# A New 5B/6T Code for Data Transmission on Unshielded Pair Cable

Miranda Mowbray, Alistair Coles,
David Cunningham
Networks and Communications Laboratory
HP Laboratories Bristol
HPL-93-49
June, 1993

codes and coding, digital communications systems

We present a new 5 binary / 6 ternary coding scheme with a similar advantageous spectral property to that of MLT3, but which has bounded running digital sum and maximum run length.

# 1 Introduction

Considerable interest has recently been shown in high speed data transmission (100 Mb/s) over unshielded twisted pair (UTP) cable [1,2,3]. The choice of signalling scheme for UTP cable is influenced by regulatory limits on electromagnetic emissions at frequencies above 30MHz [4]. MLT3 signalling has been proposed since, for random data, the bulk of this signal's energy is at frequencies below 0.25 * baudrate (e.g. below 25 MHz for 100Mb/s).

MLT3 ([1], 7.1.3) is a signalling scheme which encodes a binary string into a ternary string (using the alphabet $+$, $-$, 0). The ternary strings produced contain no substring $+-$, $-+$, $+0^k+$ or $-0^k-$ for $k > 0$, (we call these *forbidden substrings*), resulting in the advantageous spectral property mentioned above.

In MLT3, each binary 1 is encoded as a transition between ternary digits, consistent with the string containing no forbidden substrings, and each binary 0 is encoded as no transition. This encoding has two drawbacks. Among the signals which can be produced are signals with arbitrarily large maximum run length (defined as the maximum number of consecutive identical digits), and signals with arbitrarily large running digital sum (RDS, defined as the number of digits $+$ in the string minus the number of digits $-$). Since UTP cables are transformer coupled and the transmission channel therefore has no response at low frequencies, these signals are extremely susceptible to intersymbol interference in the form of baseline wander ([5], p.319). Furthermore, the lack of transitions in long runs of identical digits has a deleterious effect on clock recovery circuits.

This letter reports a new coding scheme encoding binary strings into ternary strings with no forbidden substrings, which has a bound on the RDS and maximum run length of any signal produced.

# 2 Previous Approach

There is a standard way of designing a block coding scheme which encodes a binary string, (divided into words of length $\ell$) into a ternary string, (divided into words of length $m$), such that the RDS of the encoded signal is bounded. The ternary word corresponding to a given binary word is chosen amongst one or more alternatives in such a way that if the RDS of the encoded signal produced so far is positive, then the next word transmitted will have a non-positive RDS, and if the RDS of the encoded signal produced is negative, then the next word transmitted will have a non-negative RDS. This ensures that at any point in the transmission the RDS is bounded by $\pm 3m/2$.

Unfortunately, the coding schemes produced by this approach either produce signals containing forbidden substrings, or are inefficient, or unwieldy, or have unbounded maximum run length. In the next section we shall prove the following.

**Theorem 1** *Under any 4B/5T block coding scheme using the approach just described, such that the output string produced contains no forbidden substrings, it is possible to produce a signal with arbitrarily large maximum run length.*

The proof method for Theorem 1 yields similar results when applied to 5B/6T, 6B/7T, 7B/8T and 8B/9T coding schemes, giving the following.

**Theorem 2** *Under any block coding scheme using the approach just described, such that the output string produced contains no forbidden substrings, efficiency $m/\ell$ is at least 80% and $m < 10$, it is possible to produce a signal with arbitrarily large maximum run length.*

## 2.1 Proof of Theorem 1

In order to prove Theorem 1, we first introduce some notation.
State $(n,+)$ means that the running digital sum is $n$, and the last digit sent was $+$.
State $(n,+0)$ means that the running digital sum is $n$, the last digit sent was 0, and the last non-zero digit sent was $+$.
State $(n,-)$ means that the running digital sum is $n$, and the last digit sent was $-$.
State $(n,-0)$ means that the running digital sum is $n$, the last digit sent was 0, and the last non-zero digit sent was $-$.

Say that a ternary string $S$ is *permissible from s* iff the concatenation $S'.S$ contains no forbidden substrings whenever $S'$ is a ternary string with no forbidden substrings which leaves the system in state $(n,s)$ for some $n$. (So, for example, $S$ is permissible from state $+$ iff $S$ is a ternary string with no forbidden substrings such that either the first digit of $S$ is $+$, or the first digit of $S$ is 0 and the first non-zero digit of $S$ is $-$.)

