

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

---

ECMA STANDARD

for

PRINTING SPECIFICATIONS FOR  
OPTICAL CHARACTER RECOGNITION

Mai 1968



Free copies of this standard **ECMA-15** are available from  
ECMA European Computer Manufacturers Association  
114, rue du Rhône - 1204 Geneva (Switzerland)

TABLE of CONTENT

	Page
1. Introduction	2
Purpose	2
Scope	2
Interpretation	2
Use	3
2. Spectral Requirements	3
General	3
Spectral Bands	3
3. Paper Specifications	4
General	4
Paper Reflectance	5
Average Reflectance	5
Variation in Reflectance	6
Dirt in Paper	7
Paper Opacity	7
Definition	7
Measurements	7
Classes	7
4. Characteristics of the Printed Image	7
General	7
Measuring Gauge	9
Mean Stroke Edge	9
Centreline Deviations	10
Strokewidth	11
Edge Irregularities	13
PCS	14
Definition	14
Measurements	14
Spots	16
Definition	16
PCS Measurement	16
Visual Measurement	16
Voids	17
Definition	17
PCS Measurement	18
Visual Measurement	18
Maximum PCS	19
PCS of Spots and Voids	19
PCS Voids and maximum PCS	19
Average PCS	19
Skew	19

	Page
5. Character Positioning	19
Introduction	19
Document Reference Edge	19
Clear Area	20
Printing Area	20
Margin	20
Field	21
Field Boundary	21
Line Spacing	21
Line Separation	22
Character Boundary	22
Character Reference Lines	23
Character Spacing	24
Character Separation	25
Character Misalignment	25
APPENDIX	
A 1 Paper Characteristics and Measurements	29
Spectral Properties	29
Uniformity of Paper Reflectance	30
Paper Opacity	32
Paper Gloss	33
Dirt in Paper	33
Mechanical Properties of Paper	33
A 2 Characteristics of the Printed Image	33
General	33
Rules for Design of measuring Gauges	34
Centreline Deviations	35
Strokewidth Ranges	36
PCS and visual Measurements	36
Spectral Bands for PCS	36
PCS of Spots and Voids	37
Spots remote from a Character	40
Lower-case OCR-B Characters	40
A 3 Character Positioning	40
Objectives	40
Document Reference Edge	40
Clear Area, Printing Area, Margin	41
Fields	42
Line Spacing	42
Line Separation	42
Character Boundary and Reference Lines	43
Character Spacing	43
Character Separation	43
Character Misalignment	44
A 4 General Illustration of the Characters	45
OCR-A	45
OCR-B	46



## BRIEF HISTORY

ECMA TC 4 started their standardization work in the field of Optical Character Recognition in June 1961 and developed two fonts, a stylized (Class A) numeric character set, Standard ECMA-8, and a conventional one (Class B), Standard ECMA-11.

This Standard, ECMA-15, is directed to the printing specifications required for both OCR-A and OCR-B. It has been adopted by the General Assembly of ECMA in May 1968.

Representatives of following Companies participated in the work of the Committee :

A.B. Dick  
AEG-Telefunken  
CII, Compagnie Internationale pour l'Informatique  
Compagnie des Machines Bull  
N.V. Electrologica  
English Electric Computers Ltd  
Facit Electronics AB  
Honeywell EDP Europe  
IBM-Europe  
ICT, International Computers and Tabulators Ltd  
ITT Europe  
Lamson Paragon Ltd  
NCR, The National Cash Register Company Ltd  
Ing. C. Olivetti & Co. S.p.A.  
Siemens Aktiengesellschaft  
UNIVAC Computers (Europe) Ltd  
Wiggins Teape Research & Development Ltd

In addition this Technical Committee has collaborated with the following organizations :

International Organization for Standardization (ISO)  
United States of America Standards Institute (USASI)

## 1. INTRODUCTION

### 1.1 PURPOSE

The purpose of this standard is to establish the basis for industry standards for paper and printing to be used in Optical Character Recognition (OCR) systems, and to aid in the implementation and use of such systems. It provides for the identification and measurement of and establishes specifications for the relevant parameters and gives guidance for their use.

### 1.2 SCOPE

This standard contains the basic definitions, measurement requirements, specifications and recommendations for OCR paper and print.

It applies to the ISO R... (ISO D.R. 996). Three major parameters of a printed document for OCR media are covered.

These are :

- a. The optical properties of the paper to be used.
- b. The optical and dimensional properties of the ink patterns forming OCR characters.
- c. The basic requirements related to the position of OCR characters on the paper.

The major factors of each of these areas pertinent to OCR are identified. Definitions of these items are given and bases for measurements are established. Basic specifications applicable to all OCR materials are imposed and recommendations for the implementation of an OCR system are made.

Because of the widely divergent nature of OCR applications this standard does not include all of the necessary or prudent specifications or considerations that may be necessary for a successful OCR system.

Additional restrictions will often need to be imposed and additional pertinent variables will need identification and control. Such items as document size, the mechanical properties of the paper, the degree of control necessary over possible variations and the format details of the particular application should be resolved by those concerned.

### 1.3 INTERPRETATION OF THE STANDARD

The values in this standard represent the specifications for supplies and the limit of performance for a printing



system to be used for the preparation of OCR media. They are established on the basis that they are reasonably obtainable. However, it must be recognized that many parameters are subject to variations of a statistical nature, and deviations from the specified limits may occur.

The degree to which these deviations are allowed (in cases where the specification is not already expressed in statistical terms) will depend upon the specific application and should be evaluated by the users and suppliers before a system is to be established.

Furthermore, although the limit of each parameter is given as an independent variable, a deterioration in reader performance is likely if the limits of more than one parameter are approached simultaneously. Every effort should be made to keep well within the limits.

It is not unknown for there to be a deterioration in print quality during the time which elapses between printing and OCR processing. Such changes are difficult to measure and the standard makes no distinction between the state of OCR material immediately after printing and the state immediately before reading.

#### 1.4 Use of the Standard

In using and referencing the Standard in any particular application it is necessary to specify the selection between a number of choices so that the proper portions of the Standard can be applied. These choices are selection of the font, the font size, the character repertoire, the spectral characteristics of the paper and printed images, the paper opacity, and the strokewidth tolerances.

## 2. SPECTRAL REQUIREMENTS

### 2.1 General

This section contains the definition of spectral bands of interest for OCR applications.

They must be defined since character readers operate in specific spectral regions and paper and ink characteristics change with the wavelength considered.

### 2.2 Spectral Bands

In this section a set of bands is defined as reference for the paper and printed image specification. Their use and the measuring procedures are specified in the sections Paper Reflectance (3.2), Paper Opacity (3.4) and PCS Measurement (4.7.2).

BAND	PEAK in nm	BANDWIDTH in nm, 50 % level
B 400	see below	
B 425	425 ± 5	50 or less
B 460	460 ± 5	60 or less
B 490	490 ± 5	60 or less
B 530	530 ± 5	60 or less
B 570	570 ± 10	100 or less
B 620	620 ± 10	100 or less
B 680	680 ± 10	120 or less
B 900	900 ± 50	400 or less

The bands B 425 up to B 900 represent the spectral responses required from the complete measuring instrument (light-source, filter, detector). These responses must be smooth curves without secondary peaks and with no major parts of the response curve beyond the specified 50 % points. The energy content of the illumination at wavelengths shorter than 400 nm should not exceed 5 % of that in the particular band under consideration.

The shortest wavelength band, B 400, is defined somewhat differently (see A.1.1.3), as follows :  
The light source must have a peak output at 400 ± 10 nm with a bandwidth of not more than 60 nm. The detector must have a uniform response (not less than 75 % of the peak response) over the range 365 - 500 nm.

### 3. PAPER SPECIFICATIONS FOR OCR

#### 3.1 General

The papers to be used in OCR applications should be white (see A.1.1.2), have a flat finish and low gloss and should be of high opacity.  
Fluorescent additives should be avoided. Paper for OCR should also be free from watermarks and coloured patterns.



### 3.2 Paper Reflectance

The measurements in this section deal only with diffuse reflectance. The reflected light used for measurement shall exclude specularly reflected light.

Unless otherwise specified, all reflectance values are referred to Magnesium Oxide, MgO, as the primary white standard. The reflectance of MgO is the 100 % value. Absence of any light of the wavelengths of interest is the 0 % value.

The average paper reflectance measurements shall be measured using the infinite pad method, i.e. the samples being measured must have a backing of a sufficient number of paper thicknesses of the same type of paper such that doubling the number will not change the measured value of reflectance.

The variation in paper reflectance shall be measured using the black backing method, i.e. the sample being measured must be backed with black of not more than 0,5 % reflectance.

#### 3.2.1 Average Reflectance

##### 3.2.1.1 Measurement Area

Each measurement of average reflectance shall be made over areas of at least 65 mm<sup>2</sup> (0,1 square inches). The area will be in the shape of a circle or of a regular polygon.

##### 3.2.1.2 Visual Spectrum

The average reflectance of the paper shall not be less than 60 % in the range from 425 nm to 500 nm and shall not be less than 70 % in the range from 500 nm to 700 nm.

Average reflectance may be determined either by means of spectro-photometric measurements or by a number of reflectance measurements in different spectral bands.

For white papers and slightly but uniformly coloured papers it is sufficient to measure the reflectance in the two following spectral bands :

- B 425
- B 530 or B 570 or any band peaking in between and having a band width smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution also referred to as "photopic luminosity function" satisfies this requirement.)

In doubtful cases where these two band measurements may not establish the required reflectance throughout the whole range, it is necessary to make reflectance measurements in a greater number of bands.

The following set of bands may be used for the purpose :

B 425, B 460, B 490, B 530, B 570, B 620, B 680.

Any other choice of bands may be employed provided they adequately cover the visible spectrum.

When the near infra-red (IR) spectrum is of interest, an average reflectance of 70 % in the band B 900 is required. Since white and slightly coloured papers which meet the previous specifications will usually present an average reflectance greater than 70 % in the near infra-red spectrum, reflectance measurements in this band usually are not necessary.

In cases where the near ultra-violet (UV) spectrum is considered the average reflectance should be greater than 55 % when measured in the B 400 band. White papers will usually meet this requirement.

### 3.2.2 Variation in Paper Reflectance

Variation in paper reflectance is defined as the standard deviation of reflectance measurements, taken over well separated circular areas of diameter 0,2 mm (0,008") : see A.1.2.

Two classes of variations in Paper Reflectance are specified :

- standard deviation  $\leq$  5 % of the mean reflectance (for medium opacity paper: see 3.4.3).
- standard deviation  $\leq$  3,5 % of the mean reflectance (for high opacity paper: see 3.4.3).

The specification on variation in paper reflectance must be satisfied in the following bands :

- B 425
- B 530 or B 570 or any band peaking in between and having a band width smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement.)
- B 900



In practice the measurements may usually be limited to the most critical band.

