Document # 40

CCITT SG XV

WORKING PARTY ON VISUAL TELEPHONY
SPECIALISTS GROUP ON CODING FOR VISUAL TELEPHONY

Source: BT & GEC McMichael

Title: **INTER-REGIONAL CONNECTIONS - THE STANDARDS CONVERSION PROBLEM AND BASIC CODEC PARAMETERS**

INTRODUCTION

The fundamental requirement in specifying the basic parameters of the second generation codec is to devise a system that:

(a) Is capable of providing an order of magnitude improvement in performance (bit rate or quality) than first generation codecs.

(b) Is inherently a "world" system such that inter-regional operation is a fundamental attribute.

(c) Capable of improvement and extension to applications other than simple face-to-face videoconferencing.

(d) Provides for flexibility, both for future enhancement of the codec and cost/complexity in hardware design.

(e) Provides for inter-regional operation but does not impose an unacceptable cost in terms of picture quality or price versus a codec designed for intra-regional use only: it is not the task of the specialists group to specify only an intra-regional device.

(f) For videoconferencing applications a full spatial definition 'graphics' facility needs to be possible.

At the Holmdel meeting a number of video formats and strategies for standards conversion were discussed but no consensus reached as to the best approach. This document examines further the basic problems of standards conversion and taking into account the requirements (a)-(f) above proposes a range of coding formats, with which we recommend all codecs should comply.

THE PROBLEMS

First we examine the problem of overcoming differences in field or frame frequency.

1. At both encoder and decoder a frame memory or frame buffer (as referred to in the Japanese document #21) always has to operate at the local input or output video rate to accept or provide picture display.

2. At the encoder, to change the field rate for the purpose of coding, the same frame memory can be used for interpolating the new fields as for recursive pre-filtering, provided that the output field rate F2 is less than the input rate F1 (see Figure 1). A separate frame memory must be used for

source coding, as one needs to source code the fields that are actually transmitted and therefore "time and space corrected". The requirement for a simplest possible, cheapest, implementation could be met by dispensing with store S1, then non time corrected fields could be used for coding in S2. The source coder store S2 still operates at the clock rate determined by F1 and stops when a field is to be dropped. The gap is then removed in the subsequent transmission buffer, which only needs to store a compressed frame.

So to summarise if F2 is less than or equal to F1 then one can construct a cheap encoder with a single frame store. If one requires better picture quality then a further frame store can be provided which can also be used for some pre-filtering operations such as recursive noise reduction.

3. At the decoder (see Figure 2), provided the input field rate F2 is less than or equal to the local display rate F3 then the same frame memory can be used for source decoding and temporal correction. In the diagram F-3 is at the local display clock rate but only has F2 frames per second, ie, there are some frame length gaps. The source decoder stops during the gaps. The transversal tap across the frame store is used to provide the temporal correction for display purposes.

If however F2 is greater than F3, then the source decoder store still needs to run at least at F2 because one cannot discard coded frames. This means that a further frame store or buffer is required for display, unless the frame store can have a second independent read cycle. This could considerably complicate the store design, for example, to also use the store for graphics at full definition would require around 40MHz clocking with 13.5MHz sampled images.

4. There is no simple integer relationship between PAL/SECAM and NTSC field and frame frequencies. There are however significant advantages in maintaining simple integer relationships for spatial and temporal sub-sampling ratios when carrying out source coding.

Similar arguments apply to the number of lines per picture: in this case however, the overhead in terms of extra storage, if the number of transmitted lines per frame is greater than the highest number of lines in the display, is not very great.

There are a number of other factors to consider:

1. We are specifying a codec for n x 384 kbit/s, at n = 4 or 5 then videoconferencing is not likely to be the main application. Other applications may require full motion rendition and not be limited by virtue of the coding algorithm.

2. It is desirable that in 'graphics' mode, codecs be capable of exploiting the full definition of the television system, ie 575 lines for PAL/SECAM. Future developments may avoid the vertical 'kell' factor loss from interlace by the use of a non-interlaced high rate display.

3. To depict movement, a picture repetition rate of 20-25Hz is required, although in many circumstances 15Hz has proved adequate. The picture display rate is determined by the eye's tolerance to flicker, and there is no suggestion that one should change from the existing regional standards.

