Adaptive Coding and Modulation for Phase 4 Ground

Michelle Thompson W5NYV
Phase 4 Ground Lead

Phase 4 Ground provides digital radio solutions for any payload that complies with the Phase 4 Ground Air Interface document. These projects currently include but are not limited to Phase 4B Payload, Cube Quest Challenge (CQC), Phase 3E, and terrestrial Groundsats.

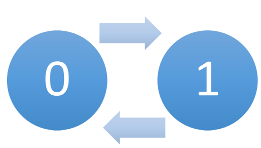
An Introduction to Coding and Modulation

In analog wireless communications, continuously varying signals are sent from transmitter to receiver. Voice, for example, is directly encoded in an analog transmission by a proportional relationship between baseband and carrier. The changes in audio that make speech intelligible to the ear are proportional to changes in either the frequency (FM), amplitude (AM), or phase (PM) of a transmitted carrier signal.

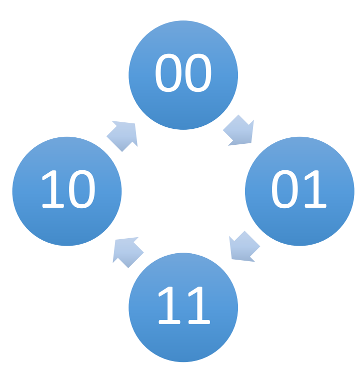
In digital wireless communications, data such as voice is represented by the digital symbols 1 and 0. Coding is the process of removing unnecessary redundancy in a signal and adding the right type of redundancy. Removing unnecessary redundancy is compression. Adding useful redundancy is channel coding. The type of channel coding we're most interested in is forward error correction coding. This is a way of coding the data where we can recover corrupted parts of the signal.

When we talk about **code rate**, we are talking about the ratio of how many bits go in to the forward error correction coder, or **encoder**, over how many go out. A rate 2/3 code takes in two bits and produces three. The extra bit is produced with mathematics especially designed to make the signal more durable as it travels from transmitter to receiver. The more bits you add, the smaller the ratio. Rates up to 1/9 are common. For a rate 1/9 code, for every bit that goes into the encoder, nine come out. As you'd expect, the more coding, the more durable the transmitted bits are against noise and interference. However, there's a cost. If you compare two signals that are transmitted at the same rate, the one with more extra bits to protect it needs more time to get through. The data rate is lower. It takes longer to transmit the same amount of data.

After the data is channel coded, the resulting bits are transmitted. The simplest type of digital waveform has two distinct states. One state corresponds to a 1, and the other state corresponds to a 0. Each of these ready-to-transmit-values is called a **symbol**. When we send one bit at a time, we have two symbols to choose from. An example of this type of modulation is Binary Phase Shift Keying (BPSK). The **modulation order** is the number of symbols we have to choose from. For BPSK it's two.



This simple BPSK modulation scheme can be dramatically improved. Sending one bit at a time is a great start, but we can do a lot better. If we use four distinct states in our transmitted



waveform, then we can send binary data two bits at a time. We now have four symbols instead of two. An example of this type of modulation is Quadrature Phase Shift Keying (QPSK). The modulation order has doubled to four.

How about 8? 16? 32? Yes, to all, and more, all the way up to 256, 512, and even 1024! Sending 1024 bits in a single sample sounds amazing. So, why don't we just send 1024 bits in a single sample all the time?

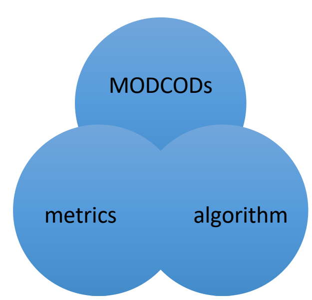
Engineering is all about trade-offs, and there's another one right here in front of us. The higher the modulation order the more power required. This means that the signal carrier power for transmitting two bits at a time must be twice that of transmitting one bit at a time, assuming that we are transmitting at the same **symbol rate**. We pay for the doubling in information capacity by having to provide double the power. As long as you have enough power, you can use more powerful modulations. If you have too much noise or not enough power, then you have to drop down to a lower modulation order.

