FROM:      J.H. Saltzer

DATE:      November 29, 1965

SUBJ:      7094 Character Set

## Background

This note is intended to supply source material to CTSS historians,
and to guide system programmers who may have occasion to make   changes
to the CTSS character code tables.

Up to the present time, the only standard character set on CTSS has
been the 7094 BCD code extended by a few carriage control characters.
Although an extended (12-bit) code character set has been used on
CTSS, no standard meanings for the extended codes have been adopted
because of the variety of typewriter devices and character sets which
have been attached to the system.

Recently, however, a number of things have happened to make an extended
standard character set desirable, to wit:

1.  The popularity of the TYPSET and RUNOFF commands has resulted
    in a large quantity of stored data based on the 1050 character
    set.  It seems worthwhile to protect our user's investment in
    these commands.

2.  The Multics programming effort will begin with programs typed
    in on the 7094 using the extended character set.  Again, the
    investment in this effort must be protected.

3.  A special "Project MAC" 1050 printing ball has been ordered;
    it has a significant character overlap with the standard Model
    35 teletype character set.  Thus now is the moment to standard-
    ize, if ever.

4.  The proposed revision of the ASCII character set contains a
    set of graphics which largely overlaps with the CTSS graphics
    (although using different codes) and we may expect that future
    typewriter devices are likely to use this graphic set.

For these reasons, the latest revision of the CTSS Programmers Guide
Section AC.2.01 describes a "CTSS character set" of 112 characters,
including 90 printing graphics, space and 21 control characters.  It
then goes on to list, device by device, transliteration ("stylization")
rules from the standard CTSS character set to the specific keyboards
now in use.  An astute reader of the tables of Section AC.2.01 will
observe some highly unusual conventions;  these conventions are all
based on the planned "MAC golf-ball" installation.

### The "Project MAC Golf-Ball"

An overriding consideration in the planning of the change-over to a
different golf ball is that memos and program files typed in before
the change-over must still type out correctly after the change. Memos
(or programs) containing characters missing from the "MAC" ball but on
the 1050 "correspondence" ball must type out correctly if a "correspondence"
ball is temporarily mounted on the machine.

This requirement automatically establishes a certain amount of inflexi-
bility in the character set:  the 86 codes which correspond to the 86
1050 "correspondence" graphics must remain in the character set before
and after the change, although the graphic _meaning_ of the codes may
change.

Following are the differences between the "Correspondence" and "MAC"
golf balls:

| CTSS Code | "Correspondence" Graphic | "MAC" Graphic |
|-----------|--------------------------|---------------|
| 105 | % | ¬ |
| 153 | ! | [ |
| 101 | ± | ] |
| 102 | � | X ^ |
| 100 | ¢ | X | |
| 155 | . (upper) | > |
| 152 | , (upper) | < |

It will be seen by inspecting the CTSS character set table that the
"MAC" Ball graphic has been adopted as the CTSS standard in each case,
while the transliteration of these graphics on the "Correspondence"
ball is indicated in the 1050 exception table.

When the MAC golf ball is installed, two trivial changes must be made
to the 1050 code tables:  at present on _input_ the upper case period
and upper case comma map into their lower case equivalents, rather than
the pointed brackets. This is a temporary measure to prevent confusion.
The input codes for these two graphics must be changed.

An unusual feature of the CTSS character set is that although there is
a "percent sign" (0112) and an "exclamation point" (0114) in the set,
the 1050 ("Correspondence" ball) percent sign and exclamation point
are really stylizations of the "logical not" (0105) and "close square
bracket" (0153) characters.

Again, to prevent confusion temporarily, the teletype percent sign is also considered a stylization of the "logical not" and produces code 105. When the MAC golf ball is installed, the possibility of confusion will vanish and the teletype percent sign should produce a CTSS percent sign code, 0112.

Here, then is a summary of code table changes needed for installation of the new Project MAC golf ball:

    1050 table:    on _input_ the upper case period becomes >
                   on _input_ the upper case comma becomes <

    Model 35 table:  % key should produce %code, 0112
                     instead of ⌐ code, 0105

At the same time, the 1050 and Model 35 exception tables in section AC.2.01 should be revised:  a Xerox copy of the revised 1050 exception table is attached.

## Phase-out of Old MIT-modified Model 35 Teletypes

When the phase-out is completed, section AC.2.01 should be re-issued with the corresponding page deleted. One character code table may also be deleted from the supervisor. (Or changed for the model 37 teletype.)

## CTSS Revised ASCII Mapping

Since there is now a standard CTSS character set, it seems certain that at some time in the future there will be a desire to perform a mapping of the CTSS graphics to and from the revised ASCII set. Since the two sets of graphics largely overlap, we can establish simple transliteration conventions. The following mapping table is proposed for three graphics which are not identical in the two sets.

| ASCII graphic | CTSS graphic | CTSS code |
|---|---|---|
| ∧ (circumflex) | \ (left slash) | 102 |
| | (vertical bar) | ↑ | 100 |
| ´ (acute accent) | " (apostrophe) | 014 |

(This is the same mapping table which has been proposed for Multics use of the 1050.)

This mapping establishes meanings for all CTSS graphics, but leaves
four ASCII graphics without any CTSS equivalents. We may define the
following new CTSS codes to correspond to these graphics.

| CTSS code | Graphic |
|-----------|---------|
| 160 | ` (grave accent) |
| 134 | ~ (tilde) |
| 173 | { |
| 174 | } |

With these conventions it is now possible to attach a model 37 teletype
to the 7094, and to provide a complete and simple CTSS-to-ASCII and
back conversion program for certain Multics applications.

075          HCF
076          HLR

1050 With MAC BALL - Same as CTSS Character Set Except as noted below:

|        | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------|---|---|---|---|---|---|---|---|
| 0000   |   |   |   |   |   |   |   |   |
| 0010   |   |   |   |   |   |   |   |   |
| 0020   |   |   |   |   |   |   |   |   |
| 0030   |   |   |   |   |   |   |   |   |
| 0040   |   |   |   |   |   |   |   |   |
| 0050   |   |   | ig |   |   |   |   |   |
| 0060   |   |   |   |   |   |   |   |   |
| 0070   |   |   |   |   |   |   |   |   |
| 0100   |   |   |   |   |   |   |   |   |
| 0110   |   |   | ig | ig | ig | ig |   |   |
| 0120   |   |   |   |   |   |   |   |   |
| 0130   |   |   |   |   |   |   | ig |   |
| 0140   |   |   |   |   |   |   |   |   |
| 0150   |   |   |   |   |   |   |   |   |
| 0160   |   | ig |   |   |   |   |   |   |
| 0170   |   |   | ig |   |   |   | ig | ig |

Notes:  1.  This table assumes a standard "MAC" ball.  Other balls
will print other graphics.

2.  Code 0017 (Interrupt) can be generated by one push of
the "Reset line" button.

3.  Code 0057 (Quit) can be generated by two pushes of the
"Reset line" button.

(This table is for publication after Project MAC type ball is installed.)