose your break placement for emphasis/grouping, and don't break author names across two lines. The latter is a good reason to enter author names with non-breaking spaces, e.g. `Andy~Buckley` so the break never splits name parts.

Using Figure, Chapter, Section, etc.: When referring to some other part of the document, always specify the type of thing you're referring to, and *capitalise it*: it is a proper noun in this context. For example, `in Section~\ref{sec:nittygritty}` or `see Tables~\ref{tab:a} and~\ref{tab:b}`. (NB. I find these sorts of `sec:`, `tab:`, etc. `\label` prefixes very helpful.) The exact word used doesn't matter, but do keep it consistent, e.g. always "Section" and "Table", or always "Sec." and "Tab.". I prefer the full form, unless a journal style insists otherwise (some even force swapping of long and short forms) but it's your thesis and your choice. Just be consistent.

Explicit use of "Ref.": Usually citations just go inline in sentences, with the idea being that you don't pronounce them at all when reading. Indeed, in many disciplines little superscript markers are used to reference them, rather than the bigger "[42]" sort of citation marker used in HEP, specifically to be less distracting. But if you are referring directly to a paper, then the citation becomes something to be pronounced. If you were to write "as demonstrated in [42]", then the reader will read it internally "as demonstrated in 42", which can be understood from context but isn't the whole story. You wouldn't refer to another bit of the document as "see 3.1" – is that a section, a table, or what? The same goes for references: when actively talking about a paper (which should be infrequent), say e.g. `as demonstrated in Ref.~\cite{xyz}` → "…in Ref. [42]". Or if you're really feeling fancy, use `\citenum` instead of `\cite` to get "…in Ref. 42".

Spell out small counting numbers: It's jarring to read "there are 2 possibilities", or "these scale-factors are compatible with 1". The reader's brain has to jump from "text mode" to "number mode" and back. For small numbers like these, spell out the words "two" and "one" to keep their train of thought on the rails.

Equations are sentence elements: Equations aren't separate from the running text, even when they are in "displayed" form. The surrounding text runs into the equation, and should interact with it as if it's a word or phrase, with appropriate commas, full-stops, etc.. Sometimes a trailing comma, full-stop or semicolon should be put in the equation itself: in these circumstances it's often a good

[3] Note that the first two of these abbreviations have two dots, but the third only has one. Sorry.

idea to sematically separate it from the mathematical symbols with a bit of space, e.g. `I = \int \mathrm{d}x \, \sin^n\!{x} \, \cos^m\!{x} \; .`[4] to give

$$I = \int \mathrm{d}x \, \sin^n\!{x} \, \cos^m\!{x} \; . \tag{1}$$

Space around displayed equations: Body text and displayed equations that connect to that text are part of a single paragraph: you hence shouldn't leave blank lines in your LaTeX source between the text bits and the equation environment. If you do have blank lines (which makes the source more readable), then there will be excessive vertical blank space before and after the equation, and an unwanted paragraph indent when the paragraph text resumes. You could use `\noindent` to suppress the latter, but fundamentally the answer is to avoid LaTeX thinking there's a paragraph break. Personally I like to use lines with just a `%` comment marker between text and displayed equations – they make the code readable without breaking the paragraph.

Multiple citations: When referencing multiple bibliographic sources at the same time, you want them to be grouped together in the same citation bracket, not a series of them one after the other. For this, use `\cite{refA,refB,...}` as opposed to `\cite{refA}\cite{refB},...`.

Use macros for repeating symbols: If you have symbols that get used many times (you do), then don't type it out every time: use `\newcommand` (or similar) to define a new LaTeX macro for your quantity and use that instead. It's always a good idea to end your macro definition with `\xspace` (from the xspace package), so you don't need to explicitly stop your macro from "eating" any white space that follows it. Now if you later spot that there's an ambiguity between two symbols, or an annoying typographic issue, it's easy to change it globally rather than resorting to search & replace.