Coding and Modulation Techniques in DVB

Traditional communications design assigns a fixed **mod**ulation and forward error correction **cod**ing (MODCOD) to a link. The MODCOD is selected to provide reliable communications under worst case conditions. For example, a microwave link that points down off a mountain is often designed to be good enough to work through rain fade and summer foliage. During clear conditions in the fall with no leaves, plenty of excess link margin is available, but a fixed system designed to work through summer thunderstorms cannot take advantage of this margin. In the Digital Video Broadcasting (DVB) world, this technique is called Constant Coding and Modulation (CCM). Phase 4 Ground uses many DVB protocols and techniques due to their high quality and widespread use in industry. Adapting these protocols to amateur radio is part of our mission.

Since it makes sense to adjust our link to better match observed conditions, one can design a system that uses a variety of MODCODs. An operator can then observe the link and then adjust the MODCOD to take advantage of better conditions. This technique is called Variable Coding and Modulation (VCM). VCM requires intervention of some sort to accomplish. In general, there is no feedback path from the receiver to the transmitter and a human is involved. But what if there was a feedback path from the receiver to the transmitter?

Adaptive Coding and Modulation (ACM) is a technique where the modulation and forward error correction are automatically changed in response to link conditions. As the link improves, higher order modulations and less coding allows increased throughput. Throughput can increase to take better advantage of available link margin.
Challenging link conditions are responded to by lower order modulation and more coding. The throughput will decrease, but the link is maintained. The adaptation is enabled by establishing the set of MODCODs to be used, listing the metrics that control the decision to change MODCODs, and defining the algorithm that produces the decision.



These three ingredients make up ACM.

With a CCM systems, severe fades can cause total loss of the link and zero throughput. VCM can address some of the challenges of severe fades, but ACM automatically turns fade margin directly into capacity. Maximizing throughput is highest with ACM.

Adaptive Coding and Modulation in Phase 4 Ground

The first challenge to an amateur-radio-centric version of ACM is that all existing implementations of ACM are proprietary. ACM is used in landline modems, 802.11, terrestrial microwave communications, and satellite links. When you are making money with subscribers, leaving margin on the table is not ideal. More efficient links mean more capacity, and more capacity means more subscribers, and more subscribers means more profit.

Most commercial ACM links generally only connect amongst themselves. There is no reason to create and maintain an open standard. Therefore, outside of the limited advice given in the implementation guidelines for DVB and a few white papers from a few companies, there is no open standard for ACM that we can simply adopt. For Phase 4 Ground we have to develop our own implementation of ACM, document it fully, and adjust it as we learn more in the field.

This is a great opportunity for amateur radio. Documenting the engineering trade-offs made in an advanced digital wireless system provides enormous educational opportunity for a wide variety of people, from interested amateurs to engineering students to satellite startups to people interested in machine learning and cognitive radio. Providing a working open-source implementation of ACM that other amateur projects can consider for adoption is a valuable engineering service.

The particular radio problem that has to be solved for space payloads is relatively straightforward. The geostationary and lunar and beyond radio environments are well-characterized. The available modulation schemes and coding rates are drawn from an established set described in the DVB standards (freely available from https://www.dvb.org). Advice from commercial and academic sources exist.

The particular radio problem that has to be solved for terrestrial Groundsats is also relatively straightforward. Groundsats are terrestrial versions of space-based payloads. They provide all the functions of an orbiting platform, but are on the ground. The control loop for terrestrial ACM has to be able to respond faster than for space. This is still well-characterized and advice exists from commercial and academic sources.

DVB allows an extreme resolution of MODCODs. Each and every frame can have a different MODCOD. This enables a link to respond very rapidly. For receiving transmissions from space, rapidly changing links are not the norm. The primary challenge is weather and rain fade or dishes not quite pointed right. For terrestrial links, changes in link quality can be more rapid, especially if the station is mobile. Terrestrial links have multipath, obstacles, noise, signal interference, and can also have rain fade and pointing problems.