The proof is through a series of Lemmas. The first Lemma is rather more general than is necessary to prove Theorem 1, but this generality is useful when extending the proof of Theorem 1 to prove Theorem 2.

**Lemma 1**

If $m > 1$ there are fewer than $2^{m-1}$ MLT-3 strings of length $m$ with non-negative digital sum whose first non-zero digit is $-$.

**Proof**

Let $\mathcal{A}$ be the set of ternary strings which begin with $-$ and end with $+$ and contain no forbidden substrings, together with the empty string. For $A \in \mathcal{A}$, let $A^*$ be the string obtained by reversing $A$ and then substituting $+$ for $-$ and $-$ for $+$. (For example, if $A = --0+$ then $A^* = -0++$.) Now define the map $F$ from MLT-3 strings of length $m$ with non-negative digital sum whose first non-zero digit is $-$, as follows.

- If $C = 0^a.A.0^b.-^c.0^d$ with $A \in \mathcal{A}$, $a,b,c,d \geq 0$, and the digital sum of $A$ is greater than zero, then $F(C) = 0^a.A^*.0^b.-^c.0^d$

- If $C = 0^a.A.0^b$ with $A \in \mathcal{A}$, $a \geq 0$, $b \geq 2$, digital sum of $C = 0$, then $F(C) = 0^a.A^*.0^{b-1}.-$
- If $C = 0^a.A.0^b.-^c.0^d.+^e.0^f$ with $A \in \mathcal{A}$, $a, b, c, d \geq 0$, $e > 0$, $f = 0$ or 1, digital sum of $C = 0$, then $F(C) = 0^a.A^*.0^b.-^{c+e+f}.0^d$

It is straightforward to check that $F$ is a well-defined 1-1 function from the set of ternary strings of length $m$ with non-negative digital sum whose first non-zero digit is $-$ and which have no forbidden substings, to the set of ternary strings of length $m$ with negative digital sum whose first non-zero digit is $-$ and which have no forbidden substrings. Moreover, there is no $C$ such that $F(C) = -^m$. Since there are $2^m$ ternary signals of length $m$ whose first non-zero digit is $-$ and which have no forbidden substrings, the result follows.

The rest of the proof of Theorem 1 is specific to 4B/5T codes, but the proof method can also be applied to codes of other sizes.

Suppose (for a contradiction) that there is a 4B/5T coding scheme using the approach described at the beginning of this section, such that the signals produced contain no forbidden substrings, and have bounded maximum run length. Let S be the set of states $(n, s)$ which can be reached at the end of a codeword. The proof proceeds by lemmas, each proving that certain states are not in $S$, until Lemma 11 shows that $S$ is empty, a contradiction.

**Lemma 2**: $S \subseteq \{(n, s) : n \geq 0 \text{ or } s \neq +0\}$.
Proof: By Lemma 1, the number of codewords of length 5 with non-negative digital sum which are permissible from state +0 is strictly less than 16. Therefore, once a 4B/5T coding scheme reaches a state $(n, +0)$ with $n < 0$, there is a non-negative probability that the running digital sum after the next codeword will be less than $n$. Hence if the coding scheme is to guarantee a bounded running digital sum by the usual method, then the coding scheme must never reach a state $(n, +0)$ with $n < 0$ at the end of a codeword.

**Lemma 3**: $(n, +) \notin S$ for $n < -1$.
Proof: Suppose that it reaches a state $(n, +)$ with $n < -1$. There are 18 codewords permissible from + which have non-negative digital sum, but of these the three codewords 00000, 0-0+0, and +0000 leave the system in state $(n, +0)$ or $(n+1, +0)$, neither of which are in $S$. There are only 15 codewords left, fewer than one for each input word. Therefore $(n, +) \notin S$, as required.

**Lemma 4**: $(n, -) \notin S$ for $n < -2$.
Proof: Similar, using Lemma 3.

**Lemma 5**: $(n, -0) \notin S$ for $n < -2$.
Proof: Similar, using Lemmas 3 and 4.

**Lemma 6**: $S \subseteq \{(0, +0), (1, +0), (2, +0), (-1, +), (0, +), (1, +), (2, +), (0, -0), (-1, -0),$

$(-2,-0)$, $(1,-)$, $(0,-)$, $(-1,-)$, $(-2,-)\}$.
Proof: By symmetrical arguments to Lemmas 1-5, there is no state $(n,-0)$ with $n>0$, $(n,-)$ with $n>1$, $(n,+)$ with $n>2$, or $(n,+0)$ with $n>2$, in $S$. The Lemma follows from this and the results proved so far.

