

THE SHAPE OF BITS TO COME

James Miller BSc, G3RUH

3 Benny's Way
COTON
Cambridge
CB3 7PS
England

ABSTRACT

This tutorial article is about bits, about bandwidth and about control of both. Lately a number of expressions have crept into amateur radio data transmission, creating both interest and puzzlement. What, for example, is the "raised cosine modulation" used on microsats AO-16/18/19? And what is the RSM-8 on Rudak/AO-21? What PSK modulation does Fuji-20 use? Come to that, what is PSK or BPSK? Read on.

Let's come clean at the outset; "raised cosine", RSM-8, PSK and DPSK are the same thing. They're all forms of PSK - "phase shift keying".

In the amateur environment we are usually trying to send data as fast as possible within a limited RF bandwidth. This limit may be set by the receiver alone, or it may be due to some statutory, aesthetic or technical requirement. So a very important part of data transmission concerns the data's spectrum.

UNDERSTANDING PSK

Let's begin with a statement and explain later:

PSK spectral considerations at RF can be analysed by examining the characteristics of an isolated bit at "DC" or baseband.

I have to start somewhere, so I'll make the assumption that you have in your head an image that "data" is the sort of regular signal you have running down a piece of cable. On a 'scope it looks like random highs and lows spaced at regular intervals. Each of these elements is called a "bit", and they flow at the "bit rate", described

as 9600 bits/sec or sometimes "9600 baud" depending on the context. See the waveform "Data In" of fig. 1.

There are many ways to modulate this data onto an RF carrier, for example ON/OFF keying, carrier FSK, two-tone AFSK and so on. Here I am concerned with double sideband modulation, DSB. That is, the data signal drives a balanced modulator whose other input is the RF carrier. The output then is simply a frequency shifted replica of the input data. (Remember what happens when the input is voice?)

Now binary data has only TWO states. It is either a "1", represented by +1 volt say, or a "0", represented by -1 volt. And that means that the modulated RF carrier also

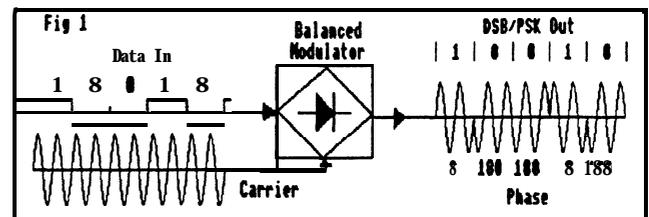


Fig 1. PSK is usually generated by a DSB balanced modulator. So DC spectrum is simply shifted to RF. Data "1"s give carrier phase 0°, data "0"s give phase 180°.

takes only TWO states; either it has phase 0 degrees or 180 degrees. This is because compared with the +1 volt level, the -1 volt inverts the carrier. That means a 180 degrees phase shift. Voila! Binary Phase-Shift-Keying or BPSK. The term PSK, often used, is actually imprecise as it embraces other shifts as well as 0/180 degrees. "Antipodal PSK" is correct but long-winded.

So, if we know the spectrum of the "baseband" data signal, then the RF spectrum is just its replica above and below the carrier frequency, because it's DSB, double sideband suppressed carrier.

If in addition, the data signal has amplitude variations, then these will translate into identical amplitude variations at RF.

This is a very important equivalence, and it's worth restating:

"Binary data DSB modulating an RF carrier" and "binary phase shift keying (BPSK)" are exactly the same thing. If we want to control or analyse the RF spectrum characteristics, we only need to control or analyse the source data characteristics.

UNDERSTANDING DATA STREAMS AND ISOLATED BITS

We now make a reasonable assumption. The data stream consists of random bits. If the data is random, for the purposes of analysis we don't then need to know anything about the content of the message stream. All we need to know is the properties of ONE ISOLATED BIT. Whatever properties we can discover about the isolated bit will also apply to the average summation of the random assortment of them that make up the stream.

so in summary, our problem of spectral considerations at RF can be analysed via the relatively simpler job of examining the characteristics of an isolated bit at source level, DC or "baseband".

As a communications theory aside, it is worth noting that with efficient communications, that data MUST in fact be

random. For if it were not, that would imply something systematic about it. So the data could actually be represented or coded more efficiently, resulting in less transmitted bits, which would then have a random distribution.