There is a simple equation for ACM. In DVB, and for ACM in particular, the symbol rate is fixed. This greatly simplifies the communications system design. After a symbol rate is determined, a set of MODCODs is selected. The bit rate expression is therefore

Bit rate = symbol rate * modulation order * code rate

There are a lot of MODCODs to choose from in DVB. For space projects, the DVB-S family is the standard to reference. For terrestrial, we look to DVB-T. S stands for Space, and T stands for Terrestrial (think "television").

Phase 4 Ground uses DVB-S2X and DVB-T2. The 2 in DVB-S2X and DVB-T2 stands for second generation. Second generation DVB-T2 and DVB-S2 is backwards compatible (with some effort) to the first generation DVB-S and DVB-T. Second generation standards provide substantial improvements over first generation.

DVB-S2X is an extension to DVB-S2 that provides additional MODCODS and some additional mechanisms. Of compelling interest to us is the additional MODCODs at the lower end of the spectrum that provide enhanced very low signal to noise (VL-SNR) operation. For CQC, VL-SNR operation will provide needed support. For Phase 4B Payload, VL-SNR allows for reasonably sized dishes and opens up the possibility of patch arrays.

A large advantage gained in choosing DVB standards is that the receiver is explicitly told, frame by frame, exactly what MODCOD has been chosen. The receiver does not have to do anything extra to use ACM. The complexity of ACM is in the transmitter.

The second challenge to an amateur-radio-centric version of ACM is that ACM assumes exactly one intended receiver. If the transmission is a QST or CQ, or intended for a roundtable talk group, or is merely open to monitoring by silent listeners, modifications to the standard ACM scheme will be needed.

Maximizing The Bit Rate

There is a very important distinction between analog and digital systems and how to interpret the guidance for best operating practices as set out in part 97.

In analog communications in amateur radio, there is a principle of conservation of power. The least amount of power should be used to ensure reliable communications in normal operations.

Part 97 : Sec. 97.313 Transmitter power standards

(a) An amateur station must use the minimum transmitter power necessary to carry out the desired communications.

Obviously, emergencies may require a different practice. In digital communications, the spirit of this guidance is best achieved with maximizing the bit rate, or throughput. Maximum bit rate ensures that the communications are achieved in the most efficient manner by providing maximum capacity. If this means transmitting at a higher power than is necessary to simply maintain a communications link, then so be it. It's better to transmit for 450 milliseconds and then almost immediately allow someone else to then use the channel than to transmit for 450 seconds on minimum power using maximum coding and the lowest modulation scheme before relinquishing that particular channel. We equate bit rate with power and assert that this complies with the spirit of part 97.

We want to maximize throughput. This means maximizing the bit rate. In order to get to maximum bit rate, the professional advice is to start out with a stable link and work upwards. Here's an excerpt from Work Microwave's website.

> Start conservatively, approach the optimum: When setting up a link it is wise to start with very conservative settings to have a stable link running in the first place. Even if the "first shot" has not the desired bandwidth efficiency, an incremental approach will be the best way to optimize the link once it is up and stable. Due to numerous parameters and conditions affecting the Es/N0, the best settings will be reached by trial and error and can hardly be predicted beforehand.
>
> "ACM Dos and Don'ts." Work Microwave, 13 Mar. 2016, https://work-microwave.com/acm-dos-donts/

The Es/N0 value is a big clue. It's a critical metric for ACM. It stands for energy per symbol divided by the noise power spectral density. We already know what symbols are. A symbol is the distinct states of the modulator. The simplest one transmits 0 and 1. Two symbols are able to be transmitted so the modulation order is 2. Next most complex is 00, 01, 11, and 10. Four symbols are able to be transmitted so the modulation order is four. Next most complex is 000, 001, 010, 011, 100, 101, 110, 111. Eight symbols are available to be transmitted so the modulation order is eight. An example of this type of modulation scheme is 8PSK.

Energy Per Bit
Es/N0 is commonly used in the analysis of digital modulation schemes, but we're going to dig deeper and look at at a quantity called Eb/N0. This is the energy per bit divided by the noise power spectral density. Eb/N0 is the normalized signal to noise ratio of our link and this value is what drives the adaptation decisions in ACM. Think of Eb/N0 as the signal-to-noise (SNR) per bit. The energy per symbol and the energy per bit are related by the following expression.