**Lemma 7**: $(2,+),(-2,-),(1,-),(-1,+) \notin S$
Proof: Similar to Lemma 3, using Lemma 6

**Lemma 8**: $(1,+),(-1,-) \notin S$
Proof: Similar to Lemma 3, using Lemmas 6 and 7

**Lemma 9**: $(2,+0) \notin S$
Proof: Suppose that $(2,+0) \in S$. There are exactly 16 codewords with non-positive digital sum which are permissible from $+0$ and take the system from state $(2,+0)$ to one of the states $(0,+0),(1,+0),(2,+0),(0,+),(0,-0),(-1,-0),(-2,-0),(0,-)$. By the previous lemmas, S is contained in this set of states. One of these 16 codewords is 00000. Since there are 16 possible input words, 00000 must be the encoding of an input word w from this state, and the end of this codeword the system is still in state (2,+0). But now the encoding of a string which takes the system to state (2,+0) and then has input word w repeated t times will contain a run of 5t consecutive zeroes. So in order to ensure that the maximum run length in the transmitted signal is bounded, it is necessary that $(2,+0) \notin S$.

**Lemma 10**: $(-2,-0),(0,+0),(0,-0) \notin S$
Proof: Similar to Lemma 9.

**Lemma 11**: $S$ is empty
Proof: By the Lemmas so far, $S \subseteq \{(0,+),(0,-),(1,+0),(-1,-0)\}$. But for each state $(n,s)$ in this set, there are fewer than 16 codewords permissible from $s$ which take the system from state $(n,s)$ to a state in the set. The result follows. □.

## 3 A new coding scheme

We now describe a coding scheme with $\ell=5, m=6$ (and so efficiency 5/6) which ensures that the RDS is bounded, but which is not designed using the above approach. The coding scheme uses the table in Fig.1, which is used to encode a binary data string in the following way.

Set $n=0$. Toss a coin: if it comes up heads then set $s=+0$, and if it comes up tails then set $s=-0$.

After having encoded a binary word, update $n$ by adding the RDS of the encoded word to it, and update $s$ as follows:

- If the encoded word ended $+$, put $s=+$; if it ended $-$, put $s=-$

- If the encoded word ended $+$ followed by at least one 0, put $s = +0$; if it ended $-$ followed by at least one 0, put $s = -0$.
- If the encoded word consisted entirely of 0s, and the current value of $s$ is $+$ or $+0$, put $s = +0$; if the encoded word consisted entirely of 0s, and the current value of $s$ is $-$ or $-0$, put $s = -0$.

To encode a binary word $b$, if $s \in \{+, +0\}$, take the word which is in row $b$ and the appropriate column of Fig.1. If $s \in \{-, -0\}$, take the the word which is in row $b$ and the appropriate column of Fig.1, and replace each occurrence of $+$ in this string by $-$, and each occurrence of $-$ by $+$. The resulting word is the encoding of $b$.

## Some properties of the coding scheme

The signals produced by the coding scheme have bounded RDS. To see why, first observe that each entry in the first two columns either has negative RDS, or has RDS zero and $+$ as its last non-zero digit. It follows that if $n > 0$ and $s \in \{+, +0\}$ then the next change to $n$ will decrease $n$.

Now observe that each word which is obtained by replacing $+$ for $-$ and $-$ for $+$ in an entry in the last two columns either has non-positive RDS, or has RDS 1 and $+$ as its last non-zero digit. Suppose that $n > 0$ and $s \in \{-, -0\}$. Then at the end of the next word either (a) $n$ is decreased, or (b) $n$ is unchanged (in particular $> 0$) and $s \in \{+, +0\}$, so that the next change to $n$ decreases it, or (c) $n$ is increased by exactly 1 and $s \in \{+, +0\}$, from which the next change to $n$ decreases it, or (d) $n$ is unchanged and $s \in \{-, -0\}$ as before. Hence the next change to $n$ either decreases $n$, or increases $n$ by exactly 1 and leaves the system with $s \in \{+, +0\}$ and $n > 0$, from which the next change to $n$ will decrease $n$.

Therefore the value of $n$ is bounded above by $m + 1$, and hence the RDS is bounded above. A symmetrical argument shows that the RDS is bounded below. Any table satisfying the conditions observed on the column entries will also give rise to a coding scheme with a bounded RDS. In this particular case the RDS is always between $\pm 5$.