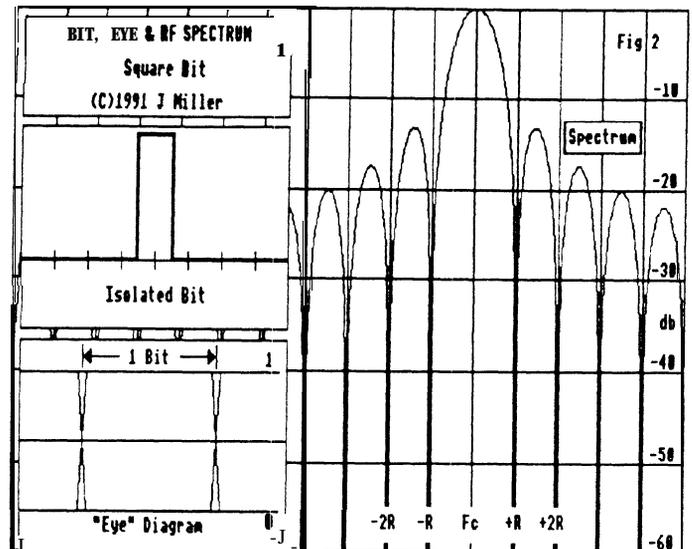


Fig 2. Rectangular bit, its eye and spectrum. See text for details.

UNDERSTANDING BIT, EYE & RF SPECTRUM PLOTS

Now let's examine some representative cases. What happens if we modulate the carrier directly with unprocessed rectangular bits?

Fig.2 shows three things a) isolated rectangular bit, b) its RF spectrum and c) what the data stream looks like on an oscilloscope. Of course the bit edges are in reality almost invisibly fast, so I have emphasised them slightly.

a) Isolated Bit. The horizontal axis is time, and is marked off in eight bit intervals. (Eight are shown). The vertical is voltage.

b) RF Spectrum. The horizontal axis is frequency, and is marked off in units of the bit rate R bps. For example, if the data rate is 1200 bit/sec (R = 1200), the vertical markers are at 1200 Hz spacing. The mid point of this spectrum is of course the RF carrier frequency. The vertical is marked off at 10 db intervals.

c) Oscilloscope diagram.

1. The horizontal axis is time, and spans a total of two bits; so it's expanded x4 compared with the isolated bit. The vertical is voltage. You are to imagine the 'scope is triggered by a local bit-rate clock. The resultant pile-up of bits is what you see displayed. It is the sum of many positive and negative isolated bits, each displaced in time by one bit.

2. If this display is the received waveform, it's what is sampled by the data detector to decide if a "1" or a "0" has been received. Usually this decision is taken at the mid-time point of the received bit by a sampler. Above the horizontal line is the regime of "1"s, below "0"s.

3. This display is also universally known as an "eye diagram". It is a concise representation of the quality of the received signal. The less confused the trace, the greater the distinction between high and low at the mid-bit sample point, then the more reliable (less error prone) bit detection will be.

4 These traces show no channel noise. That is to say, the signal to noise ratio is assumed high. If there were noise, the traces would be blurred. If the noise were momentarily large enough to cause the waveform accidentally to cross the middle horizontal line, an erroneous detection decision would be made, leading to a received bit error.

RECTANGULAR BIT

Returning to our rectangular bit shape. Two things are clear from figure 2; first, the extremely high fidelity of the data waveform and second, the price for this, profligate use of bandwidth. The diagram shows significant energy spreading well beyond 10 times the data rate. Indeed, as far away as 30 R (R = bit rate), the sidelobes are still only -40 db down. As sufferers from computer hash will know, rectangular bits are effective noise generators.

In some applications (for example spread spectrum communications) this might be the desired result. In the amateur radio environment where we want to get as much data through the limited bandwidth of our receivers, more finesse is required

Obviously we need to filter the data stream before transmission, because this will attenuate those sidelobes, and constrain the bandwidth.

SO the key issue, indeed perhaps the central problem of all data transmission, is "what form should this filter take"?

PERFORMANCE OF A SIMPLE R-C NETWORK DATA FILTER

We tentatively try a simple filter. Let's pass the data stream through an R-C network with a -3db point equal to half the data rate. The result is shown in figure 3.

Look at the spectrum; while the main lobe is hardly changed, the sidelobes are reduced by 10 db or more, which is what we wanted. Look at the isolated bit; it has become rounded. But note particularly that its duration now exceeds one bit. In fact it's roughly two bits long.

This means that successive bits overlap. In turn this brings a new potential design problem; inter-symbol interference, or ISI. Now we find that not only do we want spectrum control, but we have to do it with the constraint that successive overlappir.. bits must somehow not interfere with each other.