Don't override hyperref defaults for printing: By default the hyperref package will render links in PDF documents as (usually?) red rectangles around the hyperlinked word, indicating the live area that can be clicked on to follow the link. These boxes are, admittedly, pretty darn ugly. But they have a superpower: they don't print! By contrast, if you override the hyperref defaults, you can make the hyperlink text itself coloured like a Web link, and get rid of the ugly box…which is great when using the PDF via a screen-reader, but when you print your dead-tree thesis, all the references to sections, tables, etc. will try to print in colour. Not what you want. So this is not an absolute rule: for sharing the PDF of your thesis or paper, do whatever you like; but when printing a bound copy for yourself, stick with the hyperref defaults.

3.1 Units

Avoid italic units: Physical units are text, not maths. They often have to be used in a math context, though, which is why you should use macros for your common units (and maybe the siunitx package, with some tweaking of its default settings). Inside your macro you can use the `\text` macro to ensure that you are always writing the unit in text mode, not math mode, even if the macro is called from math mode. Units set this way will also go bold and italic to match the context, which is what you want.

[4]And lots more semantic space tweaking here!

Units and non-breaking spaces: Always tie the number to the unit with a non-breaking space: you don't want a line-break between the two, with a lonely number on the end of the line, and a lonely unit at the start of the next. The number and the unit always have to go together, just like in references to sections, tables, references, etc. The siunitx package can help you here, by specifying all unitful quantities like `\SI{12.5}{\GeV}`, which typesets as 12.5 GeV – it also allows you to globally tweak how wide the space between number and unit should be, if you like that sort of thing.

Use an upright mu for "micro" units: Use siunitx and type e.g. `\si{\micro\metre}`, rendered as "µm". Not `$\mu$m` or whatever other abomination you're using to make "μm". Ick.

3.2 Numbers and mathematics:

Ranges use an en-dash: Not a minus sign, which will add operator spacing (and use the math font): `$5-10$` looks like "$5-10$". Instead write `5--10`, which renders as "5–10". Or write "5 to 10"! siunitx's `\numrange` and `\SIrange` can also do this (and more) for you, if you like that sort of thing.

Angle brackets: Unlike round, square and curly brackets/parentheses, angle brackets, e.g. in expectation values and matrix elements are *not* bounded by their obvious keyboard-character equivalents, `<` and `>` – because those are already taken for the math-mode less-than and greater-than symbols. You don't want these as brackets: they're far too angular, add extra operator spacing, and generally look *rubbish*. $< 0|\mathcal{M}|0 >$. Instead you need more elegant, obtusely angled brackets obtained with the `\langle` and `\rangle` macros: $\langle 0|\mathcal{M}|0\rangle$. These will also respond to the automatic math-bracket sizing commands, `\left` and `\right`.

Math operator words: `\sin`, which produces "sin", not `sin` which renders as "sin". The same for `\cos`, `\log`, and many others: check the Short Math Guide. If the one you need is missing, define it easily (and with the correct spacing) via amsmath's `\DeclareMathOperator`.

d is also an operator: On the topic of operators, the total differential operator d is also – as the name suggests – an operator and not a variable. As such, it should be in upright font via the `\mathrm{...}` macro, i.e. `$\mathrm{d}f/\mathrm{d}x$` $\rightarrow \mathrm{d}f/\mathrm{d}x$, not `$df/dx$` $\rightarrow df/dx$. Although it would be bad form to do so, and you'd be better off re-thinking your variable names, the differential of a variable d makes this clear: $\mathrm{d}d$, not dd. Writing out the `\mathrm` all the time can get painful, so you might want to define some macros, e.g.

```
\newcommand{\dbyd}[2]{\ensuremath{\mathrm{d}{#1}/\mathrm{d}{#2}}\xspace}
```

to make life easier; similar things, with enhancements like optional "powers" for nth-order derivatives and partial derivatives, exist in various LaTeX packages such as physics.

Spacing in math mode: LaTeX is very good at symbol placement in equations, but not always perfect: it cannot anticipate all the semantic meaning of what you are writing. In some circumstances you can improve readability by pushing some symbols apart a little, e.g. a " " between an `\int` or `\mathrm{d}x` and the following integrand, or between distinct groups of variables and operators. In other places you might want to pull symbols a little closer for the same reason, using `\!`. For fine-tuning in extreme circumstances you can use `\mspace` or `\kern`.