Although it is necessary to keep track of $n$ and $s$ to encode, this is not necessary when decoding, because each word of 6 ternary digits can be the encoding of at most one 5 bit word. Therefore decoding can be performed by a simple lookup table.

Any signal produced by the coding scheme has at most 11 consecutive zeroes, and at most 7 consecutive $+$ or $-$ digits.

The scheme has a primitive error detection property: single errors in transmission can be detected with a checksum. Use a checksum of three digits, which is the mod-2 digit-wise sum of the binary numbers consisting of the first three digits of each input word. If there is a single error in the transmitted string, this is guaranteed to be detected by the checksum.

No signal produced by the coding scheme contains a forbidden substring. The power spectral densities of signals generated by MLT3 and the new 5B/6T coding scheme are shown in figures 2 and 3 respectively. Clearly, the d.c. component of the signal produced by the new coding scheme is zero, but otherwise the spectral densities are similar. Also shown in figures 2,3 are simulated eye diagrams at the output of an ideal raised cosine channel, modified by a first order low frequency roll off at baudrate/300. While there is some closure of the vertical eye opening at the sampling time, due to baseline wander, this effect is reduced for the new coding scheme.

# 4 Conclusion

We have presented a new 5 binary/6 ternary scheme for encoding a binary string into a ternary string without any forbidden substrings, which has a similar advantageous spectral property to that of MLT3, and which has bounded RDS and maximum run length. Such a scheme cannot be found using the standard construction method for block coding schemes with bounded RDS.

# 5 References


[1] ANSI X3T9.5 FDDI TP-PMD working draft, 1992

[2] SAUNDERS, S.: 'FDDI over copper: which cable works best?', *Data Commun. Int.*, 1991, **20**, (16), pp.57-64

[3] POWELL, D.: 'Today's technology enriches yesterday's cable', *Networking management*, 1991, **9**, pp.60-68

[4] Federal Communications Commission (USA) FCC 15J

[5] BYLANSKI, P., and INGRAM, D. G. W.: 'Digital Transmission Systems', (Peter Peregrinus Ltd., London, 1980), IEE Telecommunications Series 4

| | $s=+, n\geq 0$ <br> $s=-, n\leq 0$ | $s=+0, n\geq 0$ <br> $s=-0, n\leq 0$ | $s=+, n<0$ <br> $s=-, n>0$ | $s=+0, n<0$ <br> $s=-0, n>0$ |
|---|---|---|---|---|
| 00000 | 00000- | 00000- | 00000- | 00000- |
| 00001 | 0000-0 | 0000-0 | 0000-0 | 0000-0 |
| 00010 | 000-00 | 000-00 | 000-00 | 000-00 |
| 00011 | 00-000 | 00-000 | 00-000 | 00-000 |
| 00100 | 000-0+ | 000-0+ | 000-0+ | 000-0+ |
| 00101 | 00-00+ | 00-00+ | 00-00+ | 00-00+ |
| 00110 | 0-0000 | 0-0000 | 0-0000 | 0-0000 |
| 00111 | 0--0++ | 0--0++ | ++0-00 | 0--0++ |
| 01000 | 0-0+00 | 0-0+00 | 0-0+00 | 0-0+00 |
| 01001 | 0-00+0 | 0-00+0 | 0-00+0 | 0-00+0 |
| 01010 | 0-000+ | 0-000+ | ++0-0+ | 0-000+ |
| 01011 | +00--0 | ---000 | +00--0 | -00++0 |
| 01100 | 0-0+0- | 0-0+0- | 0-0+0- | 0-0+0- |
| 01101 | 00-0+0 | 00-0+0 | ++00-0 | 00-0+0 |
| 01110 | +000-- | ----0+ | +000-- | -000++ |
| 01111 | 000--0 | 000--0 | 0-00++ | 0-00++ |
| 10000 | +0--00 | ---00+ | +0--00 | -0++00 |
| 10001 | 0000-- | 0000-- | +00-0+ | -00+0- |
| 10010 | 00---0 | 00---0 | 0-0+++ | 0-0+++ |
| 10011 | 00--0+ | 00--0+ | +000-0 | -000+0 |
| 10100 | +0--0+ | ---0++ | +0--0+ | -0++0- |
| 10101 | 000--- | 000--- | 00-0++ | 00-0++ |
| 10110 | 00--00 | 00--00 | 0-0++0 | 0-0++0 |
| 10111 | 0---0+ | 0---0+ | +0-000 | -0+000 |
| 11000 | 0---00 | 0---00 | +00-00 | +00-00 |
| 11001 | 0--00+ | 0--00+ | ++00-- | --00++ |
| 11010 | 0--0+0 | 0--0+0 | +0-0+0 | -0+0-0 |
| 11011 | +0---- | --0000 | ++0000 | -0++++ |
| 11100 | 0--000 | 0--000 | +0-00+ | -0+00- |
| 11101 | +00--- | ---0+0 | +++0-0 | -00+++ |
| 11110 | +0---0 | --000+ | ++000- | -0+++0 |
| 11111 | 000000 | --0++0 | ++0--0 | --0++0 |