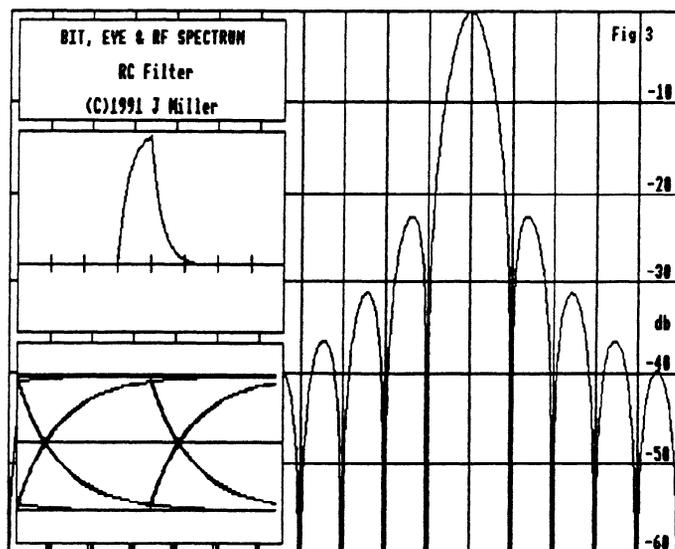


Fig 3. Rectangular bit through R-C filter has classic exponential rise and fall, and RF sidebands are reduced slightly.

Finally, look at the 'scope trace. The line at the top corresponds to the sequence 111..., that at the bottom 000... The sweeps from top to bottom are caused by ..10.. and ..01.. transitions and so on. You can also see that the traces are not confused; highs and lows of successive bits are separable, so there is no significant inter-bit interference.

INTER-SYMBOL INTERFERENCE REVEALED

In an attempt to attenuate the sidelobes more, let's now double the RC filters time constant. That is, the -3db frequency is now 1/4 the bit rate (R/4). See Figure 4.

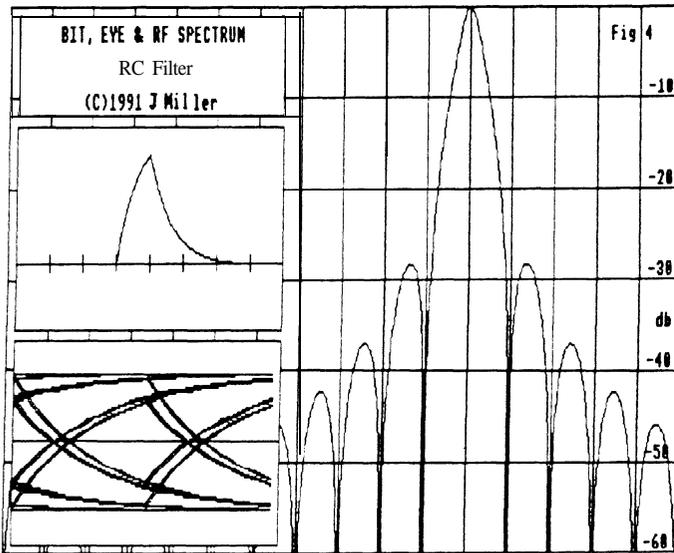


Fig 4. But too much ill-designed filtering gives inter symbol interference (I.S.I.).

Though the sidelobes are reduced, the main lobe is also rather pinched. But the real casualty can be seen in the 'scope trace. Under certain conditions (for example the sequence ..010.. or ...101...) the voltage barely gets half way up to the top line. Sure, bits are still distinct and detectable, but the noise margin is drastically reduced by some 50%.

Analytically, the reason for this is seen in the isolated bit diagram of figure 4. The bit starts off at T=0. At T=1 it reaches its peak. At T=2, it hasn't fallen back to zero again, but has a value approximately 1/3rd peak. This aberration is the sole cause of the inter-symbol

interference apparent in the 'scope trace. At T = 3,4 etc, the voltage is back to zero again.

From this observation we can formulate the requirement for a bit shape that guarantees no ISI. There should be a peak of unity at T=0. Then at all other exact bit points (T= -2, -1, +1, +2, etc) the isolated bit waveform should be exactly zero. What happens in between doesn't matter.

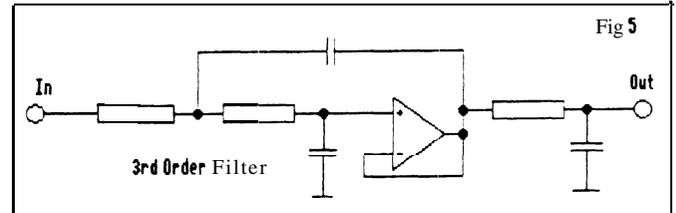


Fig 5. Packet satellite FO-20's 3rd Order Data filter is similar to this. Waveshapes as figure 6.