In doubtful cases where a single band measurement may not be sufficient to show that the specification is satisfied throughout the whole spectrum it is necessary to use the three bands.

### 3.3 Dirt in Paper

The dirt count in paper may not exceed 10 parts per million as determined by TAPPI (Technical Association for the Pulp and Paper Industry, 360 Lexington Avenue, New York, N.Y., USA) method T 437 - ts - 63.

All foreign material  $0,01 \text{ mm}^2$  (0,000012 square inches) and larger shall be counted.

### 3.4 Paper Opacity (see also A.1.3)

#### 3.4.1 Definition of Paper Opacity

Opacity (Paper Backing) is the ratio (expressed as a percentage) of the average reflectance of a specimen backed with black of not more than 0,5 % reflectance, to the average reflectance of the same specimen backed with an infinite pad.

#### 3.4.2 Measurement of Paper Opacity

Paper opacity shall be measured using B 530 or B 570 or any band peaking in between and having a band width smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement.) In choosing the class of paper opacity it is important that the recommendations given in A.1.3.2 be considered.

#### 3.4.3 Classes of Opacity

Papers acceptable for OCR fall into two classes, based on opacity.

##### 3.4.3.1 High Opacity Paper

High opacity paper has an opacity of not less than 85 %.

##### 3.4.3.2 Medium Opacity Paper

Medium opacity paper has an opacity of at least 65 % but less than 85 %.

## 4. CHARACTERISTICS OF THE PRINTED IMAGE

### 4.1 General

This section contains specifications and quality control criteria pertaining to individual OCR characters and marks, i.e. without consideration of the relationship between the

individual printed image of an OCR character and any other printing on a document. Relevant specifications for the latter are contained in section 5.

The specifications in sections 4 and 5 pertain to printed images and not to type faces.

The performance of OCR systems depends to a large extent on the print quality. Hence, every effort should be made to provide "good" print quality. i.e.:

The printed character should present as high a contrast as possible to the background document.

Strokewidths should be held as close as possible to the nominal.

There should be no voids within the stroke outline. When this cannot be prevented the number and size of individual voids should be minimized and the distance between them should be as great as possible.

There should be no extraneous ink within the clear area. When this cannot be prevented the number and size of individual spots should be minimized and the distance between them should be as great as possible.

The mean shape centreline of the printed image should be held as close as possible to the nominal. Since variations can seriously affect reading performance, type designers and print device manufacturers are cautioned to take care and use techniques which produce printed images conforming to this recommendation.

In order to achieve the print quality required for OCR it should be understood that in comparison with non OCR applications special precautions, including adjustments, maintenance, etc., may have to be taken and the ribbon life of impact printers will usually be shortened.

The reflectance specifications in this section deal only with diffuse reflectance and the reflected light used for measurement shall exclude specularly reflected light. Unless otherwise specified, all reflectance values are referred to Magnesium Oxide, MgO, as the primary white standard. The reflectance of MgO is the 100 % value. Absence of any light of the wavelength of interest is the 0 % value.



The reflectance measurements shall be made using the infinite pad method, unless otherwise specified. The sample being measured must have a backing of a sufficient number of paper thicknesses of the same type so that doubling that number will not change the measured value of reflectance. There should be a good understanding of the spectral properties of the ink, paper and the OCR scanners used. Where the spectral properties of the reader are not known, it is recommended that inks with a high absorption in all bands, from B400 to B900 inclusive, be used, e.g. carbon black pigment inks. However, it should be recognized that certain printers cannot use carbon black inks, and therefore their printing will have a high absorption only in the visual spectrum.

The requirement of human legibility imposes the use of inks which have good absorption in the visible range even where near IR and near UV spectrums are used for machine reading.

#### 4.2 Measuring Gauge

Printed images are measured using a gauge showing the minimum and maximum Character Outline Limits (COL).

These limits are constructed by superimposing the minimum and maximum stroke widths, as specified below, symmetrically about each point on the character centreline drawing. (See A.2.2)

Before making measurements, the printed image must be visually aligned to give "best fit" with the gauge.

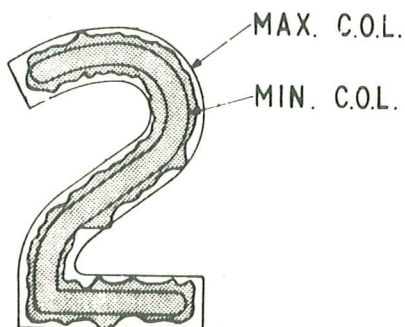
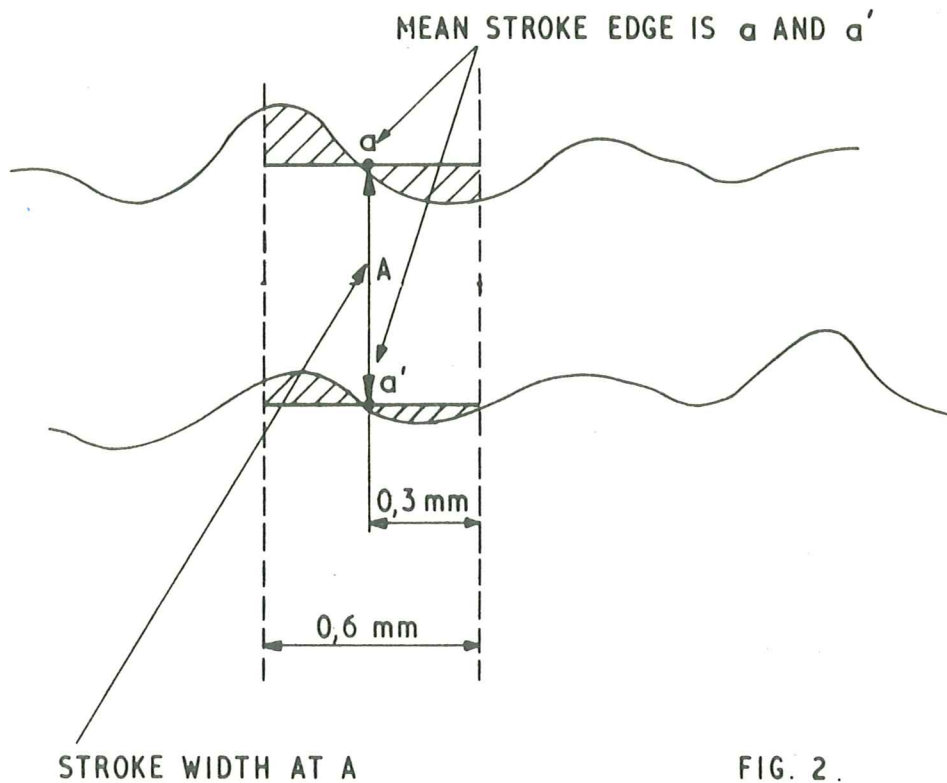


FIG. 1.

#### GAUGE IN ITS "BEST FIT" POSITION

#### 4.3 Mean Stroke Edge

Mean stroke edge is defined as the integrated average of the edge irregularities estimated visually along any length of 0,6 mm (0,024"), parallel to the COL. Mean stroke edges of a character must be contained between maximum and minimum COL.

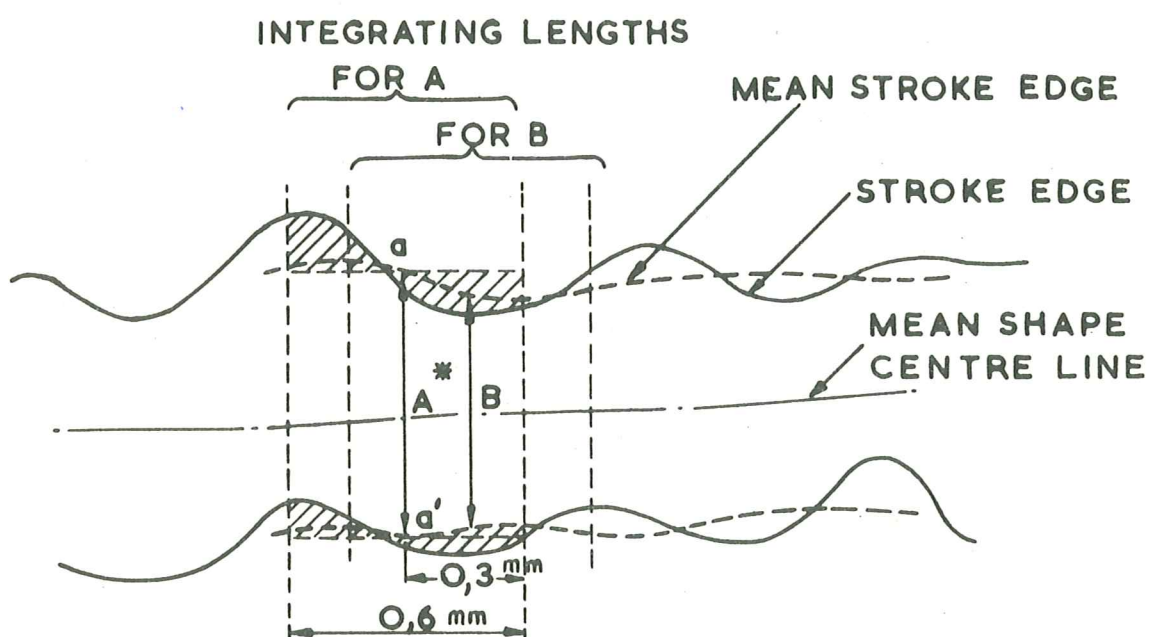


#### 4.4 Centreline Deviations

The assembly of the smoothed centrelines of the actual printed strokes is called the mean shape centreline.

The distance between any two points on the mean shape centreline of the printed image shall not differ by more than 0,075 mm (0,003") from the nominal distance between the equivalent two points on the ideal centreline.