4. Source coding and vertical filtering is far more efficient if the coded video format is non-interlaced. Conversion to a non-interlaced format involves a temporal correction at the encoder of + ¼ line interval.

At the decoder one could merely field repeat (as used in telecine) or carry out the inverse operation. In practice it is unlikely that one would see any improvement from the latter.

5. A 'handshake' system for video formats should be avoided to allow for unidirectional or broadcast type applications. This would also simplify multipoint use.

6. Manufacturers are likely to wish to produce equipment for both regions - it is far more economic to do so if a large proportion of the hardware is identical in 525 and 625 codecs.

7. In the specialists group we need to produce a specification for a 'world' codec, practically this will be a great deal easier if our discussions are limited to common video formats.

PROPOSAL

(a) Moving Pictures

A range of common intermediate coding formats be defined for moving pictures with which all codecs comply. The proposed range of formats are below:

NUMBER OF SAMPLES PER ACTIVE LINE	NUMBER OF FIELDS PER SECOND	INTERLACE	NUMBER OF LINES PER FIELD
360	30*	1:1	288
360	15	1:1	288
360	10	1:1	288

In all cases the use of a line quincunx sampling structure should be studied - this could provide enhanced definition but care will be required when considering the use of block transform techniques for source coding. Also an additional optional format of 720/30/1:1/576 could be considered for operation at primary rates.

*To be precise 29.97Hz.

(b) Graphics Mode

Provision should be made for an optional graphics facility using a 576 line picture format common to 525 and 625 codecs. Transmission may be at 64 kbit/s or at the same rate as the moving pictures. In the latter case it would be advantageous that the format be chosen such that the full definition facility can be integrated with the moving picture hardware.

DISCUSSION

1. The field frequency means that NTSC countries do not need to carry out temporal conversion or suffer any temporal distortion on intra-regional calls. PAL/NTSC countries carry out a negligible distortion 'up-down' conversion.

The use of a non-interlaced format avoids the need for an additional 'standards conversion' frame memory in the decoder: it also simplifies many source coding algorithms.

2. The number of lines per field allows PAL/SECAM countries to retain their higher definition line standard. NTSC countries need to carry out a 484 to 288 line conversion. In converting to 288 lines there will be some loss from the necessary pre-filter in comparison to field dropping to 242 but the extra 19% lines should more than compensate.

3. The hardware requirements for PAL/SECAM and NTSC are similar, both have the same number of field stores. PAL/SECAM codecs require extra multipliers for temporal conversion. NTSC line interpolators require more sets of coefficients than the PAL/SECAM ¼ line shifters. The attached diagrams show the hardware requirements in each case.

4. Provided the above philosophy is adopted in conceiving a hardware design, the hardware overhead for a 'world' codec compared with an intra-regional codec only is minimal. This is certainly not the case for a dual-mode approach, where 'world' operation will imply the addition of, at least, extra frame memories. There would also be several system problems, particularly with multipoint connections where rapid synchronisation between different standards would be difficult to achieve.

