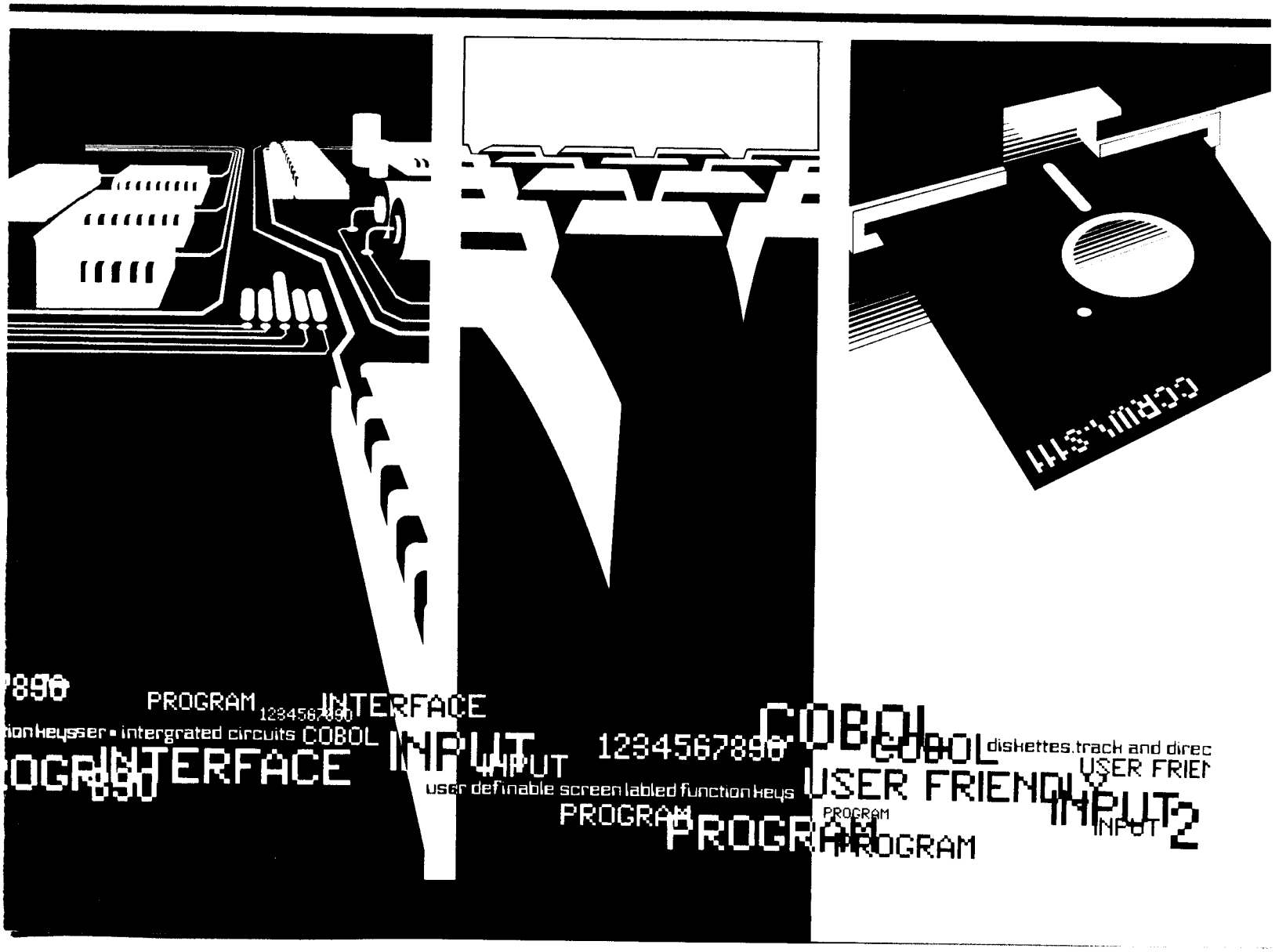
RESEARCH IN WORD PROCESSING NEWSLETTER

South Dakota School of Mines and Technology Rapid City, SD 57701 (605) 394-2481
Volume 5 Number 8 November 1987