FO-20's DATA FILTER

An example of a bit shape that meets this requirement is that transmitted by the satellite Fuji-Oscar 20. The data filter is (I believe) a 3rd order Bessel type with a -3db point at 0.532 times the bit rate (0.532 R). "Bessel" filters driven by square waves have a nice steady rise, and negligible overshoot or ringing. In electrical circuit terms, the filter is merely 3 resistors, 3 capacitors and an op-amp. The filter is shown in figure 5.

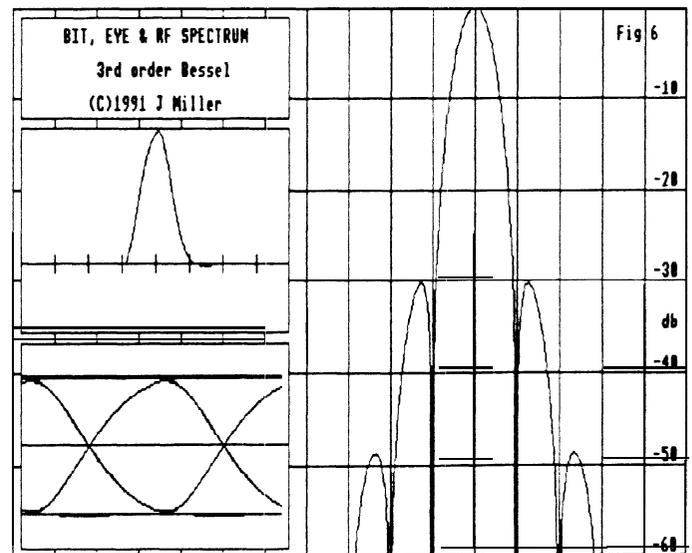


Fig 6. Characteristics of FO-20 Data

The spectrum is quite reasonably controlled; 99.9% of the energy is contained within a bandwidth of 1.75 R (as compared with 15 R of the rectangular pulse). Since the FO-20 bit rate is 1200 bps, this means that most of the signal occupies an effective RF bandwidth of the order of 2.1 kHz, so it stands a fair chance of passing through the SSB receiver filters without significant distortion.

The isolated bit shape is zero at all T mints except at one, and this leads to a nice 'scope trace. The origin of the term "eye diagram" can be seen here, as the trace is supposed to resemble a human eye. Notice that as a consequence of negligible inter-symbol overlap in the isolated bit, the 'scope trace has perfect bit convergence at the sample point.

The isolated bit has a duration of just 2 bits. This means that there can be only $2^2 = 4$ trajectories on the 'scope trace (00,01,10, and 11). The 01s and 10s show up as a very clean zero crossover. If the RF bandwidth were narrower then the isolated bit would span more than 2 bits, so the 'scope trace zero crossings would show dispersion.

AND NOW - "RAISED COSINE"

The term "raised cosine" means several things, a fact which causes a lot of confusion. A raised cosine is just that. A cosine function raised above its mid point. The formula for a raised cosine is

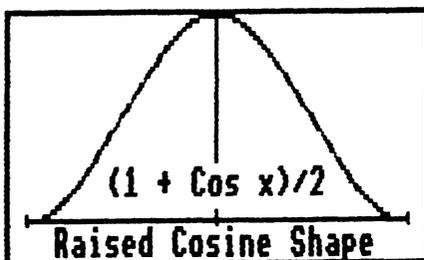


Fig 7a. A raised cosine shape. There's nothing mysterious about a raised cosine. Just get out your calculator and plot $y = (1 + \cos x)/2$ for $x = -180$ to 180 degrees!

$(1 + \cos x)/2$. Drawnasa graph, it's merely a shape. But it has a number of analytical properties that make it convenient to employ. These are simplicity, left-right and upper-lower symmetries.

A RAISED COSINE BIT

"Raised cosine" can for example, describe the shape of the isolated bit - see figure 7. Since the isolated bit is cosine shaped, the 'scope trace is composed entirely of sine waves and straight lines. Indeed, the bit sequence .10101010. actually creates a pure tone at a frequency R/2 at baseband, or at RF, two frequencies spaced by $\pm R/2$.