RCN/081

Fig 1

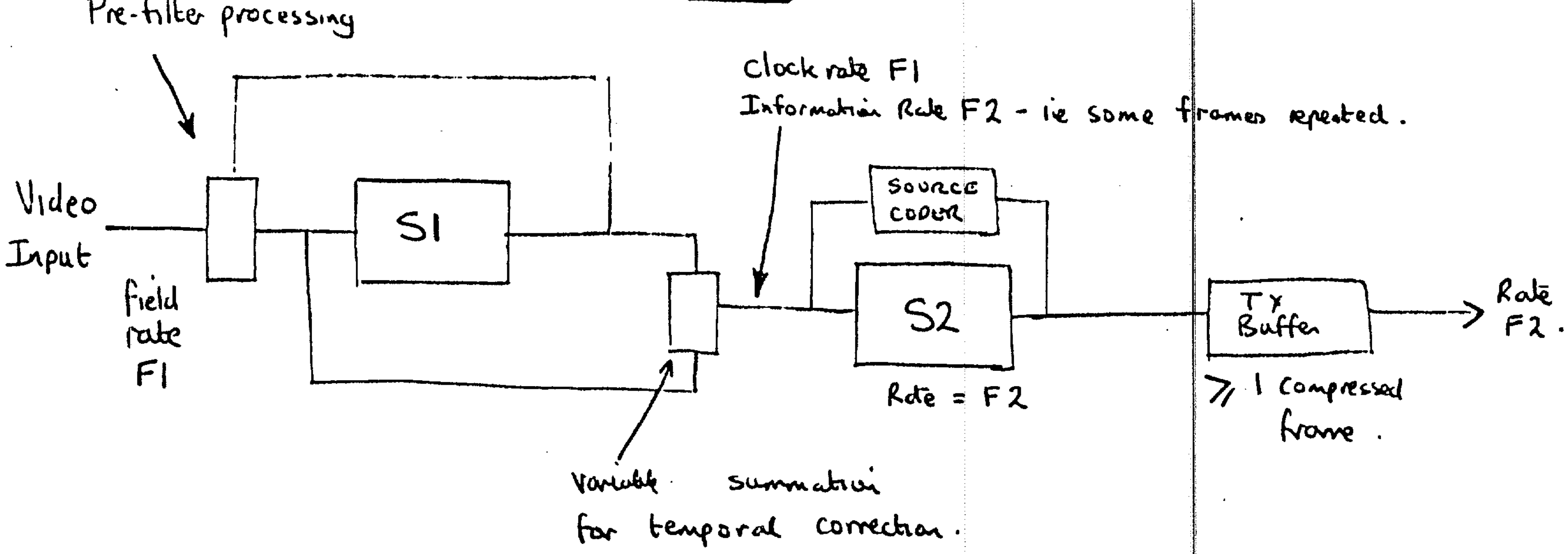
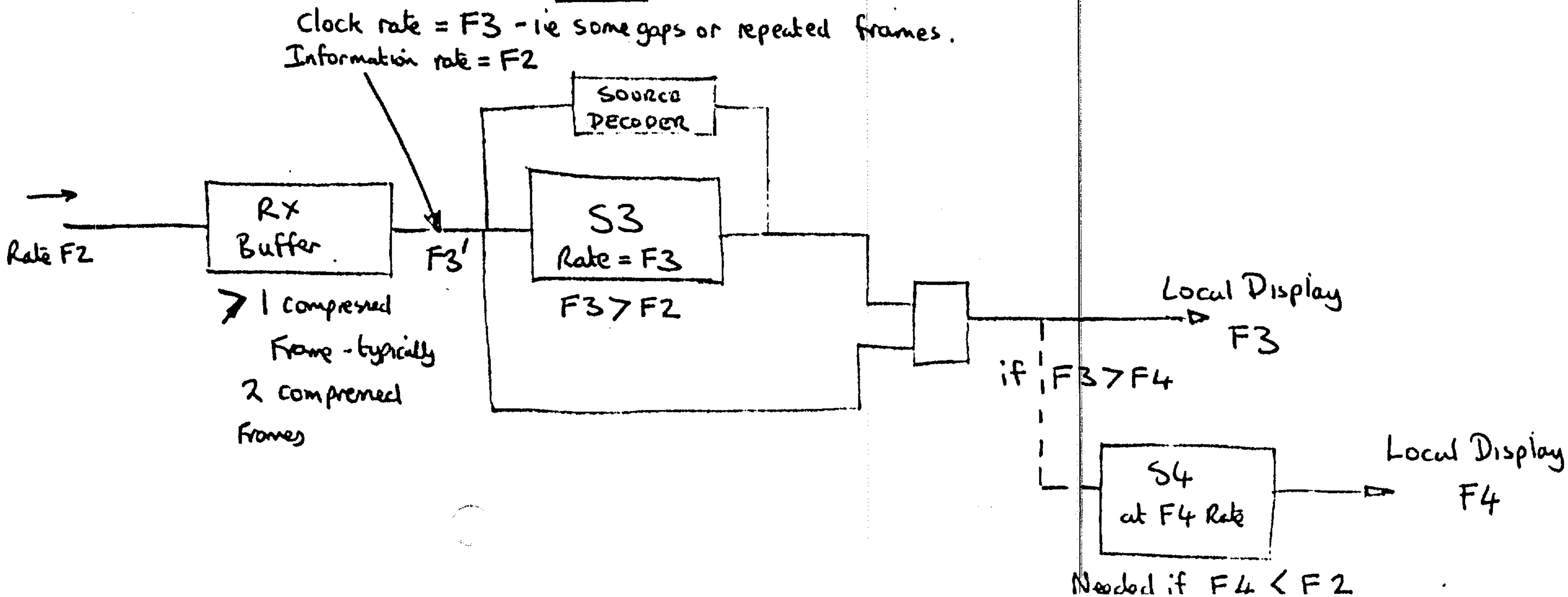

Pre-filter processing
Clock rate F1
Information Rate F2 - ie some frames repeated.
Video Input
field rate F1
S1
SOURCE CODER
S2
Rate = F2
TX Buffer
Rate F2.
≥ 1 compressed frame.
Variable summation for temporal correction.