Conference Notes
page 2



RESEARCH IN
WORD PROCESSING
NEWSLETTER

Editors
Bradford A. Morgan
James M. Schwartz

Contributing Editors
Bryan Pfaffenberger
University of Virginia
Tom Carney
University of Windsor

Production Manager
Dan Dorland
Art Director
Francis Jacobs
Typesetting/Graphics
Brenda Boyer
Photography
Dave Adams
Printing
Dick Beshara
Circulation
Wendy Painter

Accounting Manager
Sondra Wagner

Research in Word Processing Newsletter. Volume 5, Number 8. Copyright © 1987 by the South Dakota School of Mines and Technology. All rights reserved. ISSN: 0748-5484. Also indexed by **ERIC, INSPEC, COMPUTER LITERATURE INDEX,** and **SOFTWARE REVIEWS ON FILE.** The *Research in Word Processing Newsletter* is published 9 times a year (September through May) by the South Dakota School of Mines and Technology, Rapid City, South Dakota 57701-3995; (605)394-2481. Postage paid at Rapid City, South Dakota. **SUBSCRIPTION:** $18 per year (U.S.); $24 per year (Canada); $27 per year (foreign). Address all subscription inquiries and manuscripts to the Editors, *Research in Word Processing Newsletter,* SDSM&T, 501 E. St. Joseph, Rapid City, SD 57701-3995. Please allow 4 to 6 weeks for subscription processing. **POSTMASTER:** Send address changes to RWPN, South Dakota School of Mines and Technology, 501 East St. Joseph, Rapid City, South Dakota 57701-3995.

Computing Conference in Texas

The National Educational Computing Conference, NECC '88, will come together June 14-17, 1988, in Dallas, Texas, including a first-day workshop. All aspects of educational computing will be covered, and conference proceedings will be published. Papers were accepted until the November 1, 1987, deadline. Contact James L. Poirot or Cathleen A. Norris, Computer Science Department, North Texas State University, Box 13886, Denton, TX 76203-3886, or call (817) 565-2818.

Call for Papers in Computational Linguistics

The 26th Annual Meeting of the Association for Computational Linguistics will take place June 7-10, 1988, at the State University of New York in Buffalo. Papers will be accepted until January 4, 1988, on various aspects of computational linguistics, including pragmatics, discourse, semantics, syntax, and the lexicon; phonetics, phonology, and morphology; interpreting and generating spoken and written language; linguistic, mathematical, and psychological models of language, machine translation and translation aids; natural language interfaces; message understanding systems; and theoretical and applications papers of every kind. For submission details, contact Jerry R. Hobbs, ACL88 Program Chair, Artificial Intelligence Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, or call (415) 859-2229.

Literary and Linguistic Computing Conference in Jerusalem

On June 5-9, 1988, the Association for Literary and Linguistic Computing (ALLC) will hold its fifteenth international conference in Jerusalem, Israel. December 15, 1987, is the deadline for receiving abstracts of proposed papers in the following areas: ■ computational morphology, syntax and semantics ■ computational lexicography and lexicology ■ mechanized dictionaries, lexicons and grammars ■ lemmatization and parsing ■ ambiguity and its mechanical resolution ■ stylistic analysis and authorship studies ■ statistical linguistics and metrics ■ research tools: corpora, concordances, indexes and thesauri ■ full-text systems ■ natural language understanding, and ■ text processing and retrieval.

The conference will be followed immediately by the Association Internationale Bible et Informatique Second International Conference on June 9-13, 1988. More information about either conference can be obtained by contacting Hillel Weiss, Coordinator, Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel, 52100.

AN APPROACH TO MULTILINGUAL TEXTS

Christoph Gutknecht (University of Hamburg)
Frank Gerlinger (Hamburg)

1. Problem

The modern researcher who wishes to reach his optimal level of productivity can no longer ignore information-processing aids. Many word-processing programs that are presently available assist the author in the procedure of writing and editing letters, reports and manuscripts, because - to mention only the most fundamental advantages - they

a) allow the movement of parts of the text to different positions, the deletion of words, sentences, or paragraphs, and the insertion of new material wherever necessary, without the operator having to retype the whole page,

b) even check for correct spelling, grammar, and word usage, and

c) permit the setting and justification of margins and spacing for the printed version of the document.