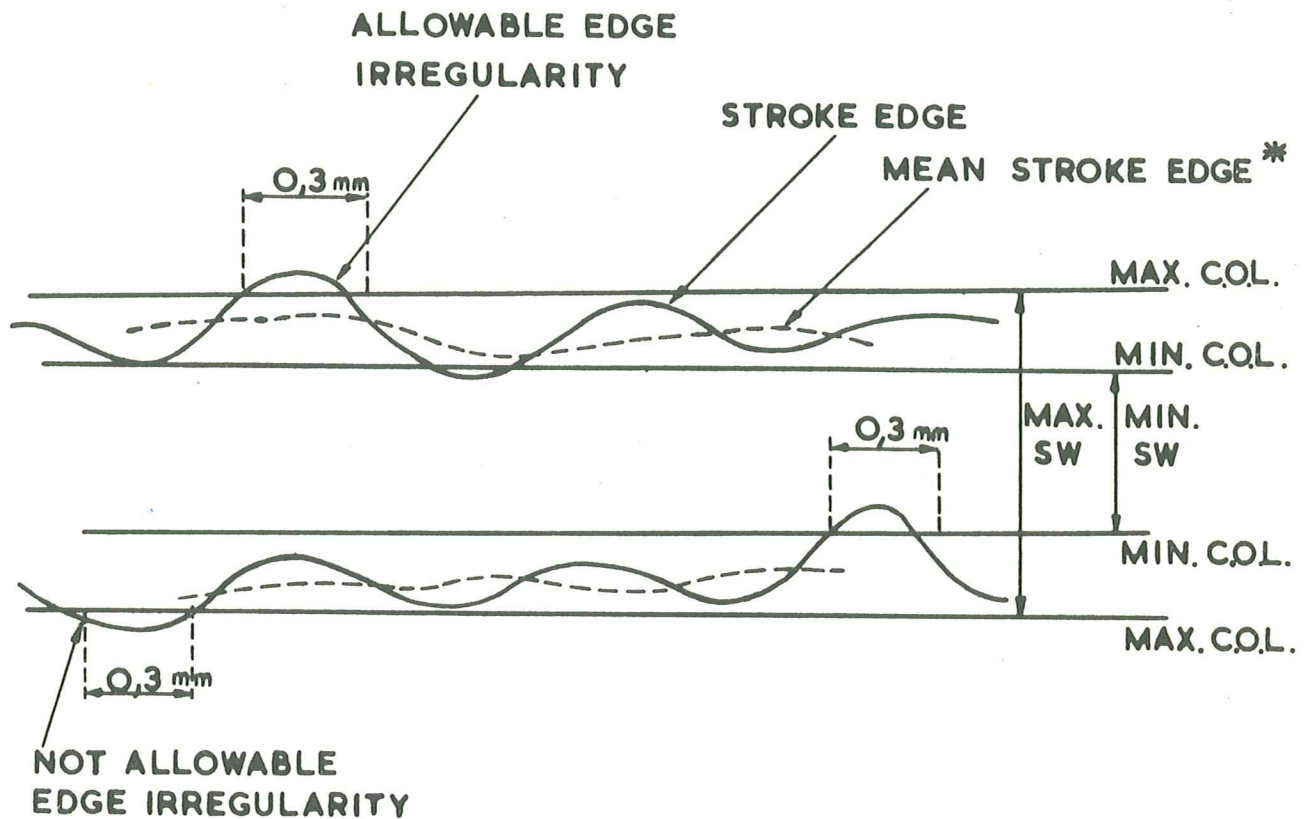


\*  
STROKE WIDTH IS A, B,....

FIG. 3

#### 4.5 Strokewidth

Strokewidth is the distance between mean stroke edges measured perpendicular to the mean shape centreline. The variations in strokewidth to be expected in the normal printed output will differ according to the type of printing device employed.



\*  
THE MEAN STROKE EDGE IS BETWEEN  
MAX AND MIN C.O.L.

FIG. 4

Two ranges of strokewidth can be identified, a small range X and a larger range Y. Range X can be tolerated without significant deterioration in reading performance. As strokewidth extends beyond range Y, the reader performance may degrade rapidly. It is expected that most printing devices for OCR will produce average printing with strokewidths in a range larger than X but smaller than Y. On the other hand, there are specially controlled printing devices and printing processes which can conveniently and economically produce average printing with strokewidth within range X. The nominal strokewidths and the tolerances about them are :



OCR - A

Size	Nominal	Range X	Range Y
I	0,35 mm (0,014")	± 0,08 mm (0,003")	± 0,15 mm (0,006")
II	0,35 mm (0,014")	± 0,08 mm (0,003")	± 0,15 mm (0,006")
III	0,38 mm (0,015")	± 0,08 mm (0,003")	± 0,18 mm (0,007")
IV	0,51 mm (0,020")	± 0,13 mm (0,005")	± 0,25 mm (0,010")

OCR - B

Size	Nominal	Range X	Range Y
I	0,35 mm (0,014")	± 0,08 mm (0,003")	± 0,15 mm (0,006")
	*0,31 mm (0,012")	± 0,08 mm (0,003")	+ 0,19 mm (0,008") - 0,11 mm (0,004")
II	0,35 mm (0,014")	± 0,08 mm (0,003")	± 0,15 mm (0,006")
	*0,31 mm (0,012")	± 0,08 mm (0,003")	+ 0,19 mm (0,008") - 0,11 mm (0,004")
III	0,38 mm (0,015")	± 0,08 mm (0,003")	± 0,18 mm (0,007")
	*0,34 mm (0,013")	± 0,08 mm (0,003")	+ 0,22 mm (0,0085") - 0,14 mm (0,0055")

For numeric applications the tolerance range for size III, range Y, may have to be widened. The following limits shall not be exceeded :

$$\begin{aligned} &+ 0,28 \text{ mm (0,011")} \\ &- 0,18 \text{ mm (0,007")} \end{aligned}$$

and the distance between the mean edges of parallel adjacent strokes shall not be less than 0,2 mm (0,008").

- \*) The strokewidth tolerances given in the table apply only to the following characters among the set of characters having nominal strokewidth 0,31 mm (0,012") for sizes I and II or 0,34 mm (0,013") for size III :

£ \$ : ; < % > ? [ ! # & , ] ( = ) \_

For the remaining characters of nominal strokewidth 0,31 mm (0,012") for sizes I and II, or 0,34 mm (0,013") for size III see Appendix A.2.9.

4.6 Edge Irregularities

Any extension of the stroke edge outside the maximum COL

should not exceed 0,3 mm (0,012"), measured visually along the maximum COL (see Fig. 4). Any extension of the stroke edge inside the minimum COL should not exceed 0,3 mm (0,012"), measured visually along the minimum COL. Edge irregularities must also meet the specifications on spots and voids (section 4.8 and 4.9).

#### 4.7 Print Contrast Signal

The contrast between a printed image and the paper on which it is printed is described by means of the Print Contrast Signal (PCS).

##### 4.7.1 Definition of PCS

The PCS is defined by the equation :

$$PCS_p = \frac{R_w - R_p}{R_w}$$

where:

$R_w$  = the maximum reflectance found within the area of interest to which the PCS of point p is referenced. (In measuring printed images, this area of interest should be a rectangle approximately twice the nominal character height by twice the nominal character width and centred on the character being measured.)

$R_p$  = the reflectance at p.

The reflectance  $R_w$  and  $R_p$  are measured within a circular area 0,2 mm (0,008") diameter.

The PCS requirements that apply to the printed image are stated in 4.11, 4.12, and 4.13.

##### 4.7.2 Measurement of PCS

The specification for PCS must be met in one or more of the following bands :

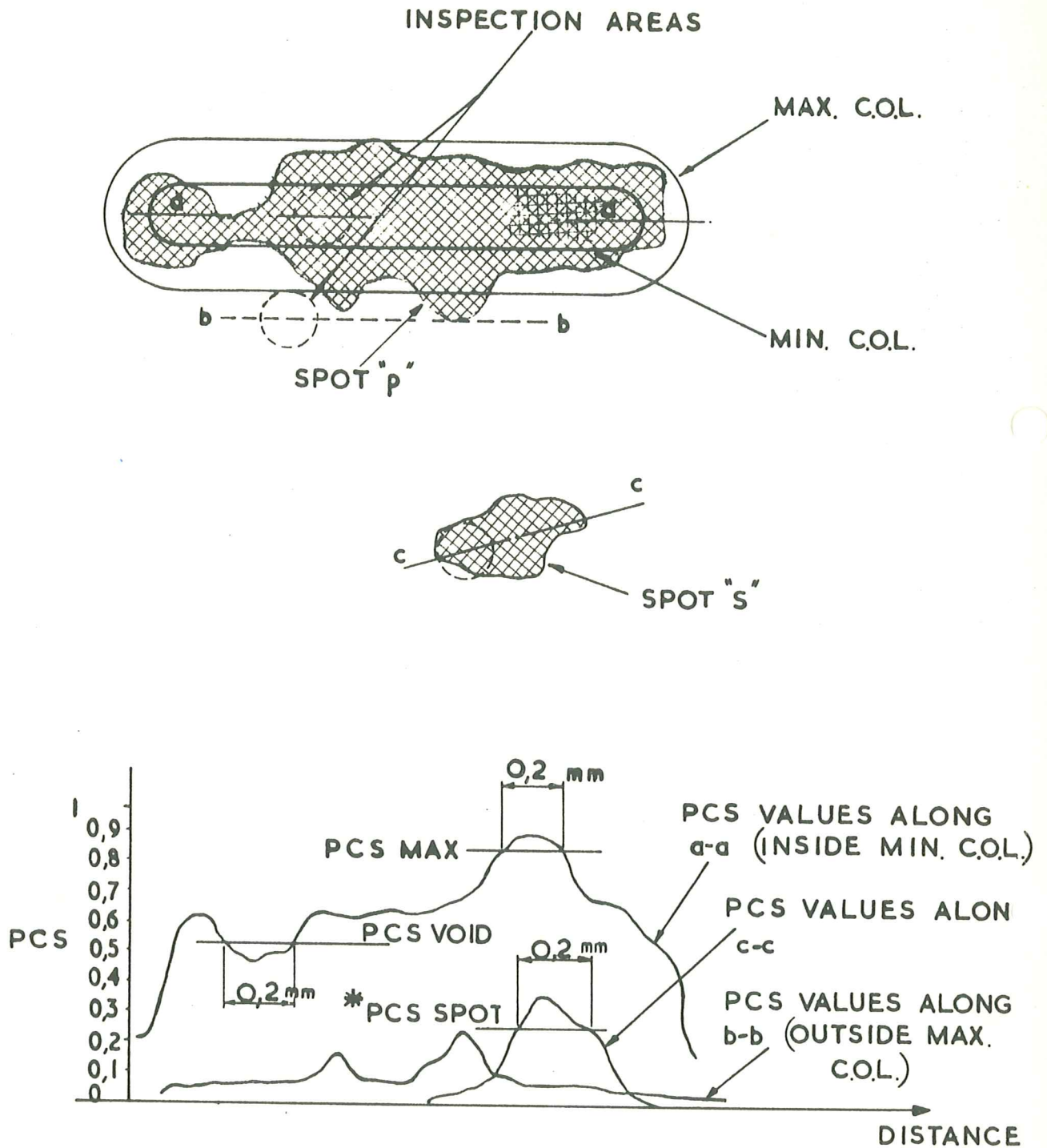
Near ultra-violet: B 425

Visual : B 530 or B 570 or any other band peaking in between and having a bandwidth not exceeding 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement.)

Near infra-red : B 900

The particular band(s) chosen will depend upon the characteristics of the reading and printing equipment in the system. It is important that the recommendations given in A.2.6 be considered.





\* IF SPOT "s" WERE NOT PRESENT THE PCS SPOT LEVEL WOULD BE DETERMINED BY SPOT "p". THE CHARACTER REPRESENTED IS A MINUS SIGN.