The isolated bit shape is very similar to FO-20's, occupying 2 bits. Not surprisingly the spectrum is very similar too. In fact 99.9% of the energy is contained within 1.69 R (compared with 1.75 R for FO-20). An important practical point to remember is that the shape of the bit is completely specified in time. This means that it can be precisely generated using real finite hardware. The microsats AO-16/WO-18/LO-19 use a filter not unlike this.

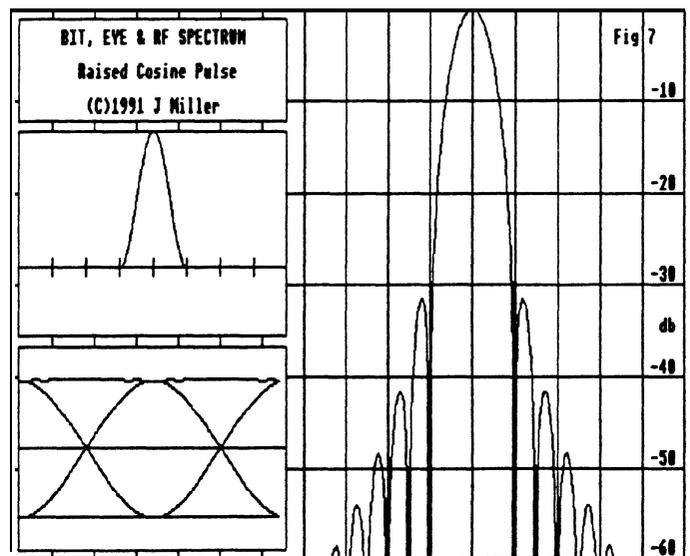


Fig 7. Raised cosine bit TIME shape used (in principle) on PacSats AO-16/WO-18/ LO-19.

THE RAISED COSINE SPECTRUM

Alternatively "raised cosine" can also be used to describe the shape of the spectrum. This is illustrated in figure 8. It doesn't look much like a raised cosine spectrum because of the logarithmic scale, but the response is unity (0db) at $f=0$, 0.5 (-6db) at $f=R/2$ and exactly zero at $f=R$, the bit rate. 99.9% of the energy is contained within $1.56 R$.

Because the spectrum is absolutely band limited to $\pm R$, the isolated bit shape that gives rise to that spectrum has infinite time duration. This means that it cannot be synthesized with real finite hardware. On the other hand, as figure 8 shows, it has negligible amplitude beyond $T = \pm 2$ bits, so that discarding the tails has little effect other than to bring in some very minor spectral sidelobes. So synthesis turns out to be highly practical.

You will also notice that the isolated bit also has precisely the desired properties for zero inter-symbol interference. Is this fortuitous? Certainly not! In fact it is a direct consequence of the two symmetries of the cosine shape mentioned earlier. That's why the raised cosine spectrum is widely used in data transmission system design.

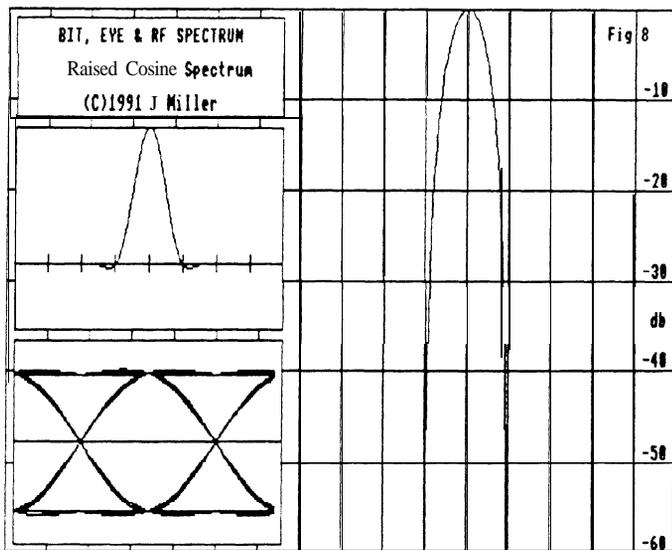


Fig 8. Raised Cosine SPECTRUM Shape. There is a family of these shapes, with the raised cosine part centred on the middle of each side of the spectrum.

9600 BAUD PACKET RADIO & UCSAT 14 -A NARROWBAND RAISED COSINE SPECTRUM

As a practical example of "raised cosine" spectrum shaping consider figure 9. Here only the middle 3/8ths of each half of the spectrum is given a raised cosine shape. From $f=0$ to $5/16 R$ the spectrum is flat (0db), from $5/16 R$ to $11/16 R$ it follows the cosine shape, and from $11/16 R$ onwards the spectrum is zero.