Also, be aware that spacing in mathematical expressions actually means something: don't drop in and out of math mode just to make symbols italic but forget about the symbols: the spacing and symbols for operators like "+" and "=" are important. If I write `$x$ + $y$ = $z$`, I get "x + y = z", whereas I really wanted the elegant, proper spacing and symbols of `$x + y = z$`, which gives $x + y = z$. That one's contrived, but it's very common to see something like `$\sqrt{s}$ = 13~\TeV`, which does *not* render as you really want it to. So put the *whole* of each mathematical expression in math mode (and drop out of it via `\mathrm` or `\text` if needed). If this gets annoying and repetitive, maybe make a macro for it: there is great power in laziness.

Negative numbers have to be in math mode to get a proper minus sign: In text, if I refer to `-1`, I'll get "-1". That's not right: I wanted a minus sign, not a short hyphen. But is LaTeX to know what I wanted. I need to go into math mode to tell it what I want: `$-1$` which gives "-1". Note that it changed the "1" symbol font as well, though: I'm not a big fan of this. The `\num` macro and similar from the amazing siunitx package will do it nicely (and many other handy things, especially for numbers direct from code output), but this may be too much overhead. It's your call: this is more fine-tuning than essential.

Bold math in headings: An annoying LaTeXism is that math-mode symbols don't adapt to the surrounding font context. The most important of these is that math symbols don't become bold, even when all the text that they are embedded in *is*. Technically this is because the text and math font modes in LaTeX's "New Font Selection Scheme" (NFSS) are handled completely separately (e.g. the boldening commands are `\textbf` and `\mathbf`, and others follow the same nomenclature), but it's still annoying. The place where this is most awkward is in section titles, as declared by e.g. `\section{Calculation of $\alpha$, $\beta$, and $\gamma$}`. The title is usually rendered bold by LaTeX, but math symbols like the Greek letters in the example above will appear unboldened, which looks rubbish. To make matters worse, if you try to force the boldening by use of `\mathbf` or `\bm` around the math stuff, it will then appear as the *only* bold symbols in the table of contents! You can't win…

…except that you can, by putting the following in your document preamble[5]:

```
\makeatletter
\g@addto@macro\bfseries\boldmath
\makeatother
```

This little gem modifies the `\bfseries` macro used to change the font variant to the bold "series" in the NFSS to also activate the bold-math mode. Math will now adapt its boldness to match the context. Hallelujah!

Upright and text-spaced words in equations: This rule applies most commonly to subscripts and superscripts on physics symbols, but applies generally to any "text" in equations. You may have the impression that LaTeX's math mode just puts all the text into italic script, but this isn't true: instead, it treats every character as a variable, so that a "word" in math mode like `$final$` is really just the product of the 5 symbols f, i, n, a, and l. There is often little distinction, but – particularly when dealing with italic letters – normal text applies "kern corrections" which change the spacing between adjacent letters to accommodate complementarities in their shapes: `VA` gives "VA" and `\textit{VA}` gives "*VA*", for example. In math mode, these corrections are switched off:

[5] See https://tex.stackexchange.com/a/124311/7649.

`$VA$` gives "VA"! You won't encounter that one very often, but x_{final} and p_{MET} are not uncommon and look rubbish: x_{final} and p_{MET} are much smarter (and note that the first also got the subscript properly shifted nearer to the x). The general rule, that you won't go far wrong with, is that "text" like acronyms and words as symbol sub- and super-scripts should be in upright "Roman" font anyway: it's not a variable but a fixed label, so it shouldn't be italic. So always wrap text in your equations in the `\mathrm` macro, which does it right[6].

Avoid inline tall fractions: LaTeX's `\frac` command makes you feel cool and like you're doing proper mathematical typesetting. It also immediately produces structures larger than the normal line depth, forcing the renderer to make the current line deeper to fit it in: this adds ugly bars of whitespace, disrupts the text, and makes reading harder. It's easier, and looks nicer, to just write a/b in inline equations. Sometimes the easy way is the right way: save `\frac` for displayed equations.

In some places, superscripts and subscripts may also induce painful line-spacing. If really unavoidable, the `\smash` macro can avoid the typographical evils by hiding the effective height from TeX's layout engine.