FIG. 5

## 4.8 Spots

### 4.8.1 Definition

Spots are defined as areas outside any maximum COL, which can be identified as contrasting with the background. They can be measured either visually or in terms of PCS. Both methods constitute an acceptable procedure for measurement of the printed image for most printing. However, where there is a substantial question of subjective visual judgment the instrumental method based on PCS must be used.

Character associated spots are those spots within a rectangle, twice the character height and twice the character width, centered on the character. For adjacent characters, spots should be related to the nearest character.

Spots remote from characters are discussed in App. 2.8.

### 4.8.2 PCS Measurements of Spots

Sizes of spots depend on the PCS level at which they are measured. For a specific (see A.2.7.1) PCS level, a spot is described as "Allowable" if it satisfies the following conditions :

1. The distance the PCS measuring aperture can be moved in a straight line, so as to give a PCS above the chosen level, nowhere exceeds 0,2 mm (0,008"). The measuring aperture must remain at all times outside the maximum COL.
2. The distance, centre to centre, of the spot from the nearest other spot, detected at the same PCS level, is at least 1,0 mm (0,040").

"PCS spots" is defined as the minimum PCS level at which all character associated spots are allowable (see Fig. 5).

### 4.8.3 Visual Measurement of Spots

A spot is allowable if it satisfies the following conditions :

1. It can be contained entirely within a circle of 0,2 mm (0,008") diameter, estimated visually.
2. Its distance centre to centre from the nearest other spot is at least 1,0 mm (0,040").



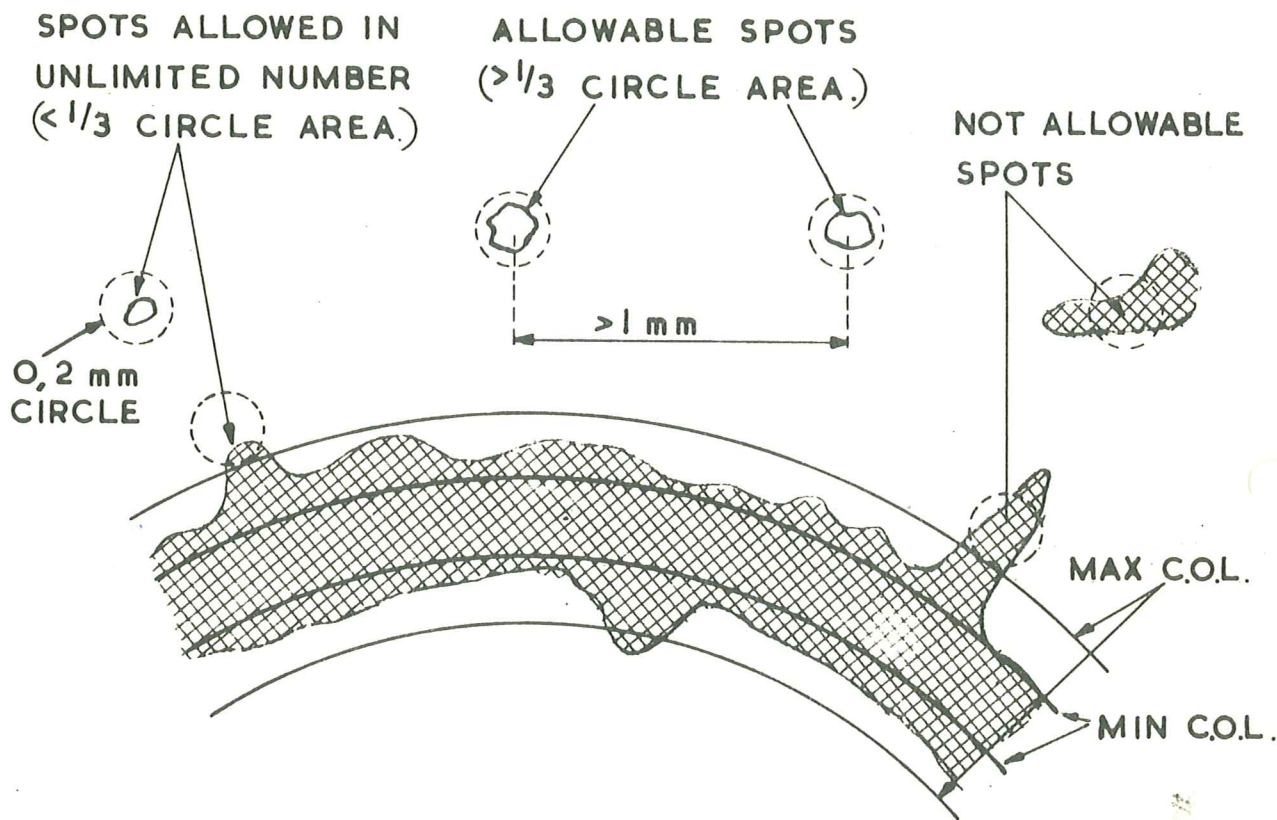


FIG. 6

Small spots or groups of small spots which are contained in a circle of 0,2 mm (0,008") diameter are allowed in unlimited number if the total area of the spot(s) contained in the circle of 0,2 mm (0,008") diameter is smaller than one third of the area of the circle (see Fig. 6).

#### 4.9 Voids

##### 4.9.1 Definition

Voids are defined as areas inside the minimum COL which can be identified as being of lower density than the printed image. Voids can be measured either visually or in terms of PCS.

Both methods constitute an acceptable procedure for measurement of the printed image for most printing. However, where there is a substantial question of subjective visual judgment the instrumental method based on PCS must be used.

#### 4.9.2 PCS Measurement of Voids

Sizes of voids depend on the PCS level at which they are measured. For a specific (see A.2.7.1) PCS level, a void is described as "allowable" if it satisfies the following conditions:

1. The distance the PCS aperture can be moved in a straight line, so as to give a PCS below the chosen level, nowhere exceeds 0,2 mm (0,008"). The measuring aperture must remain at all times inside the minimum COL.
2. The centre to centre distance of the void from the nearest other void, detected at the same PCS level, is at least 1,0 mm (0,040").

"PCS voids" is defined as the maximum PCS level at which all voids are allowable (see Fig.5).

#### 4.9.3 Visual Measurement of Voids

A void is allowable if it satisfies the following conditions:

1. It can be contained entirely within a circle of 0,2 mm (0,008") diameter, estimated visually.
2. Its distance centre to centre from the nearest other void is at least 1,0 mm (0,040").

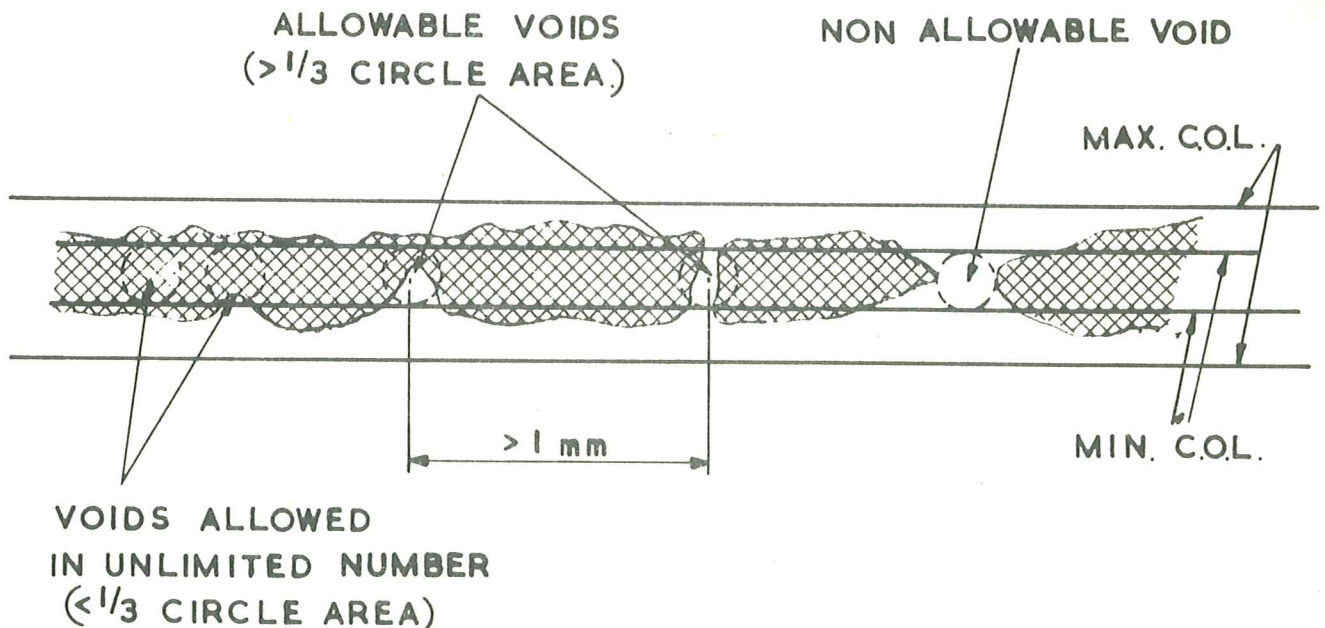


FIG. 7

Small voids or groups of small voids which are contained in a circle of 0,2 mm (0,008") diameter are allowed in unlimited number if the total area of the void(s) contained in the circle of 0,2 mm (0,008") diameter is smaller than one third of the area of the circle (see Fig. 7).

#### 4.10 Maximum PCS

The PCS max is the highest PCS level which is continuously exceeded for a scanning distance of 0,2 mm (0,008") within the maximum COL.

#### 4.11 PCS of Spots and Voids

PCS spots and PCS voids as defined in 4.8.2 and 4.9.2 must satisfy the following conditions :

1.  $\frac{\text{PCS voids}}{\text{PCS spots}} \geq 1,3$
2.  $\text{PCS voids} \geq 0,3$

It should be recognized that these are minimum values and that higher levels can generally be achieved. See Appendix A.2.7.

#### 4.12 PCS Voids and Maximum PCS

PCS voids and PCS max as defined in 4.9.2 and 4.10 should preferably satisfy the following conditions :

$$\frac{\text{PCS max}}{\text{PCS voids}} \leq 1,75$$

#### 4.13 Average PCS within a Character

In addition to the requirements stated in 4.11 and 4.12 at least 80 % of the PCS within a minimum COL (that is 80 % of the measurement made along the centreline) should preferably be greater than 0,4.

#### 4.14 Character Skew

The skew of a character is the rotational deviation of the printed image from its intended orientation relative to the document reference edge. Character skew must not exceed  $\pm 3$  degrees.

### 5. CHARACTER POSITIONING

#### 5.1 Introduction

This section contains basic specifications relating to the position of characters on a document to accommodate general requirements of OCR devices.

It does not contain all the rules which may be necessary for a particular application. (See 1.2, 1.3.)

#### 5.2 Document Reference Edge

Some specifications in this section relate to document reference edges.



These can be horizontal and/or vertical edges, preferably the bottom and/or right hand edges. Character alignment is relative to these reference edges.

