Using the `clrscode4e` Package in LaTeX 2ε

Thomas H. Cormen
thc@cs.dartmouth.edu

February 19, 2022

1 Introduction

This document describes how to use the `clrscode4e` package in LaTeX 2ε to typeset pseudocode in the style of *Introduction to Algorithms*, Fourth edition, by Cormen, Leiserson, Rivest, and Stein (CLRS 4e) [1]. You use the commands[1] in the same way we did in writing CLRS 4e, and your output will look just like the pseudocode in the text.

2 Setup

To get the `clrscode4e` package, download https://mitp-content-server.mit.edu/books/content/sectbyfn/books_pres_0/11599/clrscode4e.sty, and put it where it will be found when you run LaTeX 2ε. To use the package, include the following line in your source file:

```
\RequirePackage{clrscode4e}
```

The `clrscode4e` package itself includes the line

```
\RequirePackage{graphics}   % needed for \scalebox command
```

This line is necessary in order to get the right spacing for the == symbol that we use for equality tests. Therefore, you will need to have the `graphics` package installed and available on your system.

Because comments are in dark red, and because the `clrscode4e` package allows for a color screen behind pseudocode, the package also includes the line

```
\RequirePackage[cmyk]{xcolor}
```

You may change or remove the `[cymk]` option; we had to include it so that the book could be printed in a four-color process.

[1] We use the term "command" rather than "macro" throughout this document, though "macro" would work just as well.

3 Typesetting names

Pseudocode in CLRS 4e uses four types of names: identifiers, procedures, constants, and fixed functions. We provide commands `\id`, `\proc`, `\const`, and `\func` for these names. Each of these commands takes one argument, which is the name being typeset. These commands work both in and out of math mode. Even when these commands are used in math mode and the name given as an argument contains a dash, the dash is typeset as a hyphen rather than as a minus sign.

Identifiers: Use identifiers for variable names of more than one character. When a variable name is just a single character, e.g., the identifier i in line 1 of INSERTION-SORT on page 19, we just typeset it in math mode rather than using the `\id` command: `$i$`.

Do not typeset identifiers consisting of two or more characters, e.g., the variable *key* in line 2 of INSERTION-SORT, in this way. (See page 51 of Lamport [2].) Although LaTeX 2ε provides the `\mathit` command for typesetting multiletter identifiers, use our `\id` command instead: `\id{key}`, rather than `\mathit{key}` or—horrors!—`$key$`. Since the `\id` command may be used both in and out of math mode, the source text

```
Line~5 uses the variable \id{key} in the test $A[j] > \id{key}$.
```

will produce

> Line 5 uses the variable *key* in the test $A[j] > key$.

To see how a dash turns into a hyphen, consider line 1 of ONLINE-MAXIMUM on page 150. It contains the variable *best-score*. Typesetting this variable name by `\id{best-score}` produces a hyphen in the identifier, but typesetting it by `\mathit{best-score}` would produce $best - score$, with a minus sign—rather than a hyphen—in the identifier.

Procedures: For procedure names, use the `\proc` command. It typesets procedure names in small caps, and dashes (which occur frequently in our procedure names) are typeset as hyphens. Thus, the source `\proc{Insertion-Sort}` produces INSERTION-SORT. Since you can use the `\proc` command both in and out of math mode, the source text

```
Call \proc{Insertion-Sort} with an array $A$ and its size $n$, so that
the call is $\proc{Insertion-Sort}(A, n)$.
```

will produce

> Call INSERTION-SORT with an array A and its size n, so that the call is INSERTION-SORT(A, n).

Constants: We typeset constants such as NIL, TRUE, and RED in small caps with the `\const` command, e.g., `\const{nil}`, `\const{true}`, and `\const{red}`. The `\const` command typesets a dash within a constant name as a hyphen, so that, as on page 411, `\const{no-such-path}` will produce NO-SUCH-PATH.

Fixed functions: We typeset the names of fixed functions in plain old roman with the `\func` command, e.g., level and out-degree. By a "fixed function," we mean a function that is a specific, given function. For example, the sin function is typically typeset in roman; $\sin x$ looks right, but wouldn't $sin\ x$ look strange? Yet, on page 56, $\Theta(g(n))$ looks right, but $\Theta(\mathrm{g}(n))$ would look wrong, since g is a variable that stands for any one of a number of functions.

As with the other commands for names, a dash within a function name will typeset as a hyphen, so that `\func{out-degree}` will produce out-degree rather than out − degree. Note that LaTeX 2_ε provides commands for many fixed functions, such as sin and log; Table 3.9 on page 44 of Lamport [2] lists these "log-like" functions.