Using a word-processor to type articles for scientific journals is ideal for the pure, monolingual text passages. Thus a linguist who wants to include the following passage into a handout for a course on English phonetics is not faced with any difficulties, when typing the text:

EXAMPLE 1:
[Monolingual handout without foreign characters or special symbols.]

"A good phonologist knows all about traditional phonology, structural phonology , generative phonology, and natural phonology. A not so good phonologist will know all about the one school in which he or she happened to be reared. Both, however realise the enormous complexity of the involved phenomena and of the rules which are needed to describe them. The ESL expert who faces the practical problem is likely to be turned off by the complexity and/or controversiality of the available linguistic information. However, since the problem does need to be solved anyhow, the unsatisfied customer will try to play it 'by ear', also in the literal sense of the

phrase, and solve it, once again, without a linguist and perhaps without linguistics. As a result, the ESL person may end up spending a considerable amount of time teaching a non-speaker of English the difference between the aspirated and unaspirated /p/ or between the pronunciation of the diphthong in *ride* and *right*."

(From: Victor Raskin, "Linguistic Applications in the 1980s", in: Peter C. Bjarkmann/Victor Raskin (eds.), *The Real-World Linguist: Linguistic Applications in the 1980s*, Norwood, NJ: Ablex, 1986, p. XV.)

Difficulties arise, however, when it comes to typing multilingual texts.

In particular those writers who use keyboards other than the US-ASCII-version[1] have to deal with the inflexibility of the standard fixed-character keyboard. Whenever they need another language or different symbols, they have to use an overlay on the keyboard or have to use different keyboards.

In the case where the system lacks specific symbols or formulae these have to be added manually to the finished document[2].

2. T^3 - a multifunctional word-processing system

With the help of the multifunctional word-processing system T^3 - especially in combination with the LCBoard - the writer of scientific texts is now in a position to solve most of his problems.

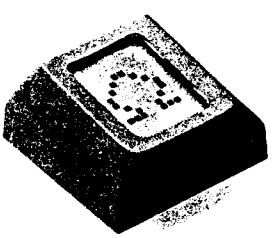
The T^3 not only provides all of the features that standard word-processors have but also offers a number of additional features superior to existing systems: (a) unlimited user-definable line and page formats, (b) macro-storage capacity not only for frequently used phrases, but also for entire command chains, (c) the ability to display all characters boldface or underlined or as a combination of the two, (d) easy-to-follow instructions because of combined menu and window technique, (e) footnote administration capability (numbering both headers and footers automatically) and (f) the possibility for multiple users to work on their own documents using the same word-processor.

In addition to these functions, the T^3 allows the keying-in and displaying of scientific texts, including possibilities for: (g) subscript and superscript of all characters of up to 25 levels, (h) display of complex formulae over a number of rows, and (i) display of integrals and roots made out of macros.

3. The "intelligent" keyboard: LCBoard (version PF 1.24 AN)

These features of the T^3 are supported by the "intelligent" LC-Board, so that it is now possible to optimise the use of the computer in both areas of display (keyboard as well as screen) and control and easily handle applications requiring access to multiple language and/or specialised character sets. These character sets, which can be freely defined by the user (and may include different alphabets such as e.g., Latin or Cyrillic), can be combined in a keyboard layout that is made up according to specific country standards or according to the user's individual preferences or requirements; thus it would, for instance, be easy to set up and use an ergonomic Dvorak-keyboard.

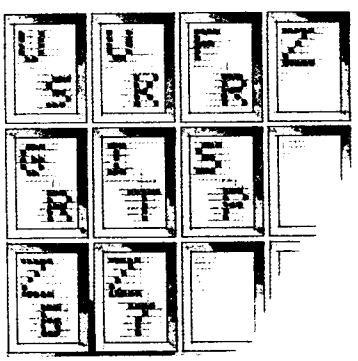
In contrast to conventional keyboards with fixed engraved or printed keys, the system-independent LCBoard has 88 LCD (liquid crystal display) keycaps (LCBlocs) that each have their individual screen - originally blank - for display and input. A wide variety of messages, commands or functions may be displayed on the LCBlocs - optionally blinking or inverse. The result is a much improved man-machine interface. Additional software allows the user to label the LCD in each keycap. The on-key microprocessor generates the chosen character.