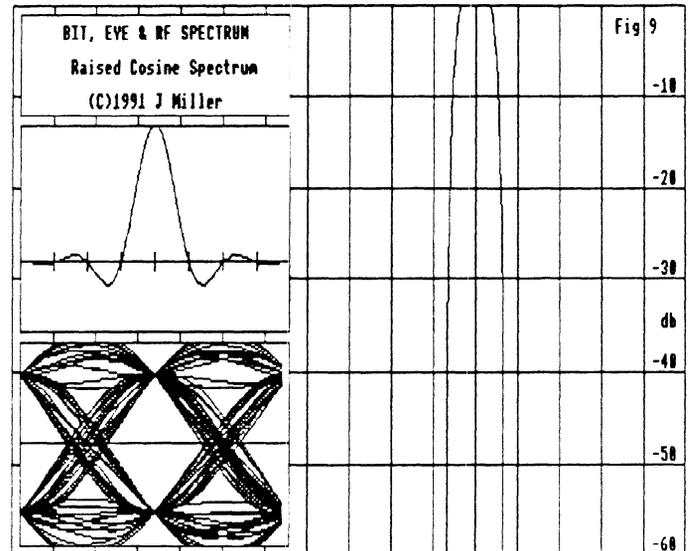


Fig 9. UCSAT 14 has almost "rectangular" raised cosine spectrum (see text). The narrow single main spectrum lobe implies a bit that is some 8 bit periods long. There is no I.S.I.

This is used in the baseband filtering of UCSAT-3/ Oscar-14 which uses the G3RUH 9600 baud Packet Radio modem [1]. The isolated bit shape spans some eight bit periods. That is to say a bit transmitted "now" will actually have an influence on up to eight other bits, four before and four after "now". Note also that the isolated bit has the desired form for minimising inter-symbol interference. A peak at $T=0$, zero at all other T points. So the received "eye" shows good convergence at the sample point.

The spectrum is almost rectangular ("brick-wall") and sidelobe free. 99.9% of the spectral energy is contained within $1.2 R$, or ± 5.6 kHz at 9600 bps.

The zero-crossing dispersion is a direct consequence of the narrow bandwidth, and great care is needed in the design of the receiver bit-clock recovery circuits that perform time bit detection. You can see that sampling only 1/8th bit away from the convergence point means bit detection half way down to the zero threshold, and so the instantaneous noise tolerance would be reduced by 6db. Narrow bandwidth bit-clock recovery circuits are essential. For implementation details see [1].

The actual signal transmitted by UO-14 is slightly different to that shown; it is pre-distorted (in time) so that when it passes through a real bandwidth limited receiver the final "eye" is as illustrated.

In the G3RUH/PacComm/Kantronics/Tasco Tereleader embodiment of this modem, transmit bit-shape is looked up from a table of values in an Eprom and passed to a DAC. This data filter is known as an F.I.R. for Finite Impulse Response. UO-14 uses a tapped digital delay line (shift register) and summing op-amp; Eproms do not travel well in spacecraft. This mechanisation of an F.I.R. is called a Transversal Filter.

In UO-14 the signal also EM modulates the carrier (as opposed to PSK). FM is simple to generate, much simpler to process on receive in the presence of mistuning such as doppler shift, and is very robust. But

FM does require slightly stronger signals, 6-8 db perhaps.

Finally note that this precision data shape could trivially generate precision BPSK via a DSB balanced modulator. Such PSK would have all the narrow-band and non-ISI attributes detailed above.

TRANSMITTER AND RECEIVER FILTERING IN SERIES

So far discussion has mostly been about spectra and bit shapes in association with transmitters. It was implicit that the receiver was infinitely wideband. For analytic purposes, if there were any RX filtering it could be assumed to be merged with the transmit filtering.

In many amateur applications this is indeed the case. For example, the current flock of PSK packet satellites confine their bit rates to that which can be passed through a typical SSB receiver. RTTY users assume the same thing.

We now ask the important question; if the overall filtering were to be deliberately split between transmitter and receiver, what should the form of the filters be?

We know from all the above what kind of overall response is required. But how should we partition it?