4 Typesetting object attributes

In the first two editions of the book, we used square brackets for object attributes. For example, we represented the length of an array A by $length[A]$. Based on requests from readers, we switched to the object-like dot-notation in the third edition and continued this notation into the fourth edition, so that we denote the length of array A by $A.length$.

You might think that you could typeset $A.length$ by `$A.\id{length}$`, but that would produce $A.length$, which has not quite enough space after the dot. Therefore, we created a set of commands to typeset object attributes. Each one may be used either in or out of math mode.

Most of the time, we use the `\attrib` command, which takes two arguments: the name of the object and the name of the attribute. Let's make a couple of definitions:

- An ***i-string*** is a string that you would use in an `\id` command, typically one or more non-Greek letters, numerals, or dashes.
- An ***x-string*** is a string that you would not use in an `\id` command, typically because it has a subscript or one or more Greek letters.
- As a special, and very common, case, a single non-Greek letter can count as either an i-string or an x-string.

The `\attrib` command works well when the object name is an x-string and the attribute name is an i-string. For example, to produce $A.length$, use `\attrib{A}{length}`. Here, we treat the object name, A, as an x-string. The attribute name, *length*, is of course an i-string.

If all your objects are x-strings and all your attributes are i-strings, then the `\attrib` command will be all you need. We provide several other commands for other situations that arose when we produced CLRS 4e.

The four basic attribute commands are `\attribxi`, `\attribxx`, `\attribii`, and `\attribix`. Each takes two arguments: the object name and the attribute name. The last two letters of the command name tell you what type of strings the arguments should be. The next-to-last letter tells you about the object name, and the last letter tells you about the attribute name. `i` indicates that the argument will be treated as an `\id`, in which case the command calls the `\id` command and also puts the right amount of space between the argument and the dot.

- You may use `\attribxi` precisely when you would use `\attrib`. In fact, `\attrib` is just a call to `\attribxi`.

- Use `\attribxx` when both the object name and attribute names are x-strings. For example, you would use `\attribxx` if the attribute name has a subscript, so that to produce $x.c_i$, you would use `\attribxx{x}{c_i}`. Another situation in which you would use `\attribxx` is when the attribute name is a Greek letter: to produce $v.\pi$, use `\attribxx{v}{\pi}`.

- If both the object name and attribute name are i-strings, then you should use `\attribii`. For example, `\attribii{item}{key}` produces *item*.*key*, and `\attribii{prev-item}{np}` produces *prev-item*.*np*.

- If the object name is an i-string and the attribute name is an x-string, then use `\attribix`. (We never had this situation arise in CLRS 4e.) But if we had wanted to produce $item.\pi$, we would have used `\attribix{item}{\pi}`.

For convenience, the `clrscode4e` package also contains commands for cascading attributes, such as $x.left.size$. These commands string together calls to the appropriate `\attribxi` and `\attribxx` commands. The number of arguments they take depends on how many attributes you are stringing together.

- When you have two attributes, use `\attribb`, which takes an object name and two attribute names: `\attribb{x}{left}{size}` produces $x.left.size$. This command assumes that the object name is an x-string and both attribute names are i-strings.

- For three attributes, use `\attribbb`, which takes an object name (an x-string) and three attribute names (i-strings): to produce $y.p.left.size$, use `\attribbb{y}{p}{left}{size}`.

- For four attributes, use `\attribbbb`, which is like `\attribbb` but with one additional attribute tacked on. We never needed to use this command in CLRS 4e.

- The `\attribbxxi` command is for one level of cascading where the first attribute given is an x-string. For example, `\attribbxxi{x}{c_i}{n}` produces $x.c_i.n$.

If your cascading attributes do not fit any of these descriptions, you'll have to roll your own command from the `\attribxx` and `\attribxi` (or `\attrib`) commands. For example, suppose you want to produce $x.left.key_i$. Because it has a subscript, key_i is an x-string, and so you should not use `\attribb`. Instead, use `\attribxx{\attribxi{x}{left}}{\id{key_i}}`. (You could replace the call of `\attribxi` by a call of `\attrib`.) Note that this call treats key_i as an attribute of $x.left$, which is correct, rather than treating $left.key_i$ as an attribute of x, which is not correct.

Edges of a graph can have attributes, too, and the `clrscode4e` package provides two commands for attributes of edges. These commands assume that the edges are of the form (u, v), where the vertices u and v are x-strings. They take three parameters: the two vertices that define the edge and the name of the attribute.

- When the attribute name is an i-string, use `\attribe`. For example, to produce $(u, v).c$, use `\attribe{u}{v}{c}`.

- When the attribute name is an x-string, use `\attribex`. For example, to produce $(u, v).c'$, use `\attribex{u}{v}{c'}`.