Figure 1: *Table for the coding scheme*

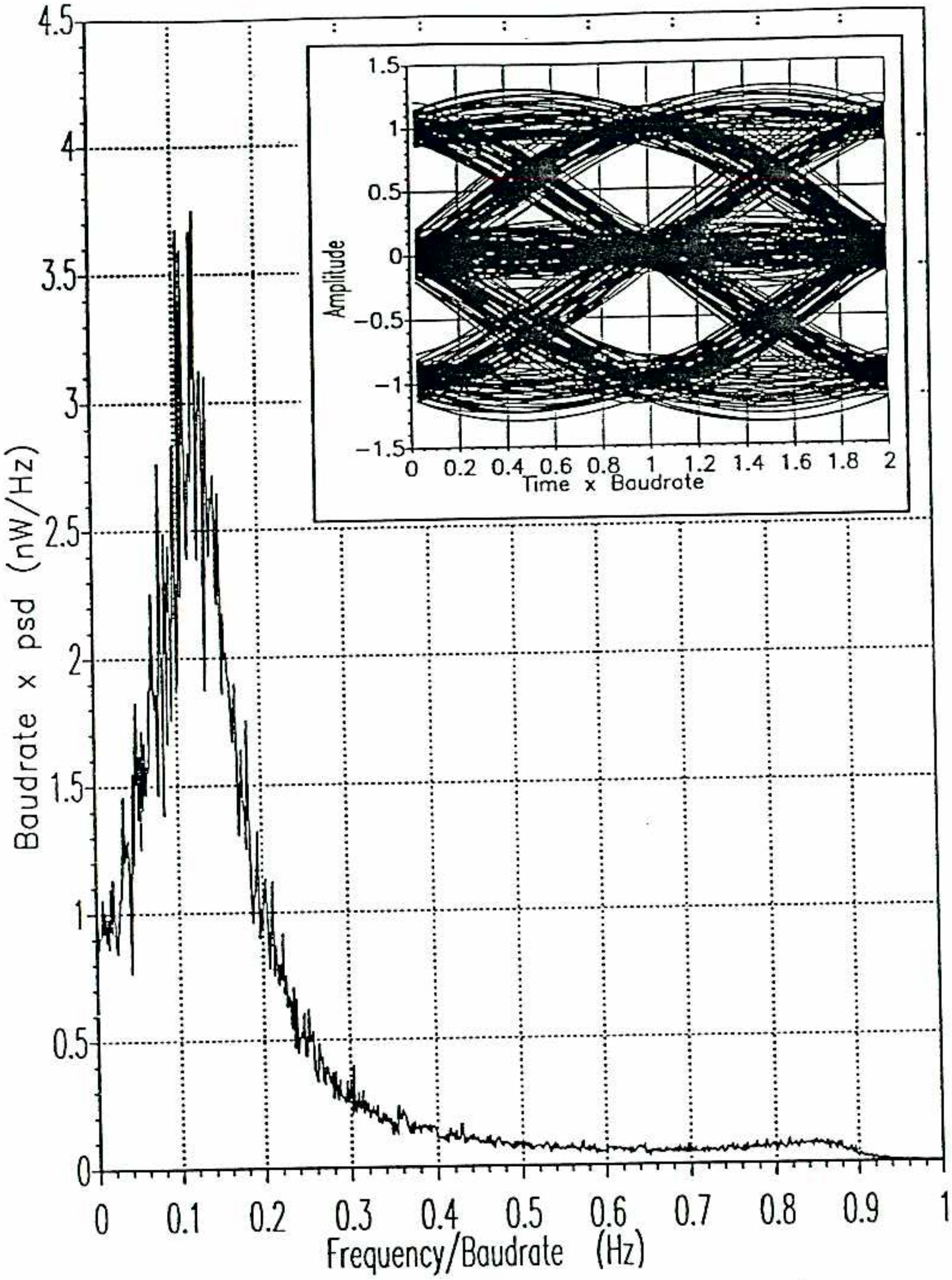


Figure 2: *Power spectral density and (inset) eye diagram of MLT3 signal*