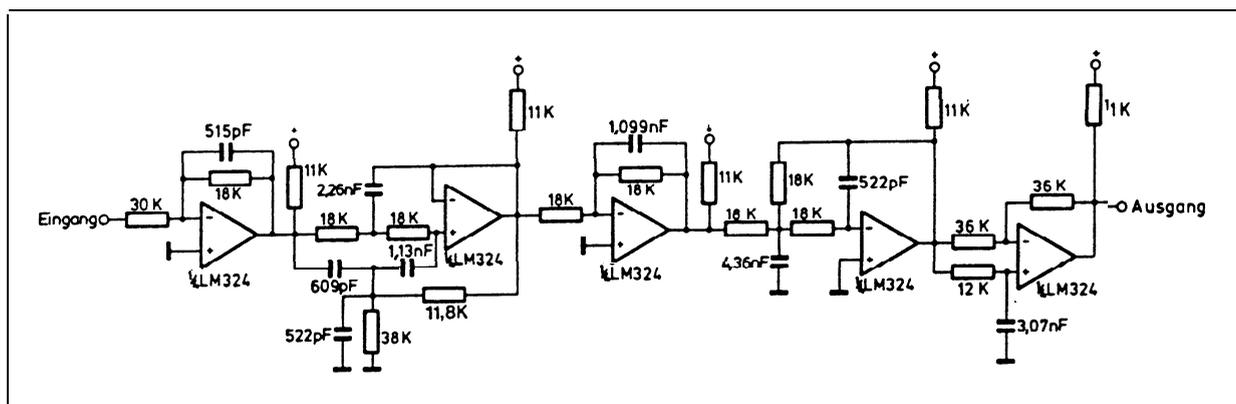


Fig 10. Data filter used in RUDAK-2 "RSM" transmitter. This comprises three cascaded sections; a 3rd order Cauer, a 3rd order Butterworth and an all-pass (refer to text). This transmit filter is used in association with an almost identical filter at the receiver. These filters are thus "matched". For spectrum and bit shapes see figure 11.

Obviously, the overall response can be apportioned in any ratio whatsoever. However one can prove that in the presence of random noise the optimum split, so as to maximise detector signal-to-noise ratio, is that with exactly half the frequency response in the transmitter, and the other half in the receiver. These are known as a "matched filter pair".

Conceptually, one can first imagine a desired overall frequency response dictated by inter-symbol interference requirements, spectral constraints and implementation ease. Then as it were, you take the "square root" of this, and put the resulting filter into both transmitter and receiver.

AO-21/RUDAK-2 "RSM" BIT SHAPING

A very good example of this is the bit shaping scheme used on one of the RUDAK-2 links of AO-21 [2]. Wittingly dubbed "RSM" Rectangular Spectrum Modulation by its creator, its transmit filter properties are shown in fig. 11 (left). Both transmitter and receiver have essentially identical

filters. This bitshape is used to phase modulate (BPSK) the RF carrier. Bit rate is $R = 9600$ bps.

99.9% of the spectral energy is contained within a bandwidth of $1.5 R$, which is consistent with the isolated bit span of about 3 bits. This means that there are about $2^3 = 8$ trajectories in the scope trace, which can be seen quite clearly. It is notable that this isolated bit shape is created using analogue filtering rather than more accurate and repeatable digital synthesis.

The transmit filter (fig. 10) has 3 sections, comprising a) a Cauer Chebyscheff type CC0315/41 which is flat up to a frequency $f = R/2$, and then plunges down to the zero at $f = R$. But above this there is a rise to only -15 db. So part b) is a 3rd order Butterworth that's smooth to the -3db point at $f = R/2$, and then -18db at $f = R$ and -36 db at $f = 2R$; c) an "all-pass" network that has a dead flat frequency response, but a tapered phase characteristic which is used to flatten out the filter delay vs. frequency curve. This makes the transmit bit shape become