5 Miscellaneous commands

The `clrscode4e` package contains four commands that don't really fit anywhere else, so let's handle them here. All four must be used in math mode.

- We denote subarrays with the ":" notation, which is produced by the `\subarr` command.[2] Thus, the source text `$A[1 \subarr i-1]$` will produce $A[1 : i - 1]$.
- We use the `\gets` command for the assignment operator. For example, line 4 of INSERTION-SORT on page 19 is `$j \gets i - 1$`, producing $j = i - 1$.
- We use the `\isequal` command to test for equality with the == symbol. For example, line 1 of PUSH on page 255 contains the test $S.top == S.size$, which we get by typesetting `$\attrib{S}{top} \isequal \attrib{S}{size}$`.
- Since the constant NIL appears frequently, the command `\nil` is equivalent to `\const{nil}`.

You might wonder why we bother with the `\gets` command when we could just typeset an equals sign directly. The answer is that in the first two editions of *Introduction to Algorithms*, we used a different symbol (a left arrow) for the assignment operator, and it made sense to use a command for that. Many readers told us that they preferred to use an equals sign for assignment—as many programming languages use—and so we made this change for the third edition. But it's a good idea to continue using the `\gets` command so that we can easily change our assignment operator should we desire to do so in the future.

Once we decided to use the equals sign for assignment, we could no longer use it for equality tests. We created the `\isequal` command for equality tests, and we decided to base it on the double equals sign used for equality tests in programming languages such as C, C++, Java, and Python. Typesetting it as == in math mode produces ==, which is too wide for our tastes. Our `\isequal` command calls the `\scalebox` command from the `graphics` package to narrow the symbol, and it puts a nice amount of space between the equals signs: ==.

6 Pseudocode boxes

We typeset pseudocode by putting it in a `codebox` environment. A `codebox` is a section of code that does not break across pages. Starting with CLRS 4e, we allow blocks of pseudocode to float, just like figures. To allow pseudocode blocks to float, use the `floatingcodebox` option in the `RequirePackage` command: `\RequirePackage[floatingcodebox]{clrscode4e}`. The `floatingcodebox` option requires that you have the `float` package installed and available on your system. Starting with CLRS 4e, we also put a light tan screen behind the pseudocode. For the light tan screen, use the `screen` option in the `RequirePackage` command: `\RequirePackage[screen]{clrscode4e}`. For floating pseudocode boxes that also have the light tan screen, use both options:
`\RequirePackage[screen,floatingcodebox]{clrscode4e}`.
The default behavior is no floating pseudocode boxes and no screen.

[2] In previous editions, we used the `\twodots` command, which produced two-thirds of an ellipsis. If you use `\twodots` with the `clrscode4e` package, LaTeX 2_ε will give you a warning.

Contents of a `codebox`

The typical structure within a `codebox` is as follows. Usually, the first line is the name of a procedure, along with a list of parameters. (Not all `codebox`es include procedure names; for example, see the pseudocode on page 504 of CLRS 4e.) After the line containing the procedure name come one or more lines of code, usually numbered. Some of the lines may be unnumbered, being continuations of previous lines. Lines are usually numbered starting from 1, but again there are exceptions, such as the pseudocode on page 504.

Using `\Procname` to name the procedure

The `\Procname` command specifies the name of the procedure. It takes as a parameter the procedure name and parameters, typically all in math mode. `\Procname` makes its argument flush left against the margin, and it leaves a little bit of extra space below the line. For example, here is how we typeset the INSERTION-SORT procedure on page 19:

```
\begin{codebox}
\Procname{$\proc{Insertion-Sort}(A, n)$}
\li \For $i \gets 2$ \To $n$
\li     \Do
            $\id{key} \gets A[i]$
\li         \Comment{Insert $A[i]$ into the sorted subarray $A[1 \subarr i-1]$.}
\li         $j \gets i-1$
\li         \While $j > 0$ and $A[j] > \id{key}$
\li             \Do
                    $A[j+1] \gets A[j]$
\li                 $j \gets j-1$
                \End
\li         $A[j+1] \gets \id{key}$
        \End
\end{codebox}
```

Multiple procedures in one box

Starting with CLRS 4e, we have a way to force multiple procedures (or more generally, blocks of pseudocode) to occupy the same box, so that they appear together on the same page. The first procedure still follows `\begin{codebox}` and the last procedure still precedes `\end{codebox}`, but use `\begin{innercodebox}` before all procedures after the first, and use `\end{innercodebox}` after all procedures except the last. For example, if you have two procedures that you want to keep together, type in the following:

```
\begin{codebox}
[First procedure goes here]
\end{innercodebox}

\begin{innercodebox}
[Second procedure goes here]
\end{codebox}
```

If you have three procedures:

```
\begin{codebox}
[First procedure goes here]
\end{innercodebox}

\begin{innercodebox}
[Second procedure goes here]
\end{innercodebox}

\begin{innercodebox}
[Third procedure goes here]
\end{codebox}
```

If you are using floating pseudocode boxes, then instead of the `codebox` environment, you can attempt to force a pseudocode box to the top of a page, the bottom of a page, where it appears within your source text, or on its own page via the environments `codeboxt`, `codeboxb`, `codeboxh`, and `codeboxp`, respecitvely. We say "attempt to force" because LaTeX 2_ε doesn't always do quite what you expect with float placement.

