Requirements for typesetting physics

Ulrik Vieth

Abstract

This paper attempts to summarize the requirements for typesetting math-mode material in physics and related sciences in accordance with the established typesetting rules applicable in those fields. Particular emphasis is placed on pointing out the subtle differences from (LA)TEX's default settings for math-mode, which are primarily geared towards the requirements for mathematics and computer science.

This is an unfinished draft of a discussion paper which was begun back in May 1997. Unfortunately, as of October 27, 1997, a number of sections still remain very sketchy and incomplete and need to be fleshed out eventually.

1 Introduction

Besides mathematics and computer science, physics is probably one of the most important areas in which (LA)TEX is heavily used for typesetting math-mode material (as opposed to applications in the humanities for typesetting body text).

While the default math-mode typesetting rules implemented in (LA)TEX are primarily geared towards the conventions applicable in North American math typesetting, as specified in authorative handbooks such as *The Chicago Manual of Style* [1] or the style guides of well respected math publishers [2], some slightly (or not so slightly) different conventions are applicable in physics and related disciplines, such as physical chemistry or engineering sciences.

This discussion paper attempts to summarize the requirements for typesetting physics, following the recommendations of IUPAP, the International Union for Pure and Applied Physics [3–7], and it also discusses various alternative conventions commonly found in physics journals or textbooks produced by publishers [8–14] who adhere to these guidelines to a greater or lesser degree.

An earlier review of the requirements for typesetting physics was presented by Jörg Knappen at the EuroTEX '92 conference [15]. The specific requirements in high-energy physics with particular emphasis on elementary particle notation were summarized in a *TUGboat* article by Michel Goossens [16] and also in Chapter 6.1 of *The LATEX Graphics Companion* [17]. Another very extensive review of the requirements for typesetting mathematics in science and technology including references to the relevant ISO standards appeared in a recent *TUGboat* article by Claudio Beccari [18].

2 Requirements for math alphabets

2.1 Symbols for physical quantities

To quote directly from the IUPAP recommendations [3, sec. 1.2.1–2], the general rules say:

> 1. Symbols for physical quantities should be single letters of the Latin or Greek alphabets with or without modifying signs.

and

> 2. Symbols for physical quantities should be printed in italic (or sloping) type.

Since this conforms exactly to the default font shape in TEX's math mode (accessible by `\mathnormal`), no special markup is needed, and there is little to worry about in this case, except for the fact that there is some disagreement concerning the preferred shape of uppercase Greek letters.

While the examples in the IUPAP document seem to indicate that uppercase Greek should be set in math italics type as well, just like anything else, this convention isn't always observed in publications, and one frequently finds the use of upright (roman) type for uppercase Greek letters instead, whether it be for laziness, lack of suitable fonts, or a policy of following a traditional house style.

To achieve the greatest possible flexibility, both conventions need to be supported at the font encoding level, and a switching mechanism at the macro level should be provided, if possible.

2.2 Subscripts to physical quantities

As a guiding principle for the printing of indices, it is recommended in [3, sec. 1.2.2] that:

> Only indices which are symbols for physical quantities should be printed in italic (sloping) type.

This implies that normal (upright) type should be used otherwise, especially for textual indices.

Thus, for example, in the energy confinement time τ_E the subscript 'E' should be printed in math italics type because it refers to the physical quantity "energy", whereas in the particle confinement time τ_p the subscript 'p' should be printed in roman (upright) type because it stands as an abbreviation for a textual expression "particle".

Some publishers' style guides [8–10] try to simplify these guidelines by suggesting to use upright type for multi-letter subscripts (which almost always represent textual material) while using math italics for single-letter subscripts (which may or may not represent a physical quantity). It is probably not a good idea to follow such a simplification since it may be misleading at times.

Concerning the font used to denote textual subscripts, care should be taken to ensure that this should always be upright (roman) type regardless of the surrounding text font, since the accidental use of text italics or slanted type (e.g. within a theorem environment) might only cause confusion. Thus, a `\text` macro like in AMS-LaTeX which has the effect of selecting the surrounding text font rather than the default roman font may only be suitable for the mark-up of intervening words between formulas, but not for textual subscripts within formulas.