### 5.3 Clear Area

A clear area is defined as that region of a document reserved for the OCR characters and the clear space around these characters. The locations and dimensions of clear areas will be determined by the nature of individual applications and the requirements specified in this section.

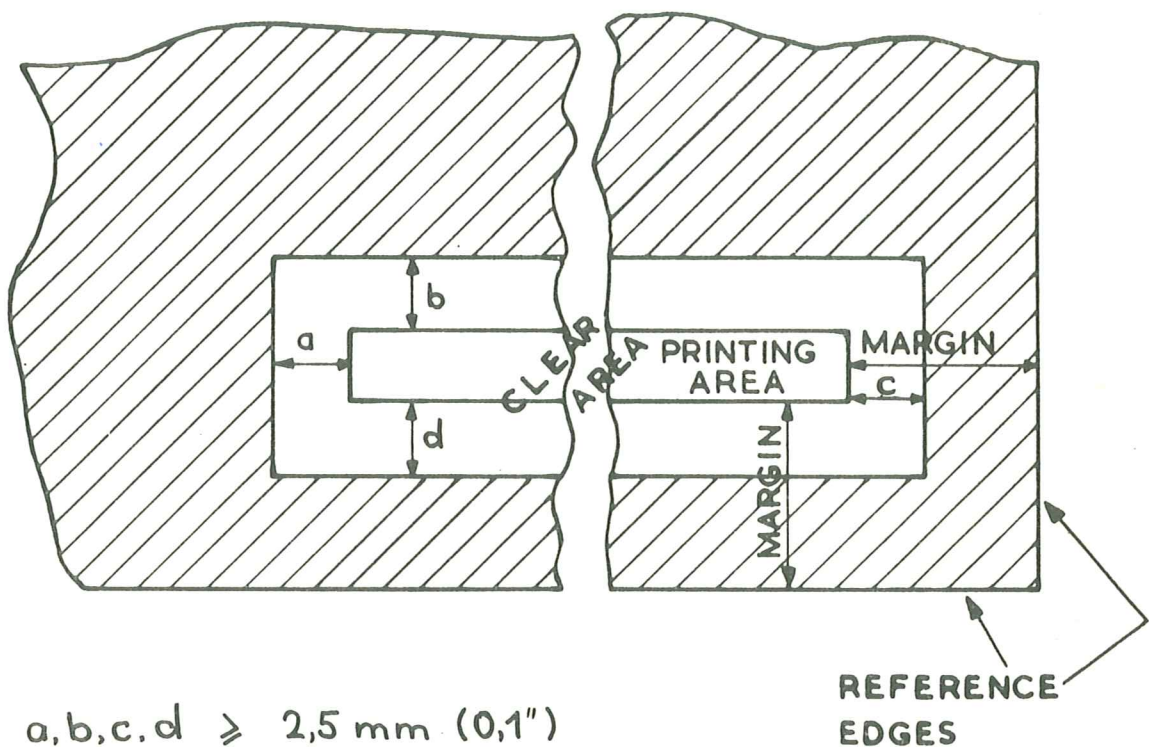


FIG. 8

### 5.4 Printing Area

A printing area is a rectangle inside the clear area, in which only OCR characters are to be printed. The sides of this rectangle must be parallel or perpendicular to a document reference edge (see Fig. 8). The distance between the corresponding boundaries of the printing area and the clear area should not be less than 2,5 mm (0,1").

### 5.5 Margin

The distance between the boundaries of the printing area and any paper edge is called the Margin (see Fig. 8).

A Margin should preferably not be less than 6 mm (0,236"). Where the choice of printing equipment imposes a smaller value, the absolute minimum is 0,36 mm (0,014"). In this case special consideration must be given to the compatibility of the print and the reading equipment.

#### 5.6 Field

A field is a specified portion of the printing area that is limited to sets of one or more characters that may be treated as a unit of information. These character sets must be located in a single line of printing. A line could comprise several fields (see A.3.4 and A.3.10).

#### 5.7 Field Boundary

A field boundary is defined as the smallest rectangle with sides parallel and perpendicular to a document reference edge, which contains all the boundaries of the component characters of the field.

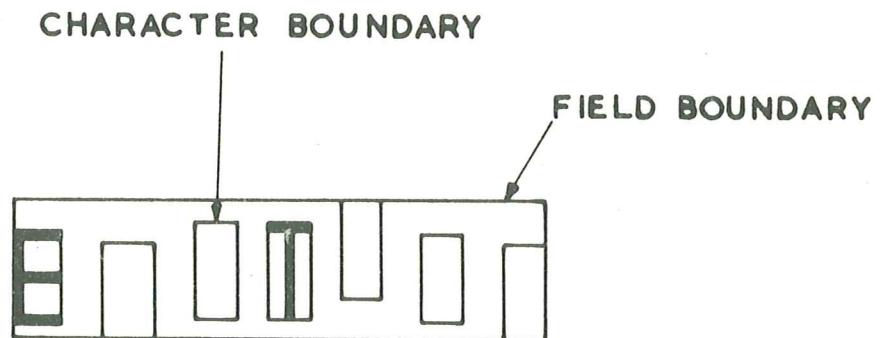


FIG. 9

#### 5.8 Line Spacing

Line spacing is the vertical distance between the average horizontal centreline positions of all OCR characters printed on one line and that of all OCR characters printed on the next line (see Fig. 10).

Nominal line spacing must be selected in such a way as to comply with the line separation tolerance (the parameters which influence line separation are: line pitch specification, line skew, vertical misalignment, character height and strokewidth).

In any case the line spacing shall not be less than 4,0 mm (0,157").

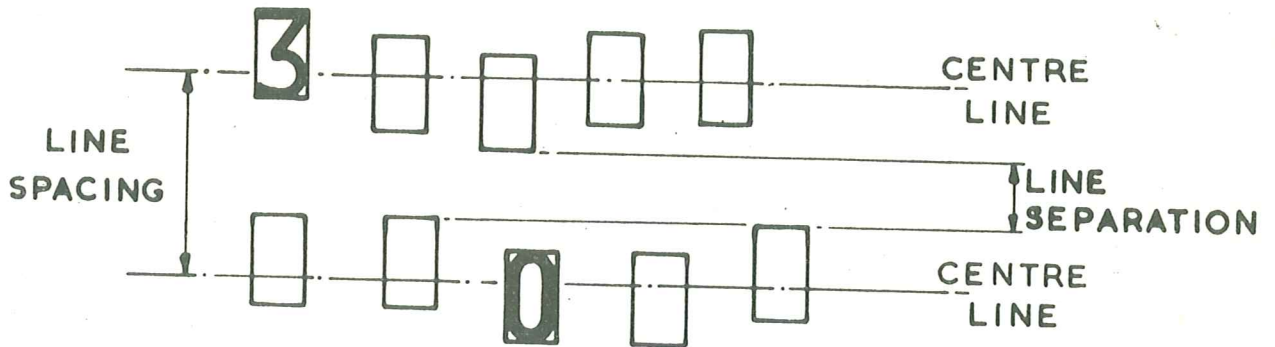


FIG. 10

### 5.9 Line Separation

Line separation is the vertical distance between the highest OCR character boundary (see 5.10) in a line and the lowest OCR character boundary in the line immediately above (see Fig. 10).

The line separation should not be less than 2,5 mm (0,1"). When closely-spaced lines are necessary (e.g. for pages) a smaller separation may be inevitable. The line separation should be maintained as large as possible by means of a reduction in vertical misalignment of the characters and by close conformity to the nominal strokewidth specification.

The minimum line separation shall not be less than the following values :

Size	I	II	III	IV
minimum line separation	0,64 mm (0,025")	1,0 mm (0,04")	1,5 mm (0,06")	2,0 mm (0,08")

If the character sizes are intermixed, the line separation limitation for any pair of lines shall be that applicable to the largest character in the two lines.

### 5.10 Character Boundary

The character boundary is defined as the rectangle with sides parallel and perpendicular to a document reference edge which is drawn tangential to the character outline and contains the character completely.

Skewed characters still have boundaries parallel or perpendicular to a document reference edge.



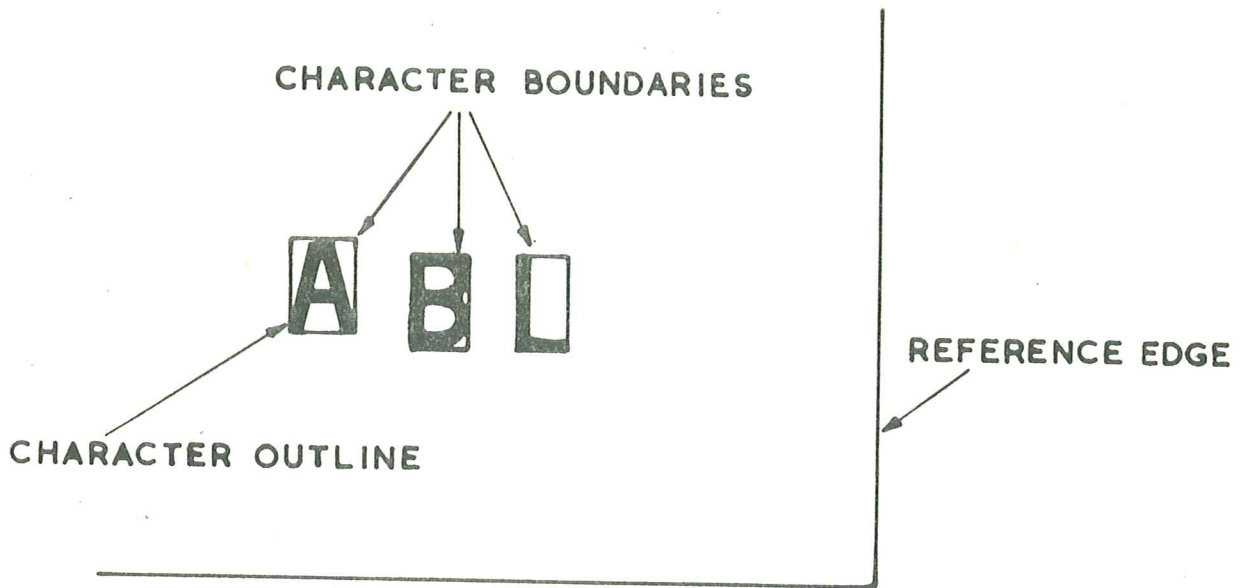


FIG. II

For the purpose of determining the boundary of the long vertical mark, only that portion of the long vertical mark will be considered which lies between the extension of the uppermost and lowermost horizontal boundaries of the adjacent character(s).

The character boundary is used to measure character and line separation and to determine field boundary.

#### 5.11 Character Reference Lines

Character reference lines are used to determine the position of a character relative to some other character or to some reference edge.

##### 5.11.1 Character Alignment Reference Line

The character alignment reference line is the horizontal centreline or the lower edge of the character boundary (see Fig. 13).

##### 5.11.2 Character Spacing Reference Line

The character spacing reference line is the vertical centreline of the character boundary.