If you have a pseudocode box that is just a little too long or wide to fit on the page, you can reduce it:

```
\reducecodeboxtrue
\codeboxreduction{XX}
\begin{codebox}
[Pseudocode goes here]
\end{codebox}
\reducecodeboxfalse
```

where `XX` is the reduction amount. For example, to reduce the size of the pseudocode box and its contents by 10%, so that it is 90% of its nominal size, use `\codeboxreduction{0.9}`.

Using \li and \zi to start new lines

To start a new, numbered line, use the `\li` command. To start a new, *un*numbered line, use the `\zi` command. Note that since a `codebox` is not like the `verbatim` environment, the line breaks within the source text do not correspond to the line breaks in the typeset output.

Commands for keywords

As you can see from the source for INSERTION-SORT, the commands `\For` and `\While` produce the keywords **for** and **while** in boldface within a `codebox`.

Sometimes you want to include a keyword in the main text, as I have done in several places in this document. Use the `\kw` command to do so. For example, to produce the previous paragraph, I typed in the following:

```
As you can see from the source for \proc{Insertion-Sort}, the commands
\verb`\For` and \verb`\While` produce the keywords \kw{for} and
\kw{while} in boldface within a \texttt{codebox}.
```

The following commands simply produce their corresponding keywords, typeset in boldface: `\For`, `\To`, `\Downto`, `\By`, `\While`, `\If`, `\Return`, `\Error`, `\Spawn`, `\Sync`, and `\Parfor` (which produces the compound keyword **parallel for**). Although you could achieve the same effect with the `\kw` command (e.g., `\kw{for}` instead of `\For`), you will find it easier and more readable to use the above commands in pseudocode. None of the above commands affects indentation.

The `\Comment` command takes as an argument the text of a comment. It produces the comment symbol //, followed by a space, followed by the argument, all in dark red. To get the comment symbol without a following space, use `\CommentSymbol`.

Loops

The INSERTION-SORT example above shows typical ways to typeset **for** and **while** loops. In these loops, the important commands are `\Do` and `\End`. `\Do` increments the indentation level to start the body. Put `\Do` on a line starting with `\li`, but don't put either `\li` or `\zi` between the `\Do` command and the first statement of the loop body. Use `\li` or `\zi` in front of all loop-body statements after the first one. `\End` simply decrements the indentation level, and you use it to end any **for** or **while** loop, or otherwise decrement the indentation level.

In the first two editions of the book, the body of a **for** or **while** loop began with the keyword **do**. Responding to requests from readers to make pseudocode more like C, C++, and Java, we eliminated this keyword in the third edition.

As you can see from the above example, I like to place each `\Do` and `\End` on its own line. You can format your source text any way you like, but I find that the way I format pseudocode makes it easy to match up `\Do`-`\End` pairs.

If you want your **for** loop to decrease the loop variable in each iteration, use `\Downto` rather than `\To`. If you want the stride to be a value other than 1, use the `\By` command. For example, line 3 of Exercise 19.2-2 on page 526 is typeset as

```
\For $i \gets 1$ \To $15$ \By $2$
```

Loops that use the **repeat**-**until** structure are a bit different. We use the `\Repeat` and `\Until` commands, as in the HASH-INSERT procedure on page 294:

```
\begin{codebox}
\Procname{$\proc{Hash-Insert}(T, k)$}
\li $i \gets 0$
\li \Repeat
\li     $q \gets h(k, i)$
\li     \If $T[q] \isequal \const{nil}$
\li         \Then
                $T[q] \gets k$
\li             \Return $q$
\li         \Else
                $i \gets i+1$
            \End
\li \Until $i \isequal m$
\li \Error ``hash table overflow''
\end{codebox}
```

Note that the `\Until` command has an implied `\End`.

Typesetting if statements

As you can see from the above example of HASH-INSERT, we typeset **if** statements with the commands `\If`, `\Then`, `\Else`, and `\End`. In the first two editions of the book, the keyword **then** appeared in pseudocode, but—again mindful of requests from our readers to make our pseudocode more like C, C++, and Java—we eliminated the keyword **then** in the third edition. The `\Then` command remains, however, in order to indent the code that runs when the test in the **if** clause evaluates to TRUE.