The display of every key (LCBloc) consists of an 8 by 12 pixel-matrix or a 20 by 8 pixel-matrix. This allows the individual key to display either any single character out of a variety of characters or up to five characters simultaneously.

3.1 Character sets

The Z-key (character set key) on the numeric keypad switches into the character set selection mode when pressed. The following lay-outs are available by pressing the corresponding key:



US: ASCII characters;
UK: British character set;
FR: French character set;
GR: German character set;
IT: Italian character set;
SP: Spanish character set.

Illustration: keys

Additional character sets can be stored in the memory of the PC or workstation and can be downloaded to the keyboard and then activated with the character set selection mode. Thus, special characters (such as those used in mathematics or APL) or alphabets (such as Greek or Cyrillic) can be displayed on the LCBoard. However, to be able to take advantage of this possibility, a word-processor is required that can work with several alphabets simultaneously: for example, the T^3 permits the user to call an operations menu from the edit-mode and select an alphabet that has previously been defined.

3.2 String-processing

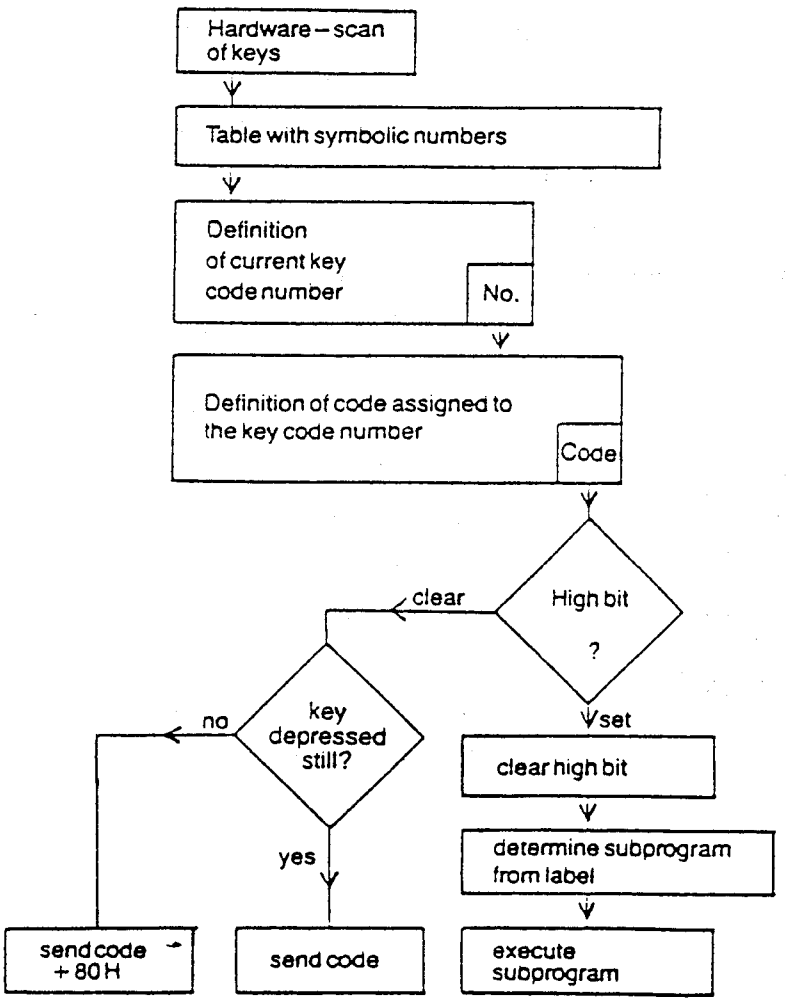
Another special key is the "string" key. A string of up to 126 key-strokes can be assigned to every key. With the help of the LCTOOL-

program a symbol can be created and made visible on each LCD-key.