$Es/N0 = Eb/N0 * \log_2(\text{modulation order})$

So for the modulations that we listed above, we have the following relationships.

$Es/N0 = Eb/N0 * \log_2(2)$ *two symbols to choose from*
$Es/N0 = Eb/N0 * \log_2(4)$ *four symbols to choose from*
$Es/N0 = Eb/N0 * \log_2(8)$ *eight symbols to choose from*

This gives us

For modulation order 2: Es/N0 = Eb/N0

The energy required to transmit a symbol of 0 or 1 is the same as required to transmit 0 or 1 bits. Makes sense!

For modulation order 4: Es/N0 = Eb/N0 * 2

The energy required to transmit a symbol of 00, 01, 10, or 11 is twice as much as required to transmit a 0 or 1. Still makes sense.

For modulation order 8: Es/N0 = Eb/N0 * 3

The energy required to transmit a symbol of 000, 001, 010, 011, 100, 101, 110, 111 is three times as much as required to transmit a 0 or 1. We are seeing the pattern.

In ACM, we have to be able to decide when we can afford to move on up to the higher order modulation schemes, which allows us to transmit more bits at once. If all the power we have available to us amounts to about as much power as required to transmit one bit, then we are stuck transmitting one bit at a time in BPSK. If our metrics tell us that we have more than three times the power required for a single bit available to us, then we can transmit a symbol that stands for three bits at once. We can go with 8PSK.

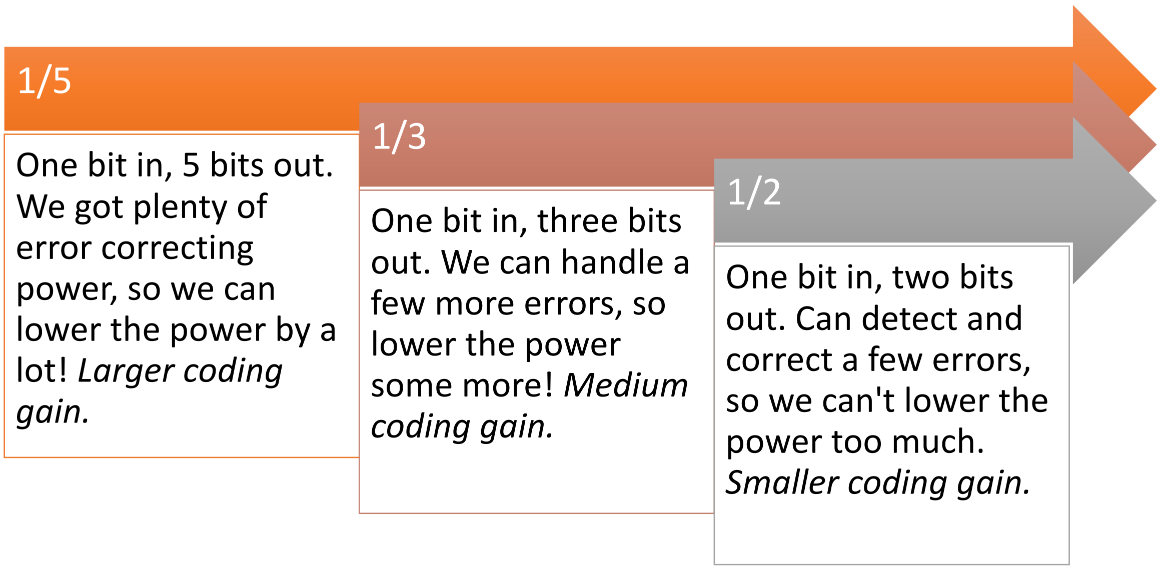
Within the modulation schemes are sets of coding rates. We've seen how spending power can increase the bit rate. How does coding fit in?

Coding Gain
There are two major types of coding. **Source coding** removes unnecessary redundancy so that source data can be more efficiently stored and handled. For example, digital music and video is source coded for compression. Otherwise the directly sampled files would be enormous.