We use `\End` to terminate an **if** statement, whether or not it has an **else** clause. For an example of an **if** statement without an **else** clause, here's the RANDOMIZED-QUICKSORT procedure on page 192:

```
\begin{innercodebox}
\Procname{$\procdecl{Randomized-Quicksort}(A, p, r)$}
\li \If $p < r$
\li     \Then
            $q \gets \proc{Randomized-Partition}(A, p, r)$
\li         $\proc{Randomized-Quicksort}(A, p, q-1)$
\li         $\proc{Randomized-Quicksort}(A, q+1, r)$
        \End
\end{codebox}
```

(It starts with `\begin{innercodebox}` because it is in the same pseudocode box as RANDOMIZED-PARTITION.) The HASH-INSERT procedure above shows how to typeset an **if** statement that has an **else** clause. Note that `\Then` and `\Else` always follow an `\li` command to start a new numbered line. As with the `\Do` command, don't put either `\li` or `\zi` between `\Then` or `\Else` and the statement that follows.

As you can see, I line up the `\End` commands under the `\Then` and `\Else` commands. I could just as easily have chosen to line up `\End` under the `\If` command instead. I also sometimes elect to put the "then" or "else" code on the same source line as the `\Then` or `\Else` command, especially when that code is one short line, such as in line 2 of the RANDOMIZED-MARKING procedure on page 808:

```
\begin{codebox}
\Procname{$\proc{Randomized-Marking}(b)$}
\li \If block $b$ resides in the cache,
\li     \Then $\attrib{b}{mark} \gets 1$
\li     \Else
\li         \If all blocks $b'$ in the cache have $\attrib{b'}{mark} = 1$
\li             \Then unmark all blocks $b'$ in the cache,
                    setting $\attrib{b'}{mark} = 0$
                \End
\li         select an unmarked block $u$ with $\attrib{u}{mark} = 0$
                uniformly at random
\li         evict block $u$
\li         place block $b$ into the cache
\li         $\attrib{b}{mark} \gets 1$
        \End
\end{codebox}
```

Our style is to put the first line within an **else** clause on the same line as the keyword **else**, as in HASH-INSERT. Doing so saves vertical space. There is one exception: starting in the fourth edition, when the first

line within an **else** clause is an **if** statement, we put the **if** statement on a new line, as in RANDOMIZED-MARKING.

Sometimes, you need more complicated "**if**-ladders" than you can get from the `\Then` and `\Else` commands. The TRANSPLANT procedure on page 324 provides an example, and it uses the `\ElseIf` and `\ElseNoIf` commands:

```
\begin{codebox}
\Procname{$\proc{Transplant}(T, u, v)$}
\li \If $\attrib{u}{p} \isequal \const{nil}$
\li     \Then $\attrib{T}{root} \gets v$
\li \ElseIf $u \isequal \attribb{u}{p}{left}$
\li     \Then $\attribb{u}{p}{left} \gets v$
\li \ElseNoIf
        $\attribb{u}{p}{right} \gets v$
    \End
\li \If $v \neq \const{nil}$
\li     \Then $\attrib{v}{p} \gets \attrib{u}{p}$
        \End
\end{codebox}
```

For an **if**-ladder, use `\Then` for the first case, `\ElseNoIf` for the last case, and `\ElseIf` followed by `\Then` for all intermediate cases. You use `\ElseNoIf` like you use `\Else` in that it follows an `\li` command, you don't follow it with `\Then`, and, because it terminates an **if**-ladder, it's followed by `\End`. I usually line up the terminating `\End` with `\If`, the `\ElseIf` commands, and `\ElseNoIf`, but the way you line it up won't change the typeset output.

The reason that we started placing the first line of an **else** clause on a new line when it's an **if** statement was to avoid confusion between **else if** (the **else** clause starts with an **if** statement) and **elseif** (a test within an **if**-ladder). (We received a few false-positive bug reports where readers misinterpreted **else if** as **elseif**.)

Indentation levels

If you look at the RECURSIVE-MATRIX-CHAIN procedure on page 389, you'll see that line 5 of the pseudocode occupies three lines on the page, and that the last two lines comprising line 5 are indented by one level. We do so with the `\Indentmore` command. The `\End` command following the indented line decrements the indentation level back to what it was prior to the `\Indentmore`. If I had wanted to indent the line by two levels, I would have used two `\Indentmore` commands before the line and two `\End` commands afterward. (Recall that `\End` simply decrements the indentation level.)