The procedure for an assignment is to (1) press the "string" key twice (a *L*ight-*E*mitting *D*iode will blink!), (2) press the (target) key to which the string is to be assigned, (3) enter the string sequence, and (4) press the string key again. When the key has been programmed, the "string" can be used by pressing both the string key (LED will light!) and the target key once, whereby the string is called and passed to the host computer.

The programming and calling of strings can additionally be performed on the SHIFT-, CTRL- and ALT-level by keeping the appropriate key pressed down during the procedure: this offers the possibility of having four different strings on each key.

Float chart: string-processing



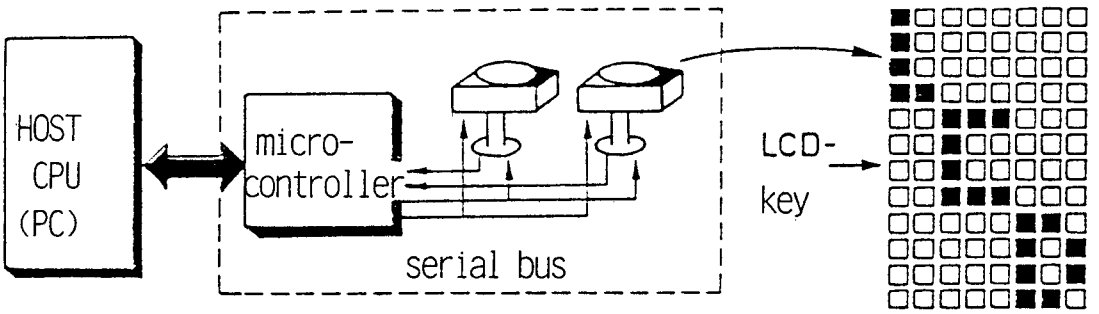


Illustration: LCBoard System Diagram

In addition to these 87 programmable keys there are 24 double-sized autostring-keys in the upper area of the LCBoard. These function-keys can display their respective functions at the levels of ALT, SHIFT, CTRL. In order to call the four different strings on these keys, one only needs to press the appropriate key once. For many frequently used software-packages, preprogrammed strings with LCD-display are already available on disk, thereby optimising the use of these software packages. A string may not only consist of repeatedly used phrases (a function which is of invaluable use for translators) but also of complex formatting procedures for text or graphics, or program sequences (macros), including procedures for teletex, modem, or plotter access. With the help of the LCBoard, the above mentioned T^3-procedure that allows the shifting from one alphabet to another can be activated by simply pressing the so-called "autostring key" once, thus giving access to the respective alphabets.