3.3 Figures and tables

Float positioning: Prefer the default float-spec, `[tbp]`, i.e. top-of-page, bottom-of-page, and separate-floats-page in decreasing order of preference. Please god don't use the "here" `[h]` or nuclear-option `[H!]` specifiers everywhere: those are for *very* extreme circumstances. People tend to gravitate to those, because they put your figure or table *right here*, like MS Word would have done. Word does that because it's crap and knows nothing about typesetting: the whole point of floating figures is that they find places to go which preserve the flow of text and aid the reader. When you thwap down a figure "right here, please", it usually breaks the flow and makes the reader's life harder – especially when the caption text is confusable with the restarting of the body text. Don't do that. `[tbp]` is a great default: override it only when you really need to.

Another common float positioning issue is keeping the floats close to the relevant text: LaTeX does its best here, but its best is not always great. In particular, it will locate floats according to where the `figure` or `table` environment is placed in the source code, rather than optimising relative to where it's `\ref`'d. Sometimes you have to place that float code in strange places to get the right positioning in the result; sometimes you need to force emptying of the float queue. The latter is particularly common if you have a *lot* of floats in a row: LaTeX can overflow its teeny 1980s memory allocation before it has time to start placing them. My favourite trick for this uses the afterpage package, which will automatically run a command after the current page has finished rendering: calling `\afterpage{\clearpage}` uses this mechanism to force clearing of the LaTeX float queue without leaving you with a half-empty page before the floats. However, *don't do this sort of fine-tuning until the very end of writing*! It is a fool's errand to spend time fiddling detailed float positioning when you may yet add more text to earlier sections that will anyway force a re-flow.

Table formatting: Use the booktabs package. Always. It's very simple: it gives you three macros, `\toprule`, `\midrule` and `\bottomrule`, which you should use for the obviously positioned hor-

[6] There is also a very useful `\text` macro (and its complement, `\ensuremath`) in the amsmath package, but this is better suited to running text than labels on symbols: see https://tex.stackexchange.com/questions/19502/is-there-a-preference-of-when-to-use-text-and-mathrm.

izontal rules in your tables: they have been defined to get the spacing and line-weighting right. There's also an `\addlinespace` that's occasionally useful. Also read booktabs' excellent manual on how table formatting should look: in short, never use vertical rules.

Alignment in tables: A cheap and reasonable rule-of-thumb is to left-align most text columns in tables, and to right-align most numerical columns. Centre-alignment looks tacky and novice: there's a reason that LaTeX doesn't centre-align any titles other than on your thesis cover page! The right-alignment of numerical columns is an easy way to align decimal places. If you have different numbers of decimal places in different rows then you'll need to work harder, e.g. with the occasional `~` or `\,` spacer. If, horror of horrors, you end up considering custom column specifiers with the decimal points built-in, and each number split into two columns, then look into the siunitx package's functionality instead: it work wonders for fancy numerical alignments, without too much pain.

3.4 Particle physics specifics

Capitalise "Standard Model": It's a proper noun, so it gets capitalised. We're not just talking about some individual's personal standard method of modelling, but the accepted HEP name for a single, unambiguous concept.

Particles in italic vs. upright: Italic is more easily consistent, except for Greek capitals, since you'll have to use math mode anyway for charge and flavour sub- and super-scripts. It's just a choice, and consistency is king, though: for example, ATLAS and LHCb prefer italics but CMS likes upright, and I wrote my thesis in upright but now prefer italics. If you like, I wrote (as thesis procrastination, many years back) the hepparticles and hepnames packages to do it all for you[7].

Heavy-quark physics objects: We're pretty clear that bottom quarks are referred to as b-quarks, but what about the hadrons they get bound into: b-hadrons or B-hadrons? B-jets or b-jets? And b-tagging or B-tagging? The answer in all these cases is the lower-case b, and for two reasons.

First, we are almost always talking about associations with the bottom quark, which apply equally to baryons like Λ_b as to the more common B-mesons. Indeed B-mesons is the one time that the upper-case B is the right way.