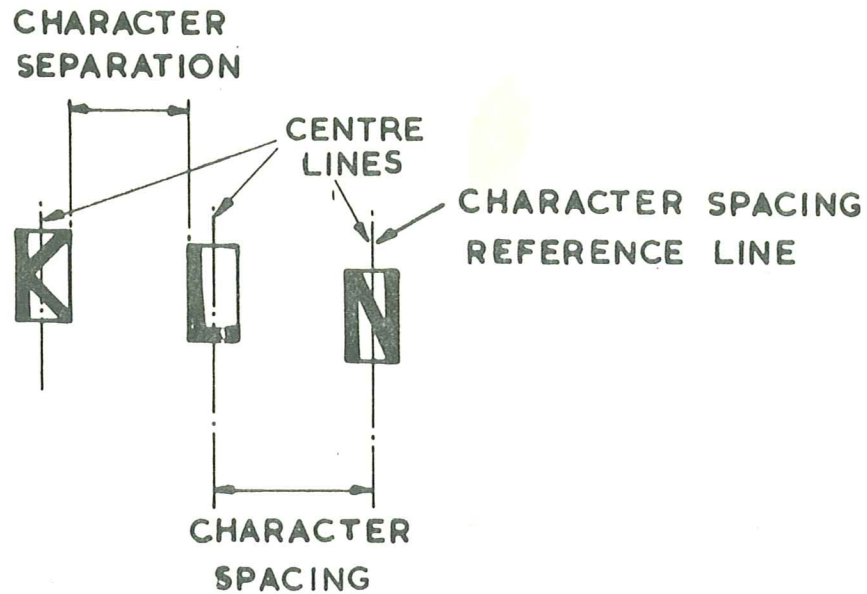


FIG. 12

### 5.12 Character Spacing

Character spacing (see Fig. 12) is the horizontal distance between the character spacing reference lines of two adjacent (including the Long Vertical Bar) corrected by the distance which would exist between the character spacing reference lines if the same two characters were superimposed in their nominal position. (This correction is derived from the nominal drawings and from the references used for the nominal alignment.)

For characters where the vertical reference line given in the nominal drawings coincides with its character spacing reference line, this correction does not apply.

Two characters are adjacent if the distance between their character spacing reference lines, corrected as mentioned above, is smaller than the following maximum values :

Size	I	II	III	IV
Maximum spacing	4,57 mm (0,180")	4,57 mm (0,180")	4,57 mm (0,180")	6,60 mm (0,260")

Character spacing of all characters shall not be less than the following specified minimum values :

Size	I	II	III	IV
Minimum spacing	2,29 mm (0,090")	2,29 mm (0,090")	2,29 mm (0,090")	3,30 mm (0,130")

Character spacing specifications will not be met when variable pitch printing is used (e.g. letterpress variable pitch type-writers; see A.3.8).

### 5.13 Character Separation

Character separation is the horizontal distance between the adjacent boundaries of any OCR character(s) and/or the Long Vertical Mark (see Fig. 12). The character separation shall not be less than the nominal strokewidth as specified in section 4.5.

### 5.14 Character Misalignment

Character misalignment is the vertical distance between the character alignment reference lines of two characters in the same line, corrected by the distance which would exist between the character alignment reference lines if the same two characters were printed in their nominal position. (This correction is derived from the nominal drawings and from the references used for the nominal alignment. However, it will not be needed, for instance, when the print is purely numeric or purely upper case alphabet. In other cases it should be determined whether a correction is necessary.) (See A.3.10.)

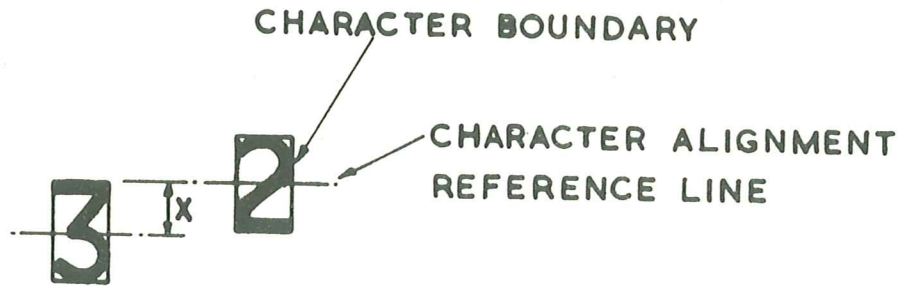
#### 5.14.1 Adjacent Character Misalignment

Adjacent character misalignment is measured according to the above procedure between the character alignment reference lines of adjacent characters (see Fig. 13). It shall not exceed the following values :

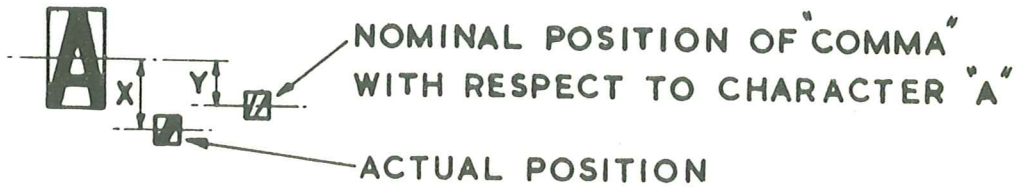
Size	I	II	III	IV
Max. adjacent char. misalignm.	0,66 mm (0,026")	0,66 mm (0,026")	0,89 mm (0,035")	1,07 mm (0,042")

This specification applies only to fields.  
(See 5.6 and A.3.4.)





$x$  = ADJACENT CHARACTER MISALIGNMENT.



$(x - y)$  = ADJACENT CHARACTER MISALIGNMENT

FIG. 13

5.14.2 Character Misalignment in a Field

Character misalignment within a field is measured according to the above procedure between the alignment reference lines of any two characters in a field (see Fig. 14). It shall not exceed the following values :

Size	I	II	III	IV
Mx. line char. misalignment	1,32 mm (0,052")	1,32 mm (0,052")	1,78 mm (0,070")	2,14 mm (0,085")

This specification applies only to fields.  
(See A.3.4 and 5.6.)



X CHARACTER MISALIGNMENT IN FIELD

FIG 14

5.14.3 Long Vertical Mark Alignment

The long vertical mark must extend beyond the top and the bottom boundaries of any adjacent character. A long vertical mark in one field should not extend nearer than 2,5 mm (0,1") to a field boundary in an adjacent line to which it does not apply.

## Appendix



## A 1 PAPER CHARACTERISTICS AND MEASUREMENTS

### A 1.1 Spectral Properties

#### A 1.1.1 Significance of Spectral Properties for OCR Documents

An OCR scanner will usually be responsive to a restricted band of optical wavelengths. Typically, these scanners respond to the near ultra-violet, the blue-green and green or the near infra-red wavelengths.

Therefore, it is a fundamental requirement that the paper used for an OCR document be a good reflector in the wavelength ranges of the optical scanner response.

#### A 1.1.2 Colour

It is strongly recommended that the paper for an OCR document be white. White paper is essentially non-selective to wavelengths of light within the range of interest for OCR scanners. Consequently if white paper is used no conflict of spectral properties will occur.

The specification excludes the use of most coloured paper, especially those with a definite and positive visual indication of colour.

If the saturation of the colour is slight, and the colour is essentially uniform throughout the OCR area on the documents, it is possible that they will comply with the specifications on average reflectance.

#### A 1.1.3 Notes on Measurements

##### A 1.1.3.1 Means of Realizing B 400

The light source specified in the body of the standard may be realized by a P16 or Q type of cathode ray tube. The detector specified in the standard may be a S4 or S11 photo-detector. The spectral response of the detector is extended to 500 nm in order to detect all the reflected energy as well as any energy resulting from fluorescent additives which may be present in the paper under evaluation.

##### A 1.1.3.2 Means of Realizing B900

To implement the B900 measurements the following components may be used :

illumination source: incandescent lamp  
sensor: silicon phototransducer

glass filter: A low frequency pass filter with cut-off at about 800 nm.

#### A 1.1.3.3 Absolute White Reference

For most practical purposes Magnesium Carbonate ( $MgCO_3$ ) which has a reflectance value very close to Magnesium Oxide ( $MgO$ ) may be used instead of  $MgO$  for the 100 % reflectance reference without significant loss of accuracy.

#### A 1.1.3.4 Fluorescent Additives

Fluorescent additives should be excluded in the manufacturing of paper for OCR use. However, it is realized that some contamination may be derived from previous operations. An effort should be made to minimize this contamination.

### A 1.2 Uniformity of Paper Reflectance

#### A 1.2.1 General

Most scanner systems for OCR will examine in detail the area containing the printed image.

The reflected light from small areas on the paper in the order of 0,1 mm (0,004") in diameter constitutes the input to a photo-detector. The presence of ink is determined by a significant change in the reflectance of these areas relative to the paper.

Paper normally has a variation of reflectance, on this small area basis, because of the formation of the fiber structure and may have similar variations due to its surface characteristics, embossment of patterns or the printing of coloured patterns.

It is important that the magnitude of any such variations be significantly lower than the magnitude of differences between paper and the printing.

Embossment of patterns or the printing of coloured patterns should be strongly avoided.

#### A 1.2.2 Measurement

##### A 1.2.2.1 Reflectance of the Black Backing

The backing of 0,5 % reflectance is most easily provided by an unlit cavity. If it is impracticable and if a solid with this low reflectance is not

available, solids with a higher reflectance (not more than 3 %) may be used without serious loss of accuracy.

#### A 1.2.2.2 Procedures

It appears that the distribution of the reflectance values measured over equal areas of a sample of paper, closely approximates to a normal curve. Such a normal distribution is defined by its mean (average paper reflectance measured with a black background) and its standard deviation  $\sigma$  (variation in paper reflectance).

If discrete measurements are made separated by at least 2 mm (0,08") they can be considered as being non-correlated. The number of observations required for a reliable determination of  $\sigma$  is then of the order of 200.

If the observations are taken in one or more continuous scans, it is required that a total scanning length of at least 20 or 40 cm (8" or 16") be covered, for high or medium opacity papers respectively, in order to avoid the influence or correlations between the reflectances of neighbouring points. This corresponds approximately to 200 non-overlapping points.

A procedure which avoids the calculation of the standard deviation and may be found more convenient in practice (still being sufficiently accurate) is as follows :

- Arrange the measurements obtained by one of the above scanning methods in a descending order of magnitude.
- Exclude highest 1/2 % and the lowest 1/2 % of the values. Calculate the ratio  $R_{max}/R_{min}$  for the remaining values.
- This ratio should not exceed 1,2 for the high opacity class and 1,3 for the medium opacity class.



For paper of wild formation with high variation in transparency the distribution of reflectance values may appreciably deviate from the normal distribution. In these cases the procedure which avoids the calculation of  $\sigma$  will be more satisfactory.

#### A 1.2.3 Recommendation

The values given in the standard should be considered as lower limits. For reliable recognition more stringent limits are recommended and are obtainable with high quality paper.