Channel coding puts back in the right type of redundancy to make the transmitted signal resilient. Forward error correction puts in additional bits that allow for both the detection and correction of errors. Better than magic!

In DVB-S2X, the forward error correcting code is called LDPC-BCH. It's an advanced **concatenated block code**. Block code means that groups of bits are gathered up and then mathematically modified with extra bits. There are other types of codes that operate on continuous streams of bits. Those types of codes operate bit-by-bit as long as there are bits in the pipeline. Each block stands alone and is decoded separately. Concatenated means that two different codes are used. The reason these two different codes are used together in DVB-S2X is because using them together cancels out weaknesses. Taken together they make a very high-performance code.

Coding gain is the measure of the difference between the Eb/N0 levels of an uncoded system when compared to a coded system, when both systems are required to provide the same bit error rate. We have the same signal energy available in either case. Coded signals allow us to correct errors, which allows us to transmit at less power.



What can with do with this extra gain? In ACM we can put it right to work in maintaining target bit error rate performance. If we know what Eb/N0 we need, and we know which codes consume that much Eb/N0 to maintain a particular performance level, then we are able to select the code that maximizes bit rate while minimizing bit error rate.

We do this by measuring Eb/N0 at the receiver. This tells us how strong the signal is. Eb/N0 is reported to the ACM controller, and the right modulation and coding is selected for that receiver. In commercial satellite, the ACM controller is centralized and is usually on the ground. For Phase 4B Payload and for Groundsats, it's planned that the controller will be onboard the satellite.

Changing the modulation is the coarse-grain control knob in ACM. Changing the code rate is the fine-grained control knob in ACM.

Putting Metrics, MODCODs, and Algorithms Together

For ACM to work, the modulator must send which MODCOD is being used at the start of each frame. The receiver must be able to handle an arbitrary change in MODCOD without any advance knowledge. The receiver must be able to work fast enough to process the packet or frame without corrupting or dropping it. This puts a lot of pressure on the receiver. This pressure can be alleviated in several ways. One example is using standardized mechanisms in DVB such as time slicing. See Wally Ritchie's paper "Using DVB-S2X and Annex M to implement low-cost Phase 4B Earth Station Terminals" for ideas on time slicing.

Another requirement is that the receiver needs to be able to measure or calculate an estimate of the link quality (Eb/N0) and then communicate this estimate to the payload. The payload must be able to process this reported metric and then adapt the data rate and change the MODCOD sent to the receiver. This maximizes the data rate, complies with the spirit of part 97, and is sparkling with efficiency and style.

Reacting to changes in channel quality makes sense. But can there be additional improvement? Yes, there can! There's a large body of research that shows how throughput and bit error rate performance changes when using linear prediction to estimate the future state of the channel based on past measurements.

There are practical limits to how quickly an ACM system can respond. In general, about 1dB per second is achievable. If something happens and the demodulator comes unlocked, then it's a good idea to go back to the lowest MODCOD. This way, you're starting over with the highest probability of re-connecting and then working your way back up to maximum bit rate.

Assuming that the receiver has acquired the satellite and done all necessary chores to receive the downlink, and assuming the receiver has the necessary authentication, and assuming the receiver can successfully determine which channels are free for transmission to the payload, the receiver now needs to determine what MODCODs it is capable of receiving.

The dish might not be pointed correctly. The receiver might be a bit noisy. The local oscillator might not be rubidium quality. There might be some atmospheric conditions that attenuate the signal. Someone could have dented the dish. A connector could be loose. Some of these factors change very slowly over time, and some of them change more quickly. All of these factors affect receive capability and all of them can be automatically accommodated with ACM.

The standard model of ACM has the receiver monitor and report its Eb/N0 to the controller. In our case, the controller can be in the payload. When Eb/N0 falls below a setpoint, the receiving station is sent a lower MODCOD. The setpoints are configured to provide a minimum level of performance. When going to a lower MODCOD, throughput is reduced but the link is maintained. Eb/N0 reports can be part of the frame structure.