To show you `\Indentmore`, here is the source code for RECURSIVE-MATRIX-CHAIN:

```
\begin{codebox}
\Procname{$\proc{Recursive-Matrix-Chain}(p, i, j)$}
\li \If $i \isequal j$
\li     \Then \Return $0$
        \End
\li $m[i,j] \gets \infty$
\li \For $k \gets i$ \To $j-1$
\li     \Do
            $q \gets \proc{Recursive-Matrix-Chain}(p, i, k)$
                \Indentmore
\zi                 $\mbox{} + \proc{Recursive-Matrix-Chain}(p, k+1, j)$
\zi                 $\mbox{} + p_{i-1} p_k p_j$
                \End
\li         \If $q < m[i,j]$
\li             \Then $m[i,j] \gets q$
                \End
        \End
\li \Return $m[i,j]$
\end{codebox}
```

(The empty `\mbox{}` commands are there so that LaTeX 2_ε will treat the operator "+" as a binary operator, rather than a unary one. A unary "+" or "−" has less space to the right of the operator than a binary one.)

Upon seeing the `\end{codebox}` command, the `codebox` environment checks that the indentation level is back to where it was when it started, namely an indentation level of 0. If it is not, you will get a warning message like the following:

```
Warning: Indentation ends at level 1 in codebox on page 1.
```

This message would indicate that there is one missing `\End` command. On the other hand, you might have one too many `\End` commands, in which case you would get

```
Warning: Indentation ends at level -1 in codebox on page 1.
```

Whenever the indentation level is nonzero upon hitting an `\end{codebox}` command, you'll get a warning telling you what the indentation level was.

If you are really picky, you can get a situation where you have two procedures in the same `codebox`, one of the procedures has 10 or more lines, but the other procedure has fewer than 10 lines. You'd like the ones digits in the line numbers to vertically align in the two procedures, but the macros for typesetting pseudocode push the line numbers to the left. The result is that the for the shorter procedure, the ones digits line up with the tens digits in the longer procedure. This situation arose in a few places, such as on page 372, where a `codebox` contains the procedures EXTENDED-BOTTOM-UP-CUT-ROD, with 10 lines, and PRINT-CUT-ROD-SOLUTION, with only four lines. The solution is to insert the command `\indentlinenumbers` just before the closing `\end{codebox}` (or, if appropriate, `\end{innercodebox}`) of the shorter procedure.

Tabs

I find that it is best to set the tab stops in the text editor to every 4 characters when typing in and displaying pseudocode source with the `clrscode4e` package. I use emacs, and to get the tabs set up the way I want them, my `tex-mode.el` file includes the line `(setq tab-width 4)`.

A `codebox` environment has a `tabbing` environment within it. Each tab stop gives one level of indentation. We designed the indentation so that the body of an **else** clause starts at just the right indentation. With one big exception, you won't need to be concerned with tabs. The primary exception is when you want to include a comment at the end of a line of pseudocode, and especially when you want to include comments after several lines and you want the comments to vertically align.

If you used the `clrscode` package from the second edition of the book, you might notice different tabbing behavior when you port your pseudocode to the `clrscode4e` package. Where the `clrscode` package used two tab stops for each level of loop indentation, the `clrscode4e` package uses just one tab stop. We made this change in the `clrscode4e` package because the third edition eliminated the keyword **then** and left-aligns **else** with its corresponding **if**.

Note that the `tabbing` environment within a codebox has nothing to do with tabs that you enter in your source code; when you press the TAB key, that's the same as pressing the space bar in the eyes of LaTeX 2_ε.

Instead of typing `\>` for a tab within a `codebox` environment, you can type `\tab`. Even better, if you want to go to a specific tab stop, you can use the `\tabto` command. For example, to go to the seventh tab stop, type `\tabto{7}`. You will find the `\tabto` command useful for vertically aligning comments. Unfortunately, you will probably have to experiment with the argument to `\tabto` to get the positioning just where you want it.

As an example, here is the PARTITION procedure on page 184:

```
\begin{codebox}
\Procname{$\procdecl{Partition}(A, p, r)$}
\li $x \gets A[r]$ \tabto{7} \Comment{the pivot}
\li $i \gets p-1$ \tabto{7} \Comment{highest index into the low side}
\li \For $j \gets p$ \To $r-1$ \tabto{7}
            \Comment{process each element other than the pivot}
\li     \Do
            \If $A[j] \leq x$ \tabto{7}
                    \Comment{does this element belong on the low side?}
\li              \Then
                    $i \gets i+1$ \tabto{8}
                        \Comment{index of a new slot in the low side}
\li                 exchange $A[i]$ with $A[j]$ \tabto{8}
                        \Comment{put this element there}
                \End
        \End
\li exchange $A[i+1]$ with $A[r]$ \tabto{7}
        \Comment{pivot goes just to the right of the low side}
\li \Return $i+1$ \tabto{7} \Comment{new index of the pivot}
\end{codebox}
```

Line 6 has `\tabto{8}` instead of `\tabto{7}` because this line is just a little too long to use `\tabto{7}`. Because lines 5 and 6 are both within the "then" part of the **if** statement in line 4, it looks good for their comments to be vertically aligned, and so line 5 also has `\tabto{7}`.