Using `\mathrm` for the latter purpose might be a better alternative, provided that the roman letters are taken from a font with normal text spacing. What would happen if math roman letters were taken from a hypothetical MC-encoded roman font with math spacing hasn't been studied so far, but it is likely to assume that some incompatibilities might arise.

2.3 Symbols for vectors and tensors

Concerning the representation of vectors and tensors the the IUPAP recommendations [3, sec. 1.2.3] suggest:

> To avoid the excessive usage of subscripts, it is often convenient to indicate vectors and tensors of the second rank by letters of a special type.

which in LaTeX lingo means using special math alphabets loaded into extra math groups. In particular, it is recommended that

(a) vectors should be printed in bold italic (sloping) type, e.g. $\boldsymbol{A}$, $\boldsymbol{a}$,

(b) tensors of the second rank should be printed in bold sans serif italic (sloping) type, e.g. $\boldsymbol{\mathsf{S}}$, $\boldsymbol{\mathsf{T}}$.

Although the recommended choice of fonts is fairly specific, the conventions used in practice often vary, and one frequently finds the use of bold upright type (`\mathbf`) for vectors, e.g. $\mathbf{A}$, $\mathbf{a}$, and bold sans serif upright type for tensors, e.g. $\mathsf{\mathbf{S}}$, $\mathsf{\mathbf{T}}$.

Still other publishers [11, 12] resort to replacement notations in the preprint styles, such as bold upright type combined with double underlining for tensors, e.g. $\underline{\underline{\mathbf{A}}}$, while using the proper typefaces in the final publications.

To some extent, all these deviations from the standard may, in the past, have been dictated by the lack of suitable fonts or the difficulties of accessing such unusual font shapes. However, if such technical constraints no longer persist, it is probably best to try to avoid any such deviations in the future and use the proper notations wherever possible.

Since, in general, a full set of Latin and Greek letters may be needed to denote vectors and tensors, a number of additional MC-encoded fonts are needed to provide suitable math alphabets, whereas using T1-encoded fonts will be insufficient for this purpose. To allow catering for all commonly used conventions, at least four additional MC-encoded fonts in bold upright, bold math italics, bold sans serif upright, and bold sans serif oblique would be needed. However, it may be sufficient to provide only a subset of the MC-encoding restricted to Greek and Latin letters or letter-like symbols, while omitting non-changeable symbols that are included in the basic MC-encoded font only for technical reasons.

Finally, if no suitable math alphabets are available, or if their use is not desirable for whatever reason, it is suggested in [3, sec. 1.2.3] that:

> A vector may be indicated by an arrow and a tensor by a double arrow above the symbol, e.g. $\vec{A}$, $\vec{\vec{S}}$.

Whereas the use of a single arrow for vectors is well-established and corresponds to the default `\vec` macro in TeX's math mode, there is a somewhat wider variety of conventions used for tensors. While the IUPAP document suggests a double arrow consisting of two arrows stacked on top of each other, the REVTeX package [8] provides a `\tensor` macro that uses a left-right over-arrow like this: $\overleftrightarrow{S}$.

In any case, establishing a consistent mark-up syntax and providing a flexible switching mechanism at the macro level to allow selecting the preferred representation of vectors and tensors would be very desirable. In particular, this would be a significant advantage with regard to document portability and reusability despite the absence of previously established standard macros to access extra math alphabets. If so, adopting already existing input notations, such as `\vec` for vectors and `\tens` for tensors, to be used regardless of their eventual representation might be the most preferable choice.

2.4 Symbols for physical units

Acoording to the IUPAP document [3, sec. 2.1]:

> Symbols for units of physical quantities should be printed in roman (upright) type.

In general, physical units are represented by single or multiple Latin letters, usually in lowercase only except for units derived from proper names which are in uppercase or mixed-case.

There are only a very small number of exceptions from this rule that require the use of Greek letters, specifically the multiplier micro ($\mathrm{\mu}$), the unit Ohm (Ω) and the inverted Ohm ($\mho$), all of which are available from a `TS1`-encoded text companion font. Finally, there is the non-SI unit Ångstrøm (Å) consisting of an accented letter which could always be accessed by resorting to `\textrm` within math mode if nothing else is provided.