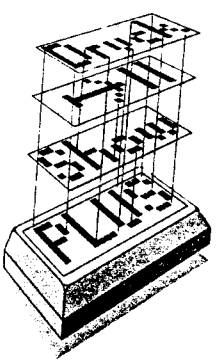
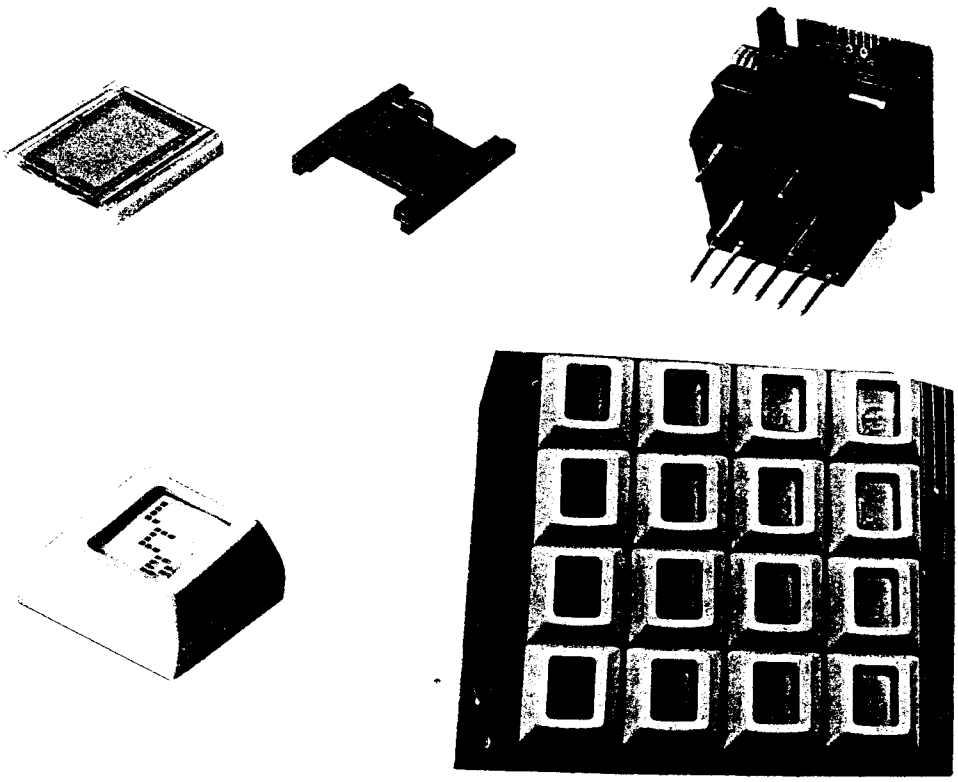
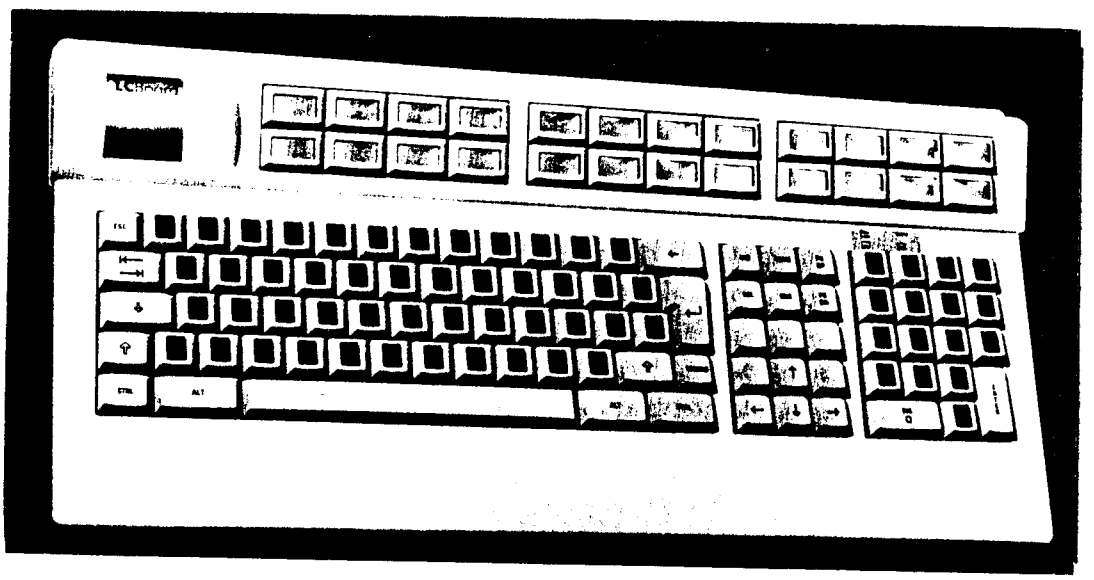


Illustration: four-level key

LCBloc with chip and integrated screen:





LCBoard (version PF 1.24 AN)

3.3 Generation of characters, character sets and whole fonts with the T^3 and LCBloc-keyboard

A great deal of the T^3's versatility is derived from its ability to display (on screen) and print many different characters. In combination with the LCD-keyboard (version PF 1.24 AN), characters of any kind can even be displayed simultaneously on the LCD-keycaps. Before discussing the technique of defining characters or letters, it is necessary to explain the following terms:

3.3.1 Character - Character Set - Font - Keyboard

Character: A character itself is an abstraction: such as the Latin letter "a" and the Rune-letter "ᚦ" ("thorn")[3]. A letter "a" will always be recognized as an "a," no matter whether it is handwritten, printed, or typeset in various ways. Thus the abstract conception of the letter "a" does not depend on the way in which it is presented.