The maximum value that the argument to the `\tabto` may take is 12. LaTeX 2_ε does not allow more tab stops than that. If you really need a tab stop to the right of the 12th tab stop, you can use `\tabto{12}` followed by one or more nonbreaking spaces (`~`).

If you use the `\tabto` command on an `\ElseNoIf` line, you will get one less tab than you expect. The MODULAR-EXPONENTIATION procedure on page 935 is a good example. We add an extra `\tab`

command after `\tabto{12}` on line 6 to actually get to the 12th tab stop. This source also shows how we use nonbreaking spaces after the last tab stop to push comments farther to the right:

```
\begin{codebox}
\Procname{$\proc{Modular-Exponentiation}(a, b, n)$}
\li \If $b \isequal 0$
\li     \Then \Return 1
\li \ElseIf $b \bmod 2 \isequal 0$
\li      \Then
             $d \gets \proc{Modular-Exponentiation}(a, b/2, n)$
                 \tabto{12}~~~\Comment{$b$ is even}
\li          \Return $(d \cdot d) \bmod n$
\li \ElseNoIf
        $d \gets \proc{Modular-Exponentiation}(a, b-1, n)$
            \tabto{12}\tab{1}~~~\Comment{$b$ is odd}
\li     \Return $(a \cdot d) \bmod n$
    \End
\end{codebox}
```

Sometimes, text in a comment spans two lines of pseudocode. For example, look at the TREE-DELETE procedure on page 325. The comment on line 8 continues onto line 9, and the comment on line 11 continues onto line 12. In each case, there is an indentation within the comment. Use the command `\tabincomment` to get that indentation. For example, here are lines 8 and 9 of TREE-DELETE:

```
\li             $\attrib{y}{right} \gets \attrib{z}{right}$
                     \tabto{9} \Comment{$z$'s right child becomes}
\li             $\attribb{y}{right}{p} \gets y$
                     \tabto{9} \Comment{\tabincomment{}$y$'s right child}
```

Referencing line numbers

The source files for CLRS 4e contain no absolute references to line numbers. We use *only* symbolic references. The `codebox` environment is set up to allow you to place `\label` commands on lines of pseudocode and then reference these labels. The references will resolve to the line numbers. Our convention is that any label for a line number begins with `\li:`, but you can name the labels any way that you like.

For example, here's how we *really* wrote the INSERTION-SORT procedure on page 19:

```
\begin{codebox}
\Procname{$\procdecl{Insertion-Sort}(A, n)$}
\li \For $i \gets 2$ \To $n$
                                          \label{li:ins-sort-for}
\li     \Do $\id{key} \gets A[i]$         \label{li:ins-sort-pick}
                                          \label{li:ins-sort-for-body-begin}
\li         \Comment{Insert $A[i]$ into the sorted subarray $A[1 \subarr i-1]$.}
\li         $j \gets i-1$                 \label{li:ins-sort-find-begin}
\li         \While $j > 0$ and $A[j] > \id{key}$
                                          \label{li:ins-sort-while}
\li             \Do
                    $A[j+1] \gets A[j]$   \label{li:ins-sort-while-begin}
\li                 $j \gets j-1$         \label{li:ins-sort-find-end}
                                          \label{li:ins-sort-while-end}
                \End
\li         $A[j+1] \gets \id{key}$       \label{li:ins-sort-ins}
                                          \label{li:ins-sort-for-body-end}
        \End
\end{codebox}
```

Note that any line may have multiple labels. As an example of referencing these labels, here's the beginning of the first item under "Pseudocode conventions" on pages 21–22:

```
\item Indentation indicates block structure.  For example, the body of
the \kw{for} loop that begins on line~\ref{li:ins-sort-for} consists
of lines
\ref{li:ins-sort-for-body-begin}--\ref{li:ins-sort-for-body-end}, and
the body of the \kw{while} loop that begins on
line~\ref{li:ins-sort-while} contains lines
\ref{li:ins-sort-while-begin}--\ref{li:ins-sort-while-end} but not
line~\ref{li:ins-sort-for-body-end}.
```

Setting line numbers

On rare occasions, we needed to start line numbers somewhere other than 1. Use the `setlinenumber` command to set the next line number. For example, in Exercise 22.2-2 on page 619, we want the line number to be the same as a line number within the DAG-SHORTEST-PATHS procedure on page 617. Here's the source for the exercise:

```
Suppose that you change line~\ref{li:dag-sp-loop-begin} of
\proc{Dag-Shortest-Paths} to read

\begin{codeboxnoscreenh}%
\setlinenumber{li:dag-sp-loop-begin}
\li \For the first $\card{V}-1$ vertices, taken in topologically
sorted order
\end{codeboxnoscreenh}
Show that the procedure remains correct.
```

(The `codeboxnoscreenh` environment produces a `codebox` without the color screen behind it and without floating it.)