Fig 2


Clock rate = F3 - ie some gaps or repeated frames.
Information rate = F2
SOURCE DECODER
RX Buffer
Rate F2
F3'
S3
Rate = F3
F3 > F2
≥ 1 compressed Frame - typically 2 compressed Frames
Local Display F3
if F3 > F4
S4 at F4 Rate
Local Display F4
Needed if F4 < F2

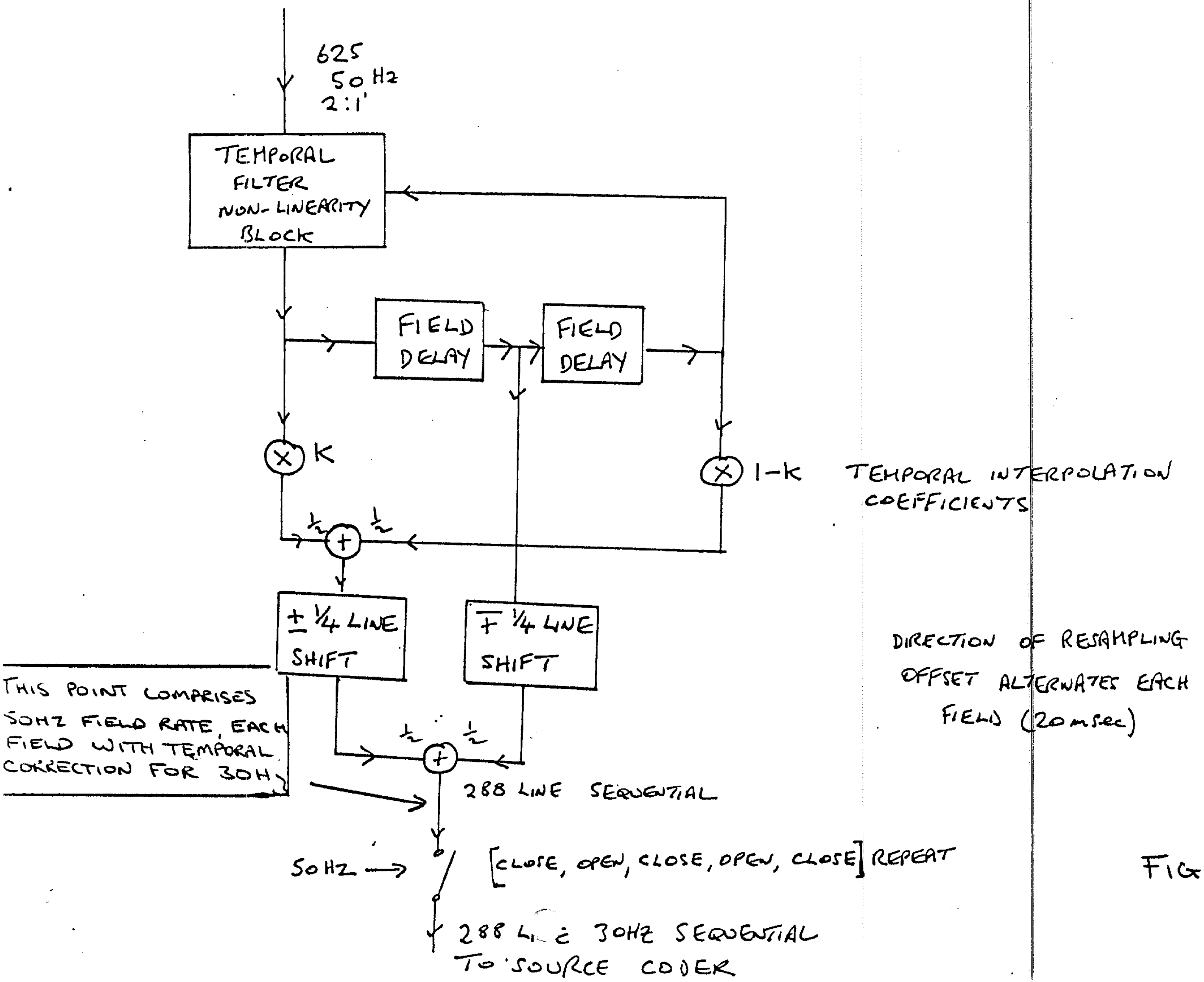
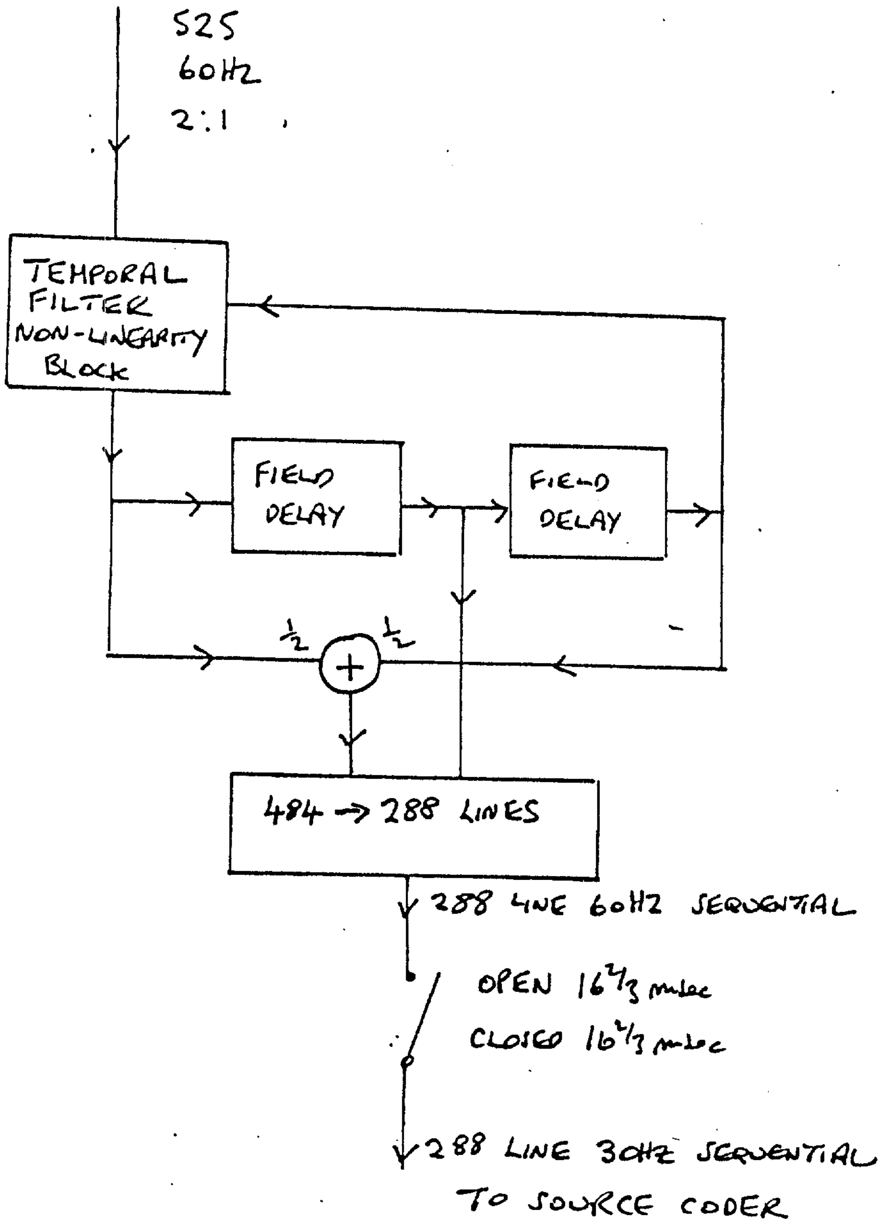