### A 1.3 Paper Opacity

#### A 1.3.1 Significance of Paper Opacity

The opacity is indicative of the change in paper reflectivity on an OCR document due to the backing material at the time of scanning. If the document transport system of the OCR device is such that a known uniform reflective surface is provided at the time of scanning, a moderately opaque paper may be usable.

However, some systems scan the document while backed by other printed documents or have a transport system that provides a non-uniform backing surface. For such cases a more opaque paper should be used, or higher PCS value should be required for OCR information.

#### A 1.3.2 Recommendations

The minimum opacity required for an OCR paper will be dependent upon the means of scanning and the application. In general, opacity is related to the basis weight of the paper; the higher the basis weight the greater the opacity. Consequently, there is a similar relationship between opacity and paper thickness, although the use of filler and coating materials have an effect.

In general, paper having Opacity exceeding 85 % should be used. Papers of lower opacity should be used only if needed for the application and after considering the scanner optical system. Papers having Opacity less than 65 % should not be used.

Many inks have the property of permeating the paper to a considerable depth. Applications requiring an OCR document to be printed on both sides may require a higher opacity or thicker paper to compensate for this effect.

#### A 1.4 Paper Gloss

##### A 1.4.1 Significance of Gloss for OCR Documents

Gloss is the property of a surface responsible for a lustrous or mirror-like appearance. It is a phenomenon related to the specular reflection of incident light. The effect of gloss is to reflect more of the incident light in a specular manner, and to scatter less. It occurs at all angles of incidence and should not be confused with grazing angle specular reflection that is often referred to as sheen.

Paper gloss is undesirable for OCR systems since it will change the effective brightness of the paper, thus affecting the print contrast signal.

##### A 1.4.2 Recommendations

Paper for OCR documents should be restricted to the low gloss varieties. The use of coated or super-calendered papers or other papers with a glossy appearance should be avoided.

#### A 1.5 Dirt in Paper

Dirt in paper refers to the presence of relatively non-reflective foreign particles embedded in the sheet. Generally these particles are quite small and infrequent in good quality paper. The frequency of their distribution is significant.

The size and lack of reflectance of the particles may be such that they will be mistaken for inked areas by an OCR scanner.

#### A 1.6 Mechanical Properties of Paper

Some mechanical properties of paper, such as tear resistance, bursting strength, folding resistance, etc. may be significant in OCR applications.

It is advisable that there be agreement on the specific papers intended to be used for such applications between users and manufacturers of OCR systems.

### A 2. CHARACTERISTICS OF THE PRINTED IMAGE

#### A 2.1 General

This standard specifies the requirements for optimum reading system performance.

The specifications should be met by all print as far as is possible in the presence of the random effects which occur in any printing process.

The design of printers and the selection of supplies should assure maximum compliance with the standard. In

any system the specifications may occasionally be exceeded, but the frequency with which this is allowed to occur should be carefully studied in the light of the reader performance required.

A 2.2 Rules for the Design of Measuring Gauges for OCR Characters

The gauges are constructed by superimposing the minimum and maximum strokewidths, as specified in the standard, symmetrically about each point on the character centreline drawings.

The following rules also apply :

1. "Internal angles" of maximum COL and "external angles" of minimum COL should be rounded with a radius of 0,1 mm (0,004").
2. When the character centreline presents a sharp corner, the "external angle" of maximum COL and the "internal angle" of minimum COL should be identical to the angle of the character centreline.

NOTE:

- "external angles" means angles which lie in the region where the angle determined by the character centreline is greater than  $180^{\circ}$ .
- "internal angles" means angles which lie in the region where the angle determined by the character centreline is smaller than  $180^{\circ}$ .
- The centreline of the character may be included in the gauge to help in finding the best fit and in measuring centreline deviations.

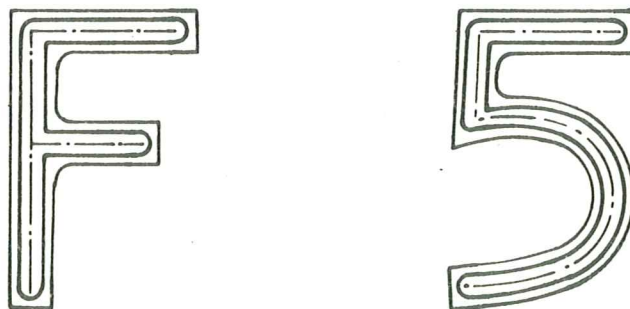


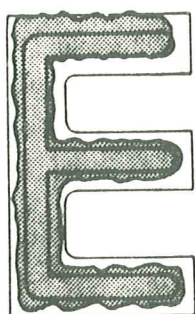
FIG 15



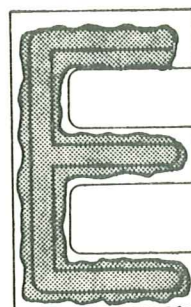
### A 2.3 Centreline Deviations

The tolerance on centreline deviations is specified in the standard in order to limit the allowed deformation in the printed character caused by uneven printing and to give guidance on the choice of tolerances in the manufacture of type.

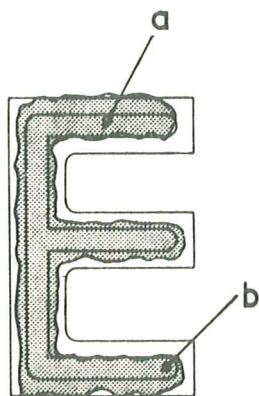
When this tolerance is met the printed edges will be symmetrically distributed along the centreline of the gauges aligned to give the "best fit".



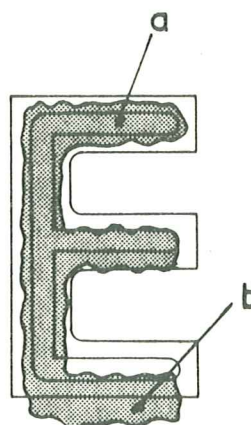
A



B



C 1



C 2

THE PRINT OF EXAMPLE A AND B SATISFIES THE SPECIFICATION  
THE PRINT OF EXAMPLES C IS OUT OF SPECIFICATION BECAUSE  
WHEN THE GAUGE IS CENTERED ON THE STROKE „a“, THE STROKE  
„b“ WILL NOT BE CONTAINED WITHIN MAX. AND MIN. C.O.L.

FIG. 16.

The tolerance may be checked by means of the gauge with the following procedure (see Fig. 16) :

1. When the gauge is centred on any stroke of the character, so that the edges are symmetrical to the corresponding centreline, all the rest of the character edges must lie between maximum and minimum COL.
2. The above condition must be satisfied centering the gauge on any stroke or portion of the character under examination.

#### A 2.4 Strokewidth Ranges

The variation in strokewidth from the nominal should be held to a minimum, since generally this could have a bearing on the reader performance.

Strokewidth Range X requires a high quality printing process and careful control of maintenance and supplies. It cannot be met by some printers in common use for OCR. However, the tolerances which these printers normally produce do not necessarily extend to the full Range Y. In such cases, printing performance should not be allowed to degrade beyond the normal level.

#### A 2.5 PCS and Visual Measurements

In drawing up the standard it was recognized that the majority of print measurements will be made visually. At the same time an objective method of measurement, closely related to the way in which optical scanners operate, was required for critical evaluations.

The PCS specification fulfils the requirements for scanner-related measurements. Every effort has been made to ensure that the two specifications are equivalent and they are intended to be useful independently. Exact correlation is not possible, however, and some differences will arise (see A 2.7 below).

#### A 2.6 Spectral Bands for PCS

For machine recognition of printed information it is necessary that a good contrast exist between the printed image and the paper. This contrast, expressed in PCS, is obtained when the paper has a good reflectance and the print is dense enough to provide a good absorption in the spectral range of interest.

Reading devices usually have a spectral response in the near UV, the visible or the near IR spectrum.

A printing ink provides good absorbance in one or more of these bands, depending on its composition. For example, black pigments tend to absorb light in all three bands, but dyes are more selective and usually yield the best absorption in the visible region.



Because of the diverse nature of printing equipment and OCR systems it is impossible to specify a single spectral range which contains the spectral responses of all reading devices and in which all printing inks would absorb sufficiently.

Which of the three specified spectral bands should be used, therefore, depends on the reading and printing devices in the application concerned. The following considerations apply:

- a. If the characteristic of all readers in the system are known, it is sufficient to choose the spectral band(s) appropriate to these readers.
- b. Printing which is required to satisfy the PCS specifications in the visible range imposes the least restriction upon the spectral characteristics of the printing inks.
- c. The only print which can meet the spectral requirements of all reading systems is that which conforms to the specification in all three bands. Print on white paper with ink of a high carbon black content will in general meet this requirement. This consideration also applies in applications where the reading systems to be used are not known when the application is to be established.

PCS measurements in the near UV region have to be implemented in the band B 425. However, it is realized that in some cases it might be convenient to use a band slightly shifted toward B 400. It must be noted that such measurements might give lower PCS values than those obtained using B 425.

## A 2.7 PCS of Spots and Voids

### A 2.7.1 Definition of PCS Spot and PCS Void

A printed image contains, in most cases, voids within the minimum COL and spots outside the maximum COL but within the neighbourhood of the characters. These spots are defined as character associated spots (see 4.8.1).

When a character is checked, signals similar to the curve (Fig. 5) will be generated. A threshold is set to decide which areas in the character field are black information and which are white.

Depending on the level of this threshold a small spot may or may not be detected as black. Similarly small voids may or may not be detected as white. For each character under examination it



is possible to find a specific PCS value for the threshold at which all spots are allowable and similarly a PCS value can be found at which all voids are allowable.

These values are called PCS spots and PCS voids respectively.

The specifications for allowable spots and voids are given in 4.8.2 and 4.9.2. The specifications on the PCS values are given in 4.11, 4.12, 4.13.

#### A 2.7.2 Significance of Spots and Voids

For machine recognition of the printed image it is essential that the print intensity of all parts should be high enough to exceed a certain minimum value and be distinguishable from the background. These requirements are covered by the specifications for PCS voids and PCS spots

The minimum PCS found within the outline of a character is a measure of the smallest useful signal that the character will produce in an OCR scanner. If the detection threshold is put above this value, the character will display voids.

Because of the distinction between allowable (small) voids and non-allowable ones, the value of PCS voids is, in general, somewhat higher than this minimum. It is the highest PCS at which all voids are still allowable, and is, broadly speaking, a measure of the contrast between the character and its background. PCS spots likewise, is not the PCS level at which spots first appear but the (lower) level beyond which they become too large to be allowable. It is related to the intensity of background noise in the region of the character.

The ratio  $\frac{\text{PCS voids}}{\text{PCS spots}}$  is therefore an indication

of the signal-to-noise ratio of the printed information, and the higher it is, the better the performance of the reading system is likely to be. The value of 1.3 should be easily attainable by the printing processes used for OCR documents. Many mechanisms, particularly those which print within the strokewidth range X, will be able to keep the ratio greater than 1.7.