If that doesn't convince, the other reason is by comparison with the charm-quark equivalents: here the baryon argument also applies, but the mesons are not so convenient as to be named C: mesons containing charm quarks are D-mesons! No-one talks about D-jets or D-tagging, so let's stick with c- and b- except when speaking specifically about the mesons.

e-V kerning: [8] Several times in this document I've mentioned kerning, the tweaking of horizontal space between characters to make them fit and flow better. This is a total geek-out topic for typography enthusiasts[9], but makes normal human beings want to claw their eyes out. Fear not, you'll learn to love it. But there's one kern you should be aware of: the one that appears in the electronvolt symbol, eV. LaTeX's fonts know a lot of standard kerns for symbols which naturally fit

[7] And also hepthesis, which might be a good starting point.
[8] Written "`e-\!V kerning`"...
[9] Guilty as charged.

together closer than their rectangular bounding boxes would suggest – "AVA", for example – but sadly no-one thought to tell the standard fonts about the ability to fit a lower-case "e" under the gull-wings of a capital "V": without tweaking, you'll get "eV", which isn't great in e.g. GeV. The default LaTeX font, Computer Modern, is worse still. I don't know of a way to automate this, or manually add standard kerns – this is built into the font, rather than under LaTeX control – so you'll need to define macros for eV and its derivatives MeV, GeV, etc. Here's mine, for this font:

```
\newcommand{\eV}{\text{e\kern-0.15ex V}\xspace}
\newcommand{\MeV}{\text{M\eV}\xspace}
\newcommand{\GeV}{\text{G\eV}\xspace}
\newcommand{\TeV}{\text{T\kern-0.1ex \eV}\xspace}
```

You can (and should) use these everywhere, including in math mode: $\sqrt{s} = 7$ TeV.

Cross section or cross-section? Oh dear, here we go. Should there be a hyphen between the pair of words we use for σ? Both ways are valid, and once again the only important rule is consistency: pick a style and use it everywhere. But personally, I'm a σ-hyphenist, and here's why.

It's about making life easy for yourself: cross-sections in HEP are so ubiquitous that we use them in more complex contexts than you would for e.g. a cross-sectional cutaway diagram. In particular, let's say that you've decided to go with unhyphenated "cross section", and then you talk about "the cross section normalisation": at least to me, this feels like a compound adjective where we'd need to add a hyphen anyway ("cross-section normalisation") to avoid inflicting a parsing difficulty on the reader. In unhyphenated land, you need to be aware to this rather subtle issue of English grammar, while if you always hyphenate your σ-word then you can slap it down everywhere without fear.

CKM matrix element subscript heights: It's harder than you think, especially in unitarity relations. The issue is that subscripts and superscripts don't always appear at the same height: the subscript depth in `$V_{cb}$` $\to V_{cb}$ is different from that in `$V^*_{cb}$` $\to V^*_{cb}$ because the superscript complex-conjugation asterisk in the second case "pushes" the subscript a bit further away to make the symbol look nice. Plain TeX has a mechanism to control all this, but it's hideously complicated. There are two options:

1. this is only really noticeable in unitarity relations where the symbols with different script heights are side-by-side, so just typeset explicitly with hidden asterisks, e.g. terms like `$V^{\phantom{*}}_{ud} V^*_{ub}$` $\to V_{ud}V^*_{ub}$;
2. or just force all subscripts to be the same height, regardless of superscripts, by using the subdepth package: just add `\usepackage{subdepth}` in your preamble.

Neutrino masses in the SM? It seems that some examiners get very exercised by statements – I suspect passed down through generations of thesis copy – that neutrinos have no mass in the Standard Model. This is a bit absurd: we've known that they have masses for around 20 years. What's missing is certainty over whether they should have Dirac or Majorana mass terms in the SM Lagrangian – a slightly different thing. Phrase it carefully, as you please: just don't make it sound like we're all denying the existence of neutrino masses.

Thanks to all those who have provided feedback and extra submissions to the collection of examiners' bugbears: Frank Krauss, Jon Butterworth, Tim Gershon, and Will Barter. More suggestions (send to `andy.buckley@cern.ch`*, please) are always welcome, but inclusion is at my discretion: keeping the issues relatively large and unambiguous, and hence this document relatively short, is key to its usefulness.*