Character set: An alphabet is a character set. In computer terms a character set usually includes the English alphanumerical characters (ASCII), special characters and various control characters.[4]

Font: A font is the realisation of a character set. In typesetting terms there are, among other fonts: *HELVETICA*, *ROMAN*, *MODERN*. Within the framework of the T^3 the different fonts refer mainly to the resolution of the dot-matrix.

Keyboard: In contrast to the hardware term "keyboard", in the T^3-system the term "keyboard" refers to a feature that can be generated by the software: an assignment of characters to actual keyboard keys. The keyboard is the tool that provides the characters which in turn can be typed into a document. The system itself only remembers a specific character (by its character set name and its position within the set) but not the software keyboard which was used for the typing of the character. Since the keyboards are completely independent from character sets or fonts, it is, for example, possible to have a keyboard which contains letters of the Greek, Russian or phonetic alphabets[5].

3.3.2 Examples of different character sets:

IBM

☺☻♥♦♣♠•◘○◙♂♀♪♫☼►◄↕‼¶§▬↨↑↓→←∟↔▲▼

@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_'

!"#$%&'()*+,-./0123456789:;<=>?

abcdefghijklmnopqrstuvwxyz{¦}~Δ

Cyrillic

№ «»

@ФИСВУАПРШОЛДЬТЩЗЙКЫЕГМЦЧНЯ[Х]^Ъ'

!Э#$%&Э()*+бъюё0123456789Ж=Ю=Ё

фисвуапршолдьтщзйкыегмцчня{Х}~

Phonetic

!"#$%*'()*ɜː,-./,eɪaɪɔɪəʊaʊɪəeəʊə :;<=>?

ɑːbcdefghiːjklmnɔːpqrstuːvwxðzæˈː

ʌ dʒə ŋ ɪ ˌɪmˌnʊ ʃtʃʊɷ θʒ[\] _'

Mathematical

∂∫ϕ⌋⌈⊥⌉ϕ±÷≠θℵϵρτ⌊⌋ιωπ∫χαστϒγηΨκ

—⊕□≈∃√⊗∉∈∞Ω||∀ΣΔ⌉Γ⌈Ψ₃Λℂºˈˌ⌋ℓ⌈t₁₂∇

λℝ ≃⌜ζξχ▽βνμ∠¬·→↑←↓⊃∩⊂∪≅≡

⌊≤≥⌋

3.3.3 Self-defined characters

The actual construction of self-defined characters is done from the operations menu by means of the font generator. Each character set has a number of versions (fonts): LCD-keycap versions, screen versions and versions for each of the printers supported. The minimum working configuration consists of three different fonts of a character set, e.g. the 8 by 12 font for the LCD-keycaps, the 8 by 16 cell screen font for use with the Enhanced Graphics Adapter and Hercules Monochrome Graphics Card, and the 30 by 50 graphics font for a laser printer or 24 nozzle ink-jet printer. For consistency, all fonts that are versions of a specific character set are given the same character set name. So the character set "Cyrillic" will always contain the same symbols in the same position. Additionally, every font group is identified by a number called the font ID. As examples, fonts for the Hercules screen font have ID 2, and fonts for Epson graphics 30 by 50 printing have ID 2701. The system provides the possibility to change every character and to create new characters and character sets.

3.3.4 Revising a character

The font revisions screen of the system's character generator displays four rows of characters at the bottom of the screen, arranged in rows of two. The upper row in each case is a reference row. It is a guide to the position in the font. When revising a printer font, these rows show the corresponding characters in the screen font for the display with the same name of the printer font being revised. It is possible to select a particular character. Pressing the Revise key (F2) modifies the character currently selected and displays it in the currently selected font (in the appropriate resolution) at the top of the screen. By moving the cursor around with the arrow keys it is possible to turn the squares on and off with the Highlight key (F1). There are various operations available when revising a character (see below). Once the revised font is saved, the appropriate character set is ready to be printed in the revised version.