625
50 Hz
2:1
TEMPORAL FILTER NON-LINEARITY BLOCK
FIELD DELAY
FIELD DELAY
K
1-K
TEMPORAL INTERPOLATION COEFFICIENTS
½
½
± ¼ LINE SHIFT
∓ ¼ LINE SHIFT
DIRECTION OF RESAMPLING OFFSET ALTERNATES EACH FIELD (20 msec)
THIS POINT COMPRISES 50HZ FIELD RATE, EACH FIELD WITH TEMPORAL CORRECTION FOR 30Hz
½
½
288 LINE SEQUENTIAL
50Hz
[CLOSE, OPEN, CLOSE, OPEN, CLOSE] REPEAT
288 L.i. 30Hz SEQUENTIAL TO SOURCE CODER

FIG 3


525
60Hz
2:1
TEMPORAL FILTER NON-LINEARITY BLOCK
FIELD DELAY
FIELD DELAY
½
½
+
484 → 288 LINES
288 LINE 60HZ SEQUENTIAL
OPEN 16⅔ msec
CLOSED 16⅔ msec
288 LINE 30HZ SEQUENTIAL
TO SOURCE CODER

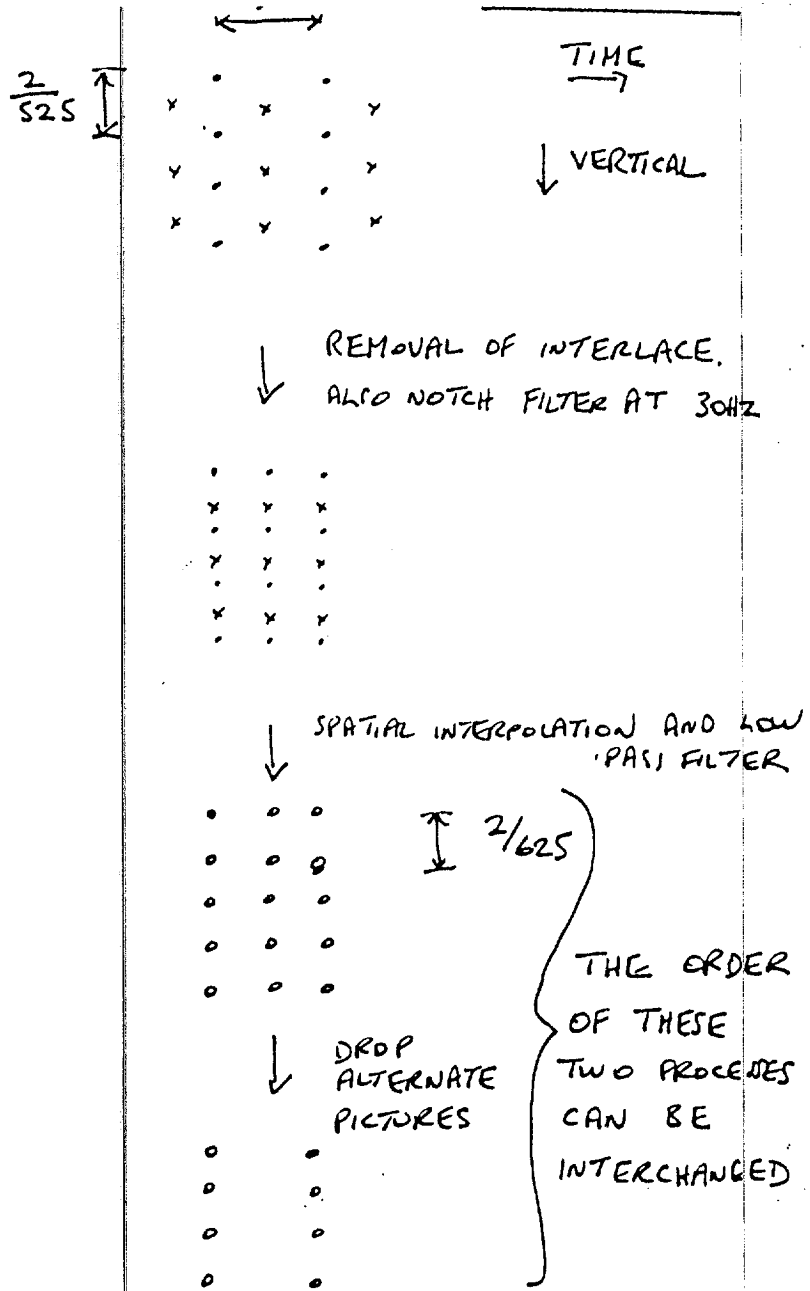

TIME
VERTICAL
2/525
REMOVAL OF INTERLACE.
ALSO NOTCH FILTER AT 30HZ
SPATIAL INTERPOLATION AND LOW PASS FILTER
2/625
DROP ALTERNATE PICTURES
THE ORDER OF THESE TWO PROCESSES CAN BE INTERCHANGED