The DAG-SHORTEST-PATHS procedure is

```
\begin{codebox}
\Procname{$\proc{Dag-Shortest-Paths}(G, w, s)$}
\li topologically sort the vertices of $G$      \label{li:dag-sp-topo-sort}
\li $\proc{Initialize-Single-Source}(G, s)$     \label{li:dag-sp-init}
\li \For each vertex $u \in \attrib{G}{V}$, taken in topologically sorted order
                                                \label{li:dag-sp-loop-begin}
\li     \Do
            \For each vertex $v \in \attrib{G}{Adj}[u]$
                                                \label{li:dag-sp-inner-begin}
\li             \Do $\proc{Relax}(u, v, w)$     \label{li:dag-sp-loop-end}
                \End
        \End
\end{codebox}
```

Even more rarely (just once, in fact), we needed to set a line number to be some other line number plus an offset. That was in the two lines of pseudocode on page 484, where the first line number had to be one greater than the number of the last line of LEFT-ROTATE on page 336. Use the `setlinenumberplus` command:

```
\begin{codeboxh}
\setlinenumberplus{li:left-rot-parent}{1}
\li $\attrib{y}{size} \gets \attrib{x}{size}$
\li $\attrib{x}{size} \gets \attribb{x}{left}{size}
        + \attribb{x}{right}{size} + 1
\end{codeboxh}
```

(The `codeboxh` environment produces a `codebox` with the color screen behind it but without floating it.) Here, the last line of LEFT-ROTATE has `\label{li:left-rot-parent}`.

Indenting long argument lists in procedure calls

You might find that you have to call a procedure with an argument list so long that the call requires more than one line. When this situation arises, it often looks best to align the second and subsequent lines of arguments with the first argument. The only place we did so was in the SUM-ARRAYS′ procedure in Problem 27-1 on page 783.

To get this style of alignment, use the `\Startalign` and `\Stopalign` commands, in concert with the `\>` command of LaTeX 2_ε. The `\Startalign` command takes an argument that is the text string that you wish to align just to the right of. Start each line that you want to indent with `\>`. Use the `\Stopalign` command to restore indentation to its state from before the `\Startalign` command.

The source code for SUM-ARRAYS′ shows how to use these commands:

```
\begin{codebox}
\Procname{$\proc{Sum-Arrays}{$'$}(A, B, C, n)$}
\li $\id{grain-size} \gets\ \ ?$ \tabto{6} \Comment{to be determined}
\li $r \gets \ceil{n / \id{grain-size}}$
\li \For $k \gets 0$ \To $r-1$
\li     \Do \Spawn $\proc{Add-Subarray}(A, B, C, k \cdot \id{grain-size}+1,$
                   \Startalign{\Spawn $\proc{Add-Subarray}($}
\>                 $\min\set{(k+1)\cdot\id{grain-size}, n})$
                   \Stopalign
        \End
\li \Sync
\end{innercodebox}
```

The second line of arguments in the call to ADD-SUBARRAY starts right under the first parameter, A, in the call. (And, once again, we use `\end{innercodebox}` because the procedure ADD-SUBARRAY follows SUM-ARRAYS′ in the same `codebox`.)

7 Reporting bugs

If you find errors in the `clrscode4e` package, please send me email (thc@cs.dartmouth.edu). It would be best if your message included everything I would require to elicit the error myself.

The `clrscode4e.sty` file contains the following disclaimer:

```
% Written for general distribution by Thomas H. Cormen, February 2022.

% The author grants permission for anyone to use this macro package and
% to distribute it unchanged without further restriction.  If you choose
% to modify this package, you must indicate that you have modified it
% prior to your distributing it.  I don't want to get bug reports about
% changes that *you* have made!
```

I have enough trouble keeping up with my own bugs; I don't want to hear about bugs that others have introduced in the package!

8 Revision history

- 19 February 2022. Initial revision of document and code.

References

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*, fourth edition. The MIT Press, 2022.

[2] Leslie Lamport. *LaTeX: A Document Preparation System User's Guide and Reference Manual*. Addison-Wesley, 1993.