We can conclude that, in general, a `T1`-encoded text font accessible with `\mathrm` or a synonym such as `\unit` would be sufficient to typeset physical units if provisions are taken to handle the exceptional cases by special macros.

Knappen [15] suggested a `\Ohm` macro for Ω and similarly one might introduce a `\micro` macro for µ. Although, some similar macros are already provided in the `textcomp` package as `\textohm` and `\textmu`, such macros should also be included in a physics-specific math-mode package to make it unnecessary to rely on other LaTeX packages.

As for the typographical treatment of physical units, it should be noted that it is frequently necessary to insert a little space between numbers and units or in between physical units where multiplication is implied. While *The TeXbook* suggests using a thinspace, some other sources even suggest a thickspace or an unbreakable word space. Whether such spacing should be inserted automatically by the macros used to mark up physical units seems highly questionable, as it is often highly dependent on the context whether or not extra spacing is needed.

As for the formatting of numbers, nearly all publishers' style guides remind the author to avoid dangling decimal points, i.e. to write out 1.0 or 0.1 instead of 1. or .1 to avoid confusion.

2.5 Symbols for chemical elements, nuclides, and particles

- to be filled in, for details see [3, sec. 4]
- symbols for chemical elements and elementary particles should be set in upright roman (`\mathrm`)
- requires access to a full set of upper- and lowercase Latin and Greek, i.e. a `MC`-encoded font
- suggested mark-up (`\chem`, `\particle`)

2.6 Quantum states and spectroscopic notations

- to be filled in, for details see [3, sec. 5]
- symbols for quantum states should be set in upright roman (`\mathrm`)
- requires access to a full set of upper- and lowercase Latin and Greek, i.e. a `MC`-encoded font

2.7 Dimensionless numerical parameters

- to be filled in, for details see [3, sec. 7.14]
- two-character mixed-cased expressions like 'Re' for the Reynolds number, 'Ma' for the Mach number, etc.
- to be set in italic (sloping) type, preferably in text italics (`\mathit`) for better kerning in two-character expressions

2.8 Other math alphabets

Script/Calligraphic letters. These are used only in a few specific contexts, typically in cases where two different physical quantities would normally be represented by the same letter. Examples include:

- '$\mathcal{E}$' when used for energy in contexts where 'E' is already used to represent the electric field.
- '$\mathcal{L}$' and '$\mathcal{H}$' when used to denote the Lagrangian or Hamiltonian density in field theory, where 'L' and 'H' represent the Lagrangian or Hamiltonian function.
- '$\mathcal{O}$' when used to denote the order of neglected terms in series expansions. Some authors (including DEK himself) prefer to use a regular math italics 'O' for this purpose, so the use of a caligraphic '$\mathcal{O}$' might be regarded as controversial.

There may be some disagreement among true believers as to whether TeX's default caligraphics letters are sufficiently scripty for this purpose or whether a different script font, such as `rsfs`, should instead be substituted for it.

There is probably no need to introduce any special mark-up tags for these relatively few cases of script or calicgraphic letters used in physics. However, it may, of course, be very convenient to define private shorthands, such as '`\L`' or '`\H`', when dealing with material where these symbols are frequently used.

Blackboard Bold letters. These are sometimes used to denote number sets, such as '$\mathbb{N}$', '$\mathbb{R}$' or '$\mathbb{C}$' exactly as they're normally used in mathematics. Again, there may be some disagreement among true believers as too their prefered shape. Also, it may again seem convenient to define private shorhands, such as '`\N`' or '`\R`', if these symbols are used rather frequently.

Fraktur letters. These are no longer used in modern physics textbooks, but they are sometimes found in reprints of historic texts (e.g. Einstein's lectures of 1922) where Fraktur was used to denote vectors,

3 Requirements for mathematical symbols

3.1 Mathematical constants

Certain mathematical constants should always be printed in upright (roman) type, especially when there is a danger that they might be confused with a similar symbol denoting a physical quantity.