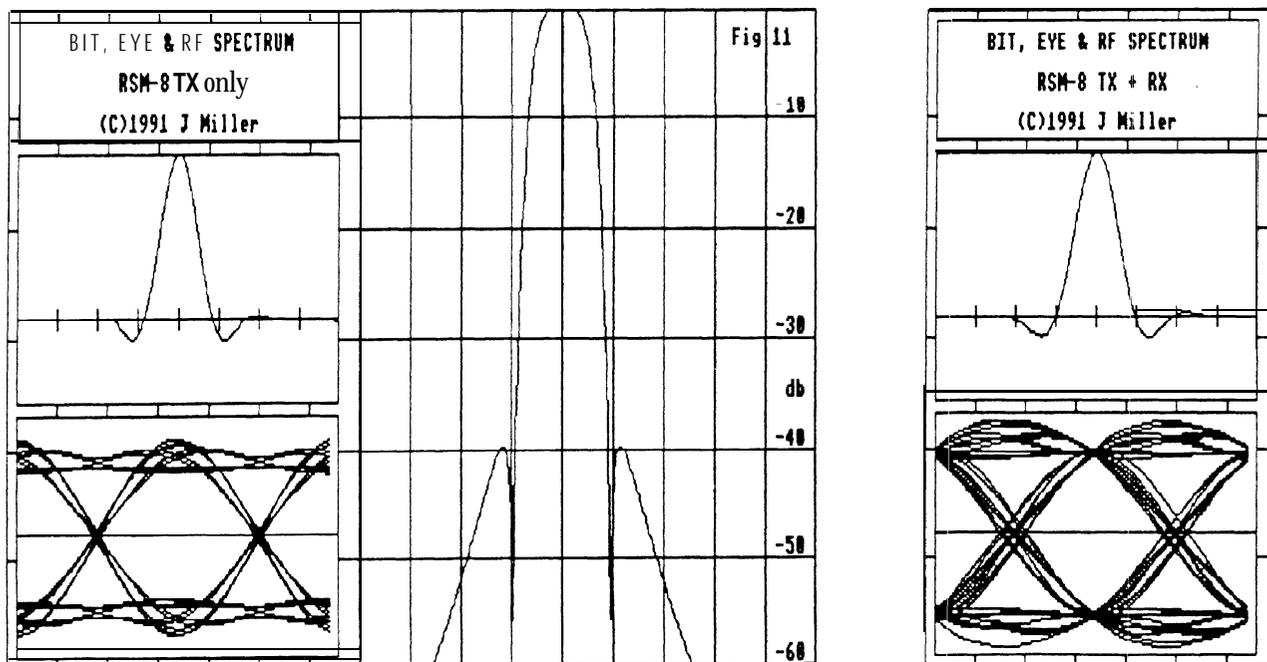


Fig 11. Performance of RUDAK-2's "RSM" TX filter (L) and signal after passing through complementary "matched" TX + RX filters (R). The eye is as wide as possible at the detection point, and regardless of neighbours, all bits have the same amplitude at this instant. (i.e. no I.S.I.)

left-right symmetric. This filter is not driven directly by the rectangular bit stream, but by short + or - pulses of polarity corresponding to the data bits' polarity.

When cascaded with an identical analogue filter in the receiver, the overall response is as per figure 11 (right). The overall isolated bit shape is almost perfect, spanning some 5 bits with remarkable symmetry. The 'scope trace shows an excellent eye. The sample point convergence has about +/- 8% scatter due to the slight non-zero values of the isolated bit shape at $T = -2$ and $T = +2$.

THE FUTURE?

The 9600 baud RSM-PSK implementation is near optimal in performance, probably within 1 db of the theoretical limit. [approx. $20 \text{ Log} (1 - 8/100)$]

It is flying on RUDAK-2/AO-21 so AMSAT experimenters (i.e. YOU) can evaluate the effectiveness of high speed PSK with or without coding [3] from the associated on-board Harris RTX2000 RISC processor. This will help determine what is practical and realistic for amateur spacecraft and other digital links of the future.

The G3RUH 9600 baud FM system is not only in daily use by many satellite stations on UO-14, but also hundreds of terrestrial packet radio links world wide. It represents a design that pushes the fastest possible binary data through a basic FM radio. 9600 bps throughput is astonishing to behold, especially on a full duplex satellite link.

Who will predict that one day even our voice transponder links will be entirely digital. Amateur ISDN? Now there's a challenge!

REFERENCES

1. Miller J R, "9600 baud Packet Radi Modem Design", Proc. 7th ARRL Computer Networking Conference, 1988 Oct, pps 135-140
2. Meinzer K & Haas W, "RUDAK 2 - The Radio Links", Amsat-DL Journal No. 1/17, March 1990 pps 9-12. (German). English translation in Oscar News No. 83, 1990 June, pps 16-21
- 3 Miller J R, "Shannon, Coding and the Radio Amateur", Oscar News No. 81, 1990 Feb, pps 11-15

CONTACTS

AMSAT-UK, London E12 5EQ, England. (Oscar News).

AMSAT-DL, Holderstrauch 10, W-3550 Marburg 1, Germany. (A-DL Journal)

PacComm Inc, 3652 W Cypress St, Tampa, FL 33607-4916, USA. (NB-96 Modem)

Kantronics Co Inc, 1202 E 23rd St, Lawrence, KS 66046, USA. (DE9600 modem).

Tasco Tereleader, 3 8 Minami-Youchi, Higasibata, Anjo-City, Aichi 444-12, Japan. (TMB-965 modem).