FIG 4

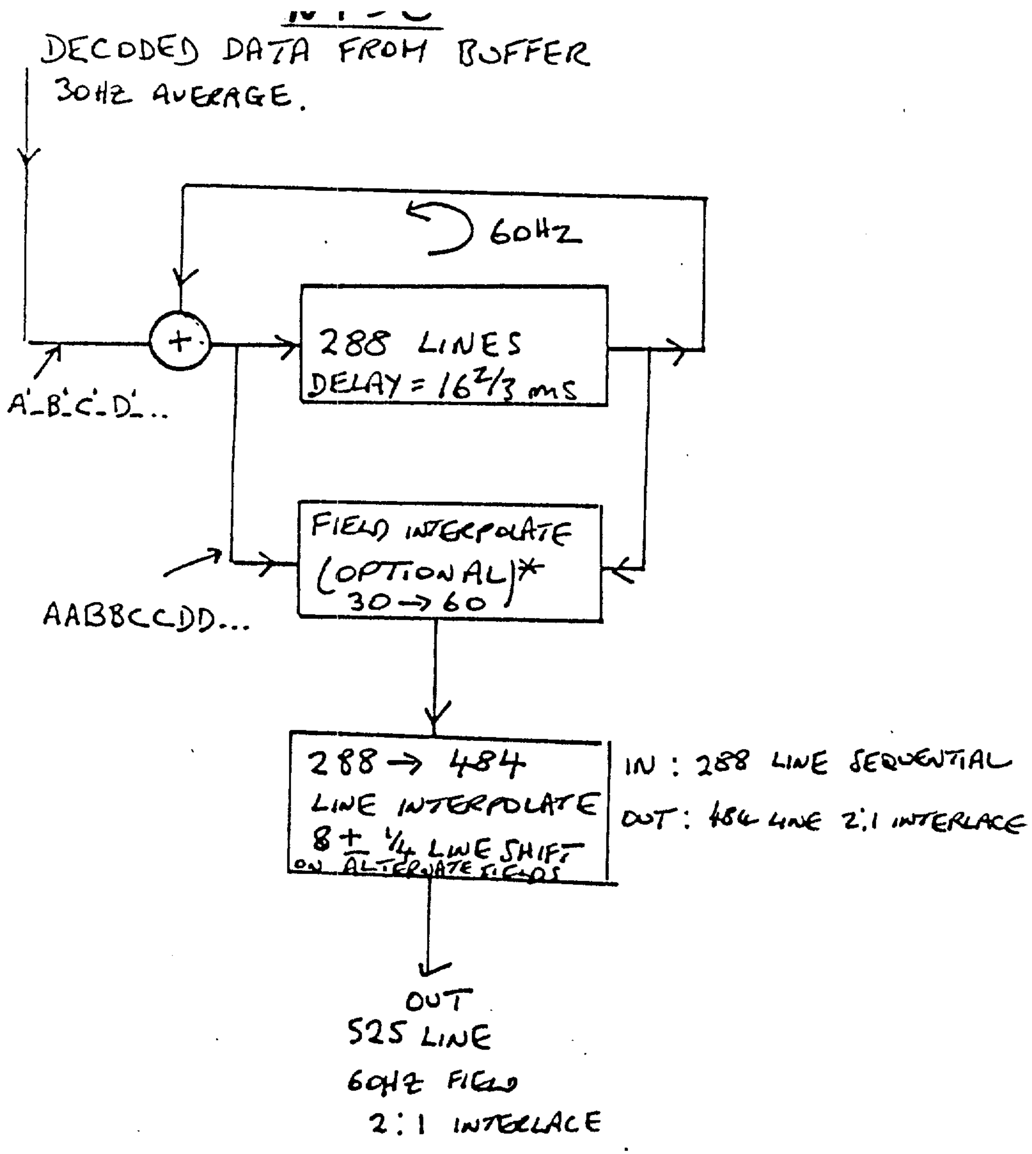


FIG 5

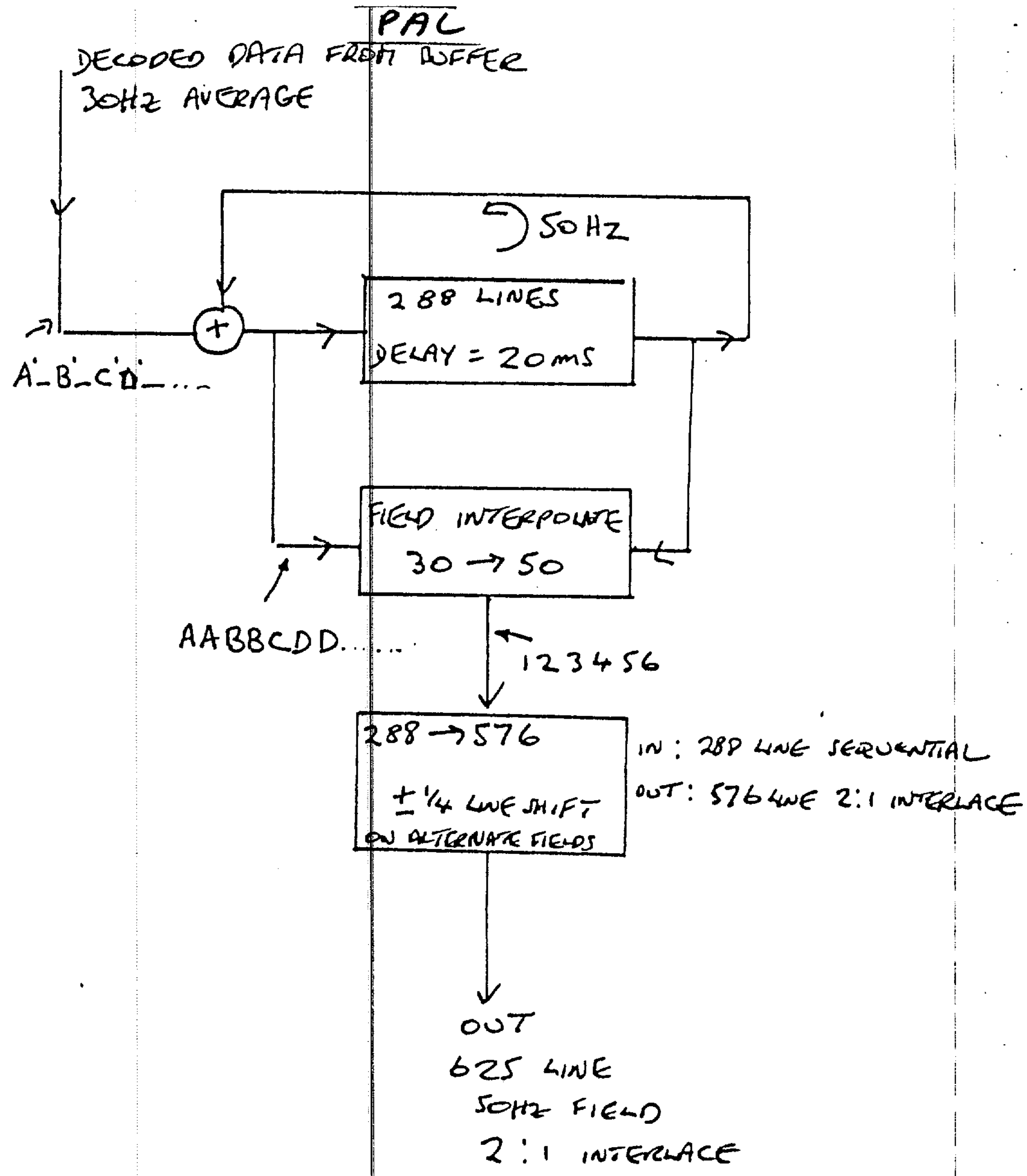


FIG 6

FIELD REPEATING AS PROVIDED BY THE DECODING LOOP SHOULD BE ADEQUATE.

A', B' ETC = UPDATE VALUES

A, B ETC : RECONSTRUCTED PICTURES - 30HZ SAMPLING