In particular, this convention applies to the following mathematical constants:

- 'π' in its usual meaning ($\pi = 3.14159\ldots$). This convention is rarely observed at present, presumably for lack of suitable fonts.
- 'e' when used as the base of the exponential function or natural logarithm ($\mathrm{e} = 2.718\ldots$). In this context, 'e' is almost invariably followed by a possibly non-trivial exponent, but never by a subscript. It is sometimes set off from the preceeding material by a little space at the author's or publisher's discretion.

  Several publisher's macro packages define a control sequence `\e` for this purpose, usually defined simply as a synonym for `\mathrm{e}`. Elsevier's macros, however, define

  ```
  \def\e{\mathop{\mathrm{e}}\nolimits}
  ```

  which also affects the spacing.
- 'i' when used as the imaginary unit ($\mathrm{i} = \sqrt{-1}$). Some sub-fields in physics such as electrical engineering may use 'j' instead to avoid confusion with the AC-current $i(t)$.

  Some publisher's macro packages define a control sequence `\i` as a synonym for `\mathrm{i}`, thereby replacing or overwriting the usual meaning of text mode dotless-i, which is not applcable in math mode anyway. It may be possible to retain both meanings by employing a suitable switching macro depending on the mode.

3.2 Derivatives, increments, variations, etc.

In addition to mathematical constants, several Latin letters should be printed in upright (roman) type, when used in a well-defined context as mathematical operators. In particular, this applies to:

- 'd' when used to denote the integration variable in an integral, a total derivative, or as a total (exact) differential.

  Some publisher's macro packages define a control sequence `\d` as a synonym for `\mathrm{d}`, thereby replacing or overwriting the usual command for the dot-under accent in text mode. It may be possible to retain both meanings by employing a suitable switching macro.
- 'D' when used a differential along geodesics in general relativity (relatively rare).
- '∂' when used as a partial derivative. This convention is rarely observed so far, presumably for lack of suitable fonts. If an upright `\partial` becomes available in a new math font encoding, it may be reasonable to make this the default choice when in a physics context.
- 'δ' when used as a variational derivative, or as an inexact differential, e.g. $\mathrm{d}U = \delta Q + \delta W$.
- 'Δ' when used to denote a finite increment.
- 'ð' and 'þ' when used as a weighted derivative in quantum field theory (rare).

Some slashed or barred variants of 'd' or 'δ' are ocassionally used in thermodynamics to distinguish between (exact or inexact) differental quantities in reversible or irreversible processes. However, such a notation presumably might be an author's invention and thus might be considered non-standard.

3.3 Operators in vector caculus

- Nabla (gradient, divergence, rotation/curl)
- Laplace operator (Delta)
- d'Alembert operator ("Quabla")

4 Requirements for special notations

4.1 Special brackets, braces, etc.

- (anti-)commutator in quantum theory: $[p, q]_\pm$
- Poisson- or Lie-brackets in theoretical mechanics: $[p, q]$ or $\{p, q\}$
- bra-ket notation in quantum theory, quantum states, averages, projection operator: $\langle x|H|x\rangle$
- ensemble averages, time averages: $\langle x\rangle$, $\bar{x}$
- directions, planes, coordinates in crystals

4.2 Barred letters

- $\hbar$ (`\hbar`) and/or $\hslash$ (`\hslash`), both representing a short-hand notation for $h/2\pi$ where h is Planck's constant. This notation is very well-established and widely used in quantum theory.
- $\lambdabar = \hbar/(m_\mathrm{e}c)$, representing the Compton wavelenght. A suitable macro is provided as `\lambdabar` in the RevTEX macro package.
- ι, used in a sub-field of fusion plasma physics to denote $\iota/2\pi$ where ι is the rotational transform of magnetic field lines. Another sub-field of plasma physics prefers to use a "safety factor" defined as $q = 2\pi/\iota = 1/\iota$ instead, thus making it unnecesary to use unusual notation.

4.3 Slashed letters in QED

- Feynman slash notation

4.4 Operators in quantum mechanics

- Operators vs. classical quantities
- Complex conjugate states, denoted by 'a^*' in quantum mechanics where '$\bar{a}$' would normally be used for the complex conjugate in a math context.
- (Self-)adjoint operators, denoted by 'a^+' or '$a^\dagger$'. This example indicates the use of the dagger as a math symbol, not just as a footnote symbol.