Digital communications performance can be defined by maximum allowable bit error rate. DVB is designed to provide very low error rates. The standard of performance for DVB is called quasi-error-free. DVB allows one uncorrected error per hour of video broadcast viewing. This is a very high standard that works out to a bit error rate of about $1*10^{-10}$ to $1*10^{-11}$.

When you establish the values for Eb/N0 that you're going to allow, they have to be made based on what bit error rate you can tolerate. Quasi-error-free bit error rate in DVB is many orders of magnitude lower than, say, the maximum bit error rate for GSM ($1*10^{-3}$) and lower than the data-centric maximum bit error rate for 3G data ($1*10^{-6}$). Voice is more forgiving than data which is more forgiving than digital video broadcasting. Selecting a baseline bit error rate of $1*10^{-6}$ is not out of line.

Once you have a bit error rate that you want to keep below, then you calculate a table of Eb/N0 values that would cause you to move to a better MODCOD. "Better" could mean higher or lower depending on whether you were doing great or having trouble with the link.

Anyone that's ever worked with set points knows that there's a potential for oscillating when the measured value is very close to the set point. A common approach with ACM is to put in 0.3dB or more of hysteresis. Going up requires a bit more SNR than coming down. This doesn't just prevent oscillating between two MODCODs but can also help maintain demodulation lock. You don't want your radio to work any harder than it has to.

We want the operator to see as much information about the metrics and the link as they desire. Our goal is to provide quality presentations of signal-to-noise ratios, states of lock, channel occupancy, system status, Usersynchronous log visualizations, symbol rate, modulation constellation, data rate, bit error rate, and more. Metrics such as these and more are presented by an application that can be run or not, depending on the preferences of the operator. Some systems provide a bit error rate tester as an application. This can help debug situations of synchronization loss, unexpected bursts of bit errors, or other problems. If the operator doesn't want to see any of this, then they don't have to. It should "just work" without intervention, and provide clear error or failure messages if anything goes wrong.

When a higher MODCOD is selected, the available data rate is increased. This usually isn't a problem. When a lower MODCOD is selected, the available data rate is decreased. This can be a problem. Congestion control must be considered and implemented to avoid losing packets or frames. Buffers and FIFOs to the rescue!

Is maximizing the bit rate enough? What about latency? While ACM considered in the abstract doesn't minimize or maximize latency, the use of DVB-S2X can offer some relief over DVB-S2. Latency will be one of the biggest challenges to operator experience on the Phase 4B Payload. It is impossible to go faster than the speed of light, and the round-trip delay of at least 240mS is substantial. There are things that we can do to mitigate latency such as reducing buffer size and using shorter frame lengths. Providing voice memo as an alternative to real-time voice is another.