As PCS voids diminish, the print contrast diminishes also, until in the limit it is no greater than the level of reflectance irregularities

in the paper. An increase in PCS voids will tend to improve reading system performance. This can be achieved, for instance, at some extra cost, by a reduction in the allowed duration of ribbon life. Many mechanisms, particularly those which can print within the stroke-width  $X$ , will be able to, and should, maintain PCS voids above 0.4.

It is essential that a majority of the characters have a PCS higher than the lowest PCS voids which the specification allows (0,3). This requirement is covered by the specification for the average PCS within a character.

The variation in PCS level within a single character is a measure of the unevenness of printing within that character. Even if the minimum PCS within the character is adequate, the reliability of reading will decrease as the variation in PCS increases. For example, in a reading machine which automatically adjusts its threshold level based on the average level of the printing, an extremely wide variation can cause the threshold to rise to such a level that the lower PCS areas are not recognized. This requirement is covered by the specification for PCS voids and PCS max.

In addition to these requirements the following conditions should be satisfied :

1. Characters should not display allowable voids and allowable spots at the same PCS level.
2. The PCS in the area between maximum and minimum COL should not be higher than the PCS within the minimum COL.

It is important to note one practical difference between the visual and PCS methods of assessing spots. By the visual method, the assessment depends only on the dimensions and spacing of the spots, irrespective of the density of the associated character. In PCS terms, however, a spot associated with a dense character may pass the specification while an identical spot associated with a weak character will be rejected.

With respect to 4.8.3 and 4.9.3 the following considerations are relevant :

Small spots and groups of small spots are acceptable in unlimited number when the visual method is used. This is because experience indicates that either distortions of this kind are rarely those which define the value of PCS spots for a character (i.e. that larger faults will also be found) or that the value of PCS spots will be



well below the specified limit. In cases of doubt the PCS measurement must be employed. A similar situation applies to the visual measurement of small voids.

#### A 2.8 Spots Remote from a Character

The neighbourhood of a character is defined in 4.7.1 as an area twice the nominal character height by twice the nominal character width and centred on the character being measured. The PCS level and frequency of spots in the neighbourhood of a character are specified in 4.8 and 4.11.

The size and frequency of spots remote from a character should also be strictly controlled. The size of spots should be minimized; printing smudges and regular patterns of dots should be avoided.

Many reading operations are started upon detection of the first black point and if a spot occurs larger than 0,2 mm (0,008") then the recognition process may begin. It is advisable that spots greater than 0,2 mm (0,008") be prevented.

#### A 2.9 Recommendation for lower-case OCR-B Characters

For the following set of characters a higher print quality is required, both in terms of PCS and strokewidth. PCS voids should be above 0,4. The ratio  $\frac{\text{PCS voids}}{\text{PCS spots}}$  should be kept greater than 1,7.

Strokewidth variations should be maintained within range X.

a b c d e f g h i j k l m n o p  
q r s t u v w x y z m ð ø æ @ "  
' ` ^ ~ ,

### A 3 CHARACTER POSITIONING

#### A 3.1 Objectives of the Character Positioning Requirements

Character positioning specifications (Format Rules) are needed to ensure that each OCR character on a document is "seen" by the reading device without interference from the other OCR characters or from non-OCR matter. The Format rules given in the standard (which are explained in the following sections) are to be taken as minimum requirements and may need to be supplemented by further rules for specific systems.

#### A 3.2 Document Reference Edge

The document used in an OCR system must be moved and



suitably positioned for printing and reading the OCR information. One or more document edges are used to provide a reference for these operations. Because of the diverse nature of OCR documents, it may sometimes be convenient to specify one reference edge (e.g. for journal rolls); for others it may be necessary to specify two edges (e.g. for cheques the bottom right hand edges are usually specified).

The tolerance on the distance between the average horizontal centreline on a line of OCR characters and a top or bottom reference edge may be vital to the satisfactory functioning of the system. No dimension for this tolerance is given in the specification, since system requirements differ widely, but its importance must not be overlooked.

### A 3.3 Clear Area, Printing Area and Margin

OCR printing must be isolated from all other printing or patterns on the document in order to allow the reading device to distinguish the OCR information more readily. This isolation is provided by maintaining a "border" of blank paper between the OCR information and the remainder of the document. From this arises the distinction between the "Printing Area", which must include all of the OCR characters and the larger "Clear Area" which includes the Printing Area and must be free from any other printing or embossing. If the distance between the boundary of the Clear Area and that of the Printing Area approaches the minimum specified, due account must be taken of printing tolerances (vertical misalignment, etc.) and expected paper dimensional changes. It is good practice in document design to provide as generous a Clear Area as possible.

The boundary of the Printing Area should be kept well within the paper edges, i.e. the margins should be large. This has the advantage, among others, that a moderate degree of edge mutilation can take place without impairing readability. There are some special cases, however, where the small size of the document may make large margins impracticable and the boundary of the Printing Area may then have to lie close to the document edge(s). Relaxation of the specification in this respect is only permissible when it has been established that all OCR devices in the system can handle these documents.

The dimensions of the printing area and its position relative to the document edge(s) are important for readers that have limited line finding capabilities.

A 3.4 Fields

For the purposes of this standard a field is any group of characters printed at the same time on the same line by the same mechanism.

A 3.5 Line Spacing

Line spacing is only significant for multiple line documents. It is the intent of the standard to limit the number of lines of printing that may occur within a given vertical distance. This limitation is necessary in addition to the requirements of line separation, since characters in a line may all be less than full character height, (e.g. symbols, such as minus). In such cases the line spacing must be maintained to permit printing of full height character.

The maximum line packing density permitted by the tolerances given in the body of the standard for the four character sizes are approximately :

Size	I	II	III	IV
Lines per 25,4 mm (1")	6	5	4	3

However, for these values to be acceptable the tolerances on the parameters influencing Line Separation must be below the maxima specified, which apply for wider spacing. (The parameters which influence Line Separation are: line pitch tolerance, vertical misalignment, character height and strokewidth.)

In general, line spacing should be kept as large as possible consistent with the other requirements of the system.

A 3.6 Line Separation

Line separation defines the isolation required between successive lines of OCR information.

Some documents may require and permit more dense spacing of lines of OCR information than can be accommodated with the recommended line separation of 2,5 mm (0,10"). See A 3.5 above.

An absolute minimum value of line separation for each of the four character sizes is given. Where this minimum is approached an effort should be made to ensure as large a Line Separation as possible, by controlling



character alignment, character strokewidth, and if possible, line spacing.

#### A 3.7 Character Boundary and Character Reference Lines

The Character Boundary is defined for the actual printed image under examination rather than for an ideal character. This is done in order that the limits assigned to the separation between characters and lines shall be realistic and applicable to any quality of print.

In addition to the Character Boundary, which varies in extent according to the strokewidths (wide or narrow), it is necessary to specify vertical and horizontal reference lines, which depend only on the position of the printed character and which serve for measurements of character spacing and misalignment.

Two alternative definitions of the character alignment reference line are given. In marginal cases where the two methods may yield different results, the one giving the smaller misalignment is to be accepted.

#### A 3.8 Character Spacing

It is the object of the character spacing requirement of the standard to define the lateral relationship of any pair of characters side by side in the same field in such a way that the maximum and minimum character separation requirements can be met.

As mentioned in 5.12, the specification on Character Spacing will not be met when variable pitch or variable set width printing is used (e.g. variable pitch typewriters, letterpress). Since these types of printing use wide variation in the character width and spacing they may impose difficulties for OCR devices, special consideration must be given to the compatibility of the print and the reading equipment.

#### A 3.9 Character Separation

It is a primary requirement of OCR that characters side by side in the same line shall be isolated by a clearance of unprinted paper. This separation constitutes a vertical band (of width not less than the nominal strokewidth, as defined in section 4.5) which may not be intruded upon by any part of the character outline.

In order to satisfy the minimum character separation requirement, in difficult cases where the nominal character spacing is close to the minimum, the following points need particular attention :



- a) Strokewidth variation
- b) Character skew
- c) The difference that exists for certain characters between the character spacing reference line and the vertical reference line given in the character drawings. For the Class B character L (size I), for instance, this distance is as much as 0,18 mm (0,007").

#### A 3.10 Character Misalignment

The vertical misalignment of characters should be limited to reduce the cost and complexity of OCR devices, to an extent that is compatible with normal and relatively unsophisticated printing equipment.

The misalignment may be due to :

- a) misalignment of individual print faces,
- b) misalignment of the document in the printer, causing a complete group of characters printed at one time to be displaced vertically and/or tilted (skew),
- c) local distortion or folding of the document before, during or after printing.

This section of the specification limits the degree of misalignment of adjacent characters, with an overall limit on the misalignment of any two characters in a field. The allowable misalignment within a complete line (more than one field) of OCR characters is not specified but will be determined by the reading device(s) in the system.

Misalignment in this case could be the effect of the fields being printed at different times by different printing devices. Therefore it is important to determine the potential misalignment and the requirements for a specific application in order that specifications and controls can be established.

The measurement of actual character misalignment is complicated if symbols, lower case characters, and others which are not of the normal height are present. For such characters the alignment reference lines as defined will be at different levels even when there is no misalignment. Appropriate correction factors can be obtained from the character drawings.

A 4. GENERAL ILLUSTRATION OF OCR-A AND OCR-B CHARACTERS

These reproductions are for illustrative purposes only. For exact dimensions of the characters, see Standards ECMA-8 and ECMA-11, available free upon application to ECMA, 114 Rue du Rhône, 1204 Geneva, Switzerland.

A 4.1 Illustration of OCR-A characters (numeric set according to ECMA-8), scale 5 : 1

0 1 2 3 4 5 6 7 8 9

√ 4 H

|

A 4.2 Illustration of OCR-B characters (complete set according to ISO D.R. 996) scale 5 : 1.

A 4.2.1 Characters of OCR-B having a strokewidth of 0,35 mm (0,014") for sizes I and II and of 0,38 mm (0,015") for size III :

A B C D E F G H

I J K L M N O P

Q R S T U V W X

Y Z \* + , - . /

0 1 2 3 4 5 6 7

8 9

Ä Ö Å Ñ Ü Æ Ø

↑ ≤ ≥ × ÷ 0 α



A 4.2.2 Characters of OCR-B having a strokewidth of 0,31 mm (0,012") for sizes I and II and of 0,34 mm (0,013") for size III, which are considered in 4.5.

£ \$ : ; < % > ? [ ! # & , ]

( = ) \_

A 4.2.3 Characters of OCR-B having a strokewidth of 0,31 mm (0,012") for sizes I and II and of 0,34 mm (0,013") for size III, which are considered in A 2.9.

a b c d e f g h

i j k l m n o p

q r s t u v w x

y z m ð ø æ

" ' \ ^ ~ ˇ  
,

@