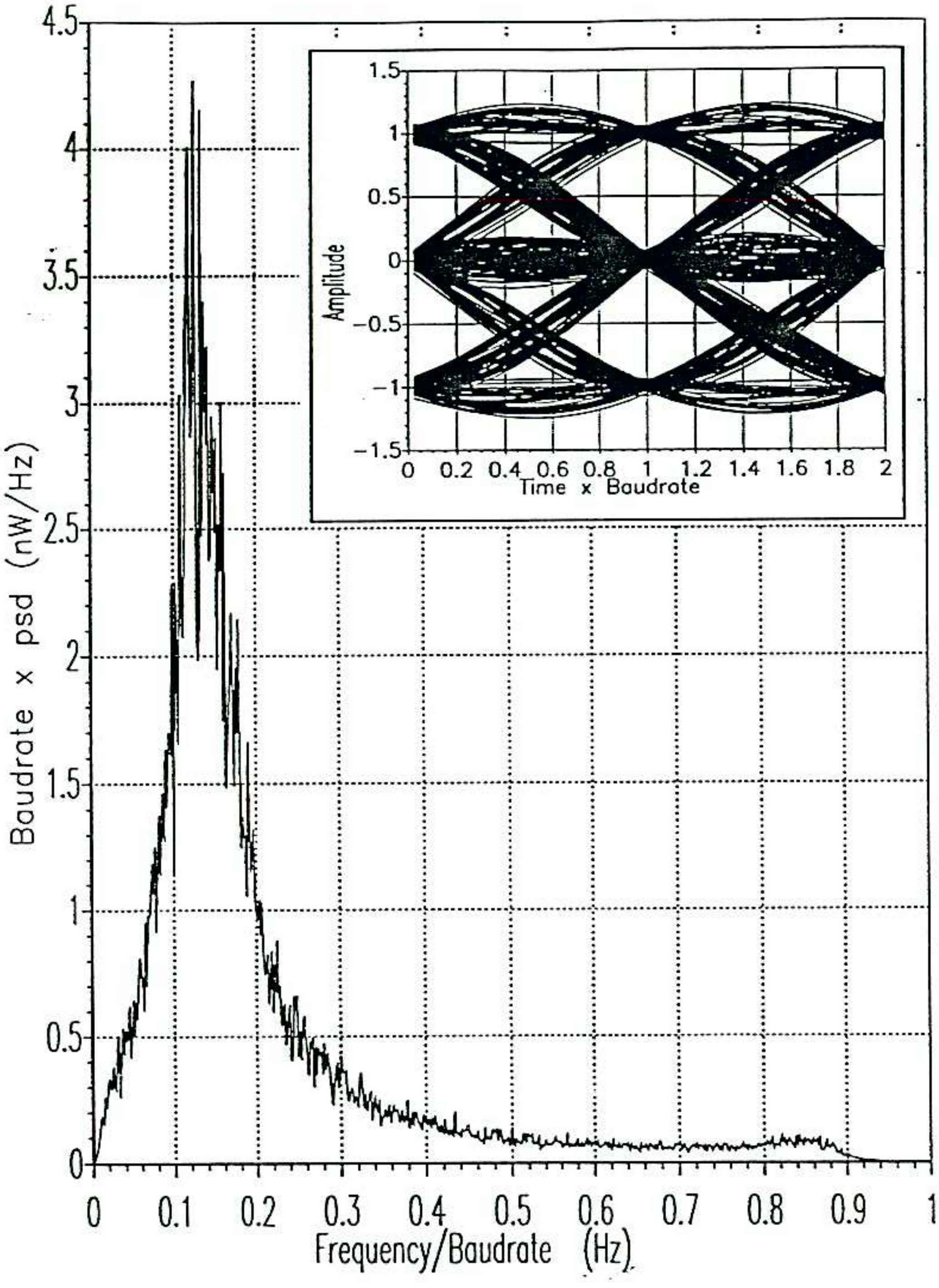


**Figure 3:** *Power spectral density and (inset) eye diagram of new 5B/6T signal*