5 Summary

The following math alphabets are needed in the `MC`-encoding providing access to both Latin and Greek letters in a given shape:

- italic (sloping) type (`\mathnormal`), for symbols denoting physical quantities
- upright (roman) type (`\mathrm`), for particle symbols or chemical elements
- bold italic (sloping) type, for vectors (recommended notation)
- bold sans serif italic (sloping) type, for tensors of the second rank
- bold upright type (`\mathbf`), for vectors (widely-used alternative notation)
- bold sans serif upright type, for tensors (widely-used alternative notation)

The following math alphabets are needed only for Latin letters and may thus be implemented either using a `T1`-encoded or `MC`-encoded font:

- upright (roman) type (`\mathrm`), for textual subscripts
- upright (roman) type (`\mathrm`), for physical units (with few exceptions)
- text italic type (`\mathit`), for dimensionless parameters

References

[1] University of Chicago Press. *The Chicago Manual of Style*. University of Chicago Press, Chicago and London, 13th revised edition, 1982.

[2] Ellen Swanson. *Mathematics into Type*. American Mathematical Society, revised edition, 1979.

[3] International Union of Pure and Applied Physics, S.U.N. Commission. Symbols, units and nomenclature in physics. *Physica A*, 93:1–60, 1978. Document U.I.P. 20 (1978).

[4] International Union of Pure and Applied Physics, S.U.N. Commission. Symbols, units and nomenclature in physics. *Physica A*, 146:1–67, 1987.

[5] ISO. Quantities and units. Technical Report ISO 31-0–ISO 31-13, International Organization for Standardization, Geneva, Switzerland, 1992.

[6] Barry N. Taylor. The International System of Units (SI). NIST Special Publication 330, National Institute of Standards and Technology, Washington, August 1991. vi+56 pp., available from `http://www.nist.gov/`.

[7] Barry N. Taylor. Guide for the Use of the International System of Units (SI). NIST Special Publication 811, National Institute of Standards and Technology, Washington, April 1995. x+74 pp., available from `http://www.nist.gov/`.

[8] American Physical Society. *The REVTEX Input Guide*. American Physical Society, September 1996. For REVTEX 3.1.

[9] American Physical Society. *Physical Review Letters REVTEX Electronic-Submission Compuscript Program*. American Physical Society, March 1996.

[10] American Physical Society. *Physical Review Style and Notation Guide*. American Physical Society, revised edition, February 1993.

[11] Institue of Physics Publishing. *Preparing an article for publication in an IOP journal using LATEX*. Institue of Physics Publishing, Bristol, 1994.

[12] Institue of Physics Publishing. *Using the AMS fonts with LATEX 2.09 and the IOP journals's style file*. Institue of Physics Publishing, Bristol, 1994.

[13] Elsevier Science. *Preparing articles with LATEX*. Elsevier Science, Amsterdam, Oxford, New York, Lausanne, Tokyo, 1995.

[14] IAEA Publishing Team. *How to prepare a Nuclear Fusion article*. International Atomic Energy Agency, Vienna, 1996.

[15] Jörg Knappen. Changing the appearance of math. In *EuroTEX '92, Proceedings of the 7th European TEX Conference*, pages 212–216, Prague, Czechoslovakia, September 14–18 1992.

[16] Michel Goossens and Eric van Herwijnen. The elementary Particle Entity Notation (PEN) scheme. *TUGboat*, 13(2):201–207, July 1992.

[17] Michel Goosens, Sebastian Rahtz, and Frank Mittelbach. *The LATEX Graphics Companion*. Addison-Wesley, 1997.

[18] Claudio Beccari. Typesetting mathematics for science and technology according to ISO 31/XI. *TUGboat*, 18(1):39–48, March 1997.

 Task 2: Requirements analysis

[19] Michel Goosens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, 1994.

◇ Ulrik Vieth
Heinrich-Heine-Universität
Institut für Theoretische Physik II
Universitätsstraße 1
D-40225 Düsseldorf
Germany
vieth@thphy.uni-duesseldorf.de
URL: http://www.thphy.uni-duesseldorf.de/~vieth/