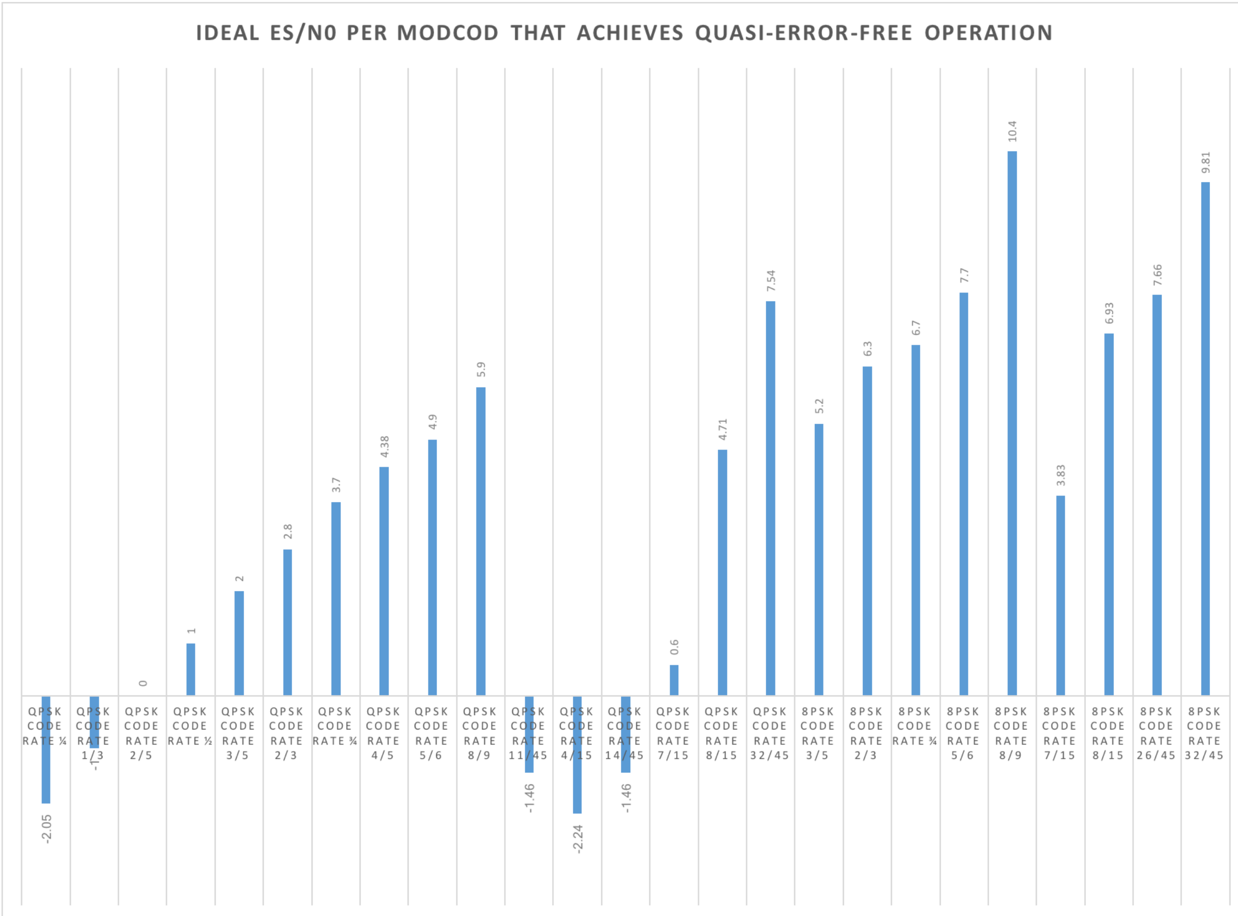
Proposed Adaptive Coding and Modulation Scheme

Here's the current proposal for MODCODs, metrics, and algorithm for ACM for Phase 4 Ground. This is an open design that is going into prototyping and testing. The expectation is that this proposal will be reviewed, refined, and retuned to maximize bitrate and avoid commonly encountered challenges. Some challenges are anticipated and have been mentioned above. Others we will certainly discover along the way.

There are choices of frame size in DVB-S2 and DVB-S2X. The CCSDS (Consultative Committee for Space Data Systems) RF Modulation and Channel Coding Workshop, among other individuals and groups, recommends the short frame size for near space-earth transmissions. A selection of the short frame size MODCODs that we believe will work best for Phase 4B Payload are presented in the table below. Short frame size is 16200 bits. Frame size and the presence or absence of pilot signals is communicated in the TYPE field of the physical layer header. Each MODCOD has an identification code. The decimal value of that code, which goes into the PLS field of the physical layer header, is the first column. Ideal Es/N0 is ideal energy per symbol divided by the noise power spectral density in dB in order to achieve a frame error rate of 10^{-5}. This is quasi-error-free operation with no impairments. In other words, very ideal!

PLS Code	MODCOD Name	Rate	Ideal Es/N0
1	QPSK code rate ¼	1/4	-2.05
2	QPSK code rate 1/3	1/3	-1.00
3	QPSK code rate 2/5	2/5	0
4	QPSK code rate ½	½	1
5	QPSK code rate 3/5	3/5	2
6	QPSK code rate 2/3	2/3	2.8
7	QPSK code rate ¾	¾	3.7
8	QPSK code rate 4/5	4/5	4.38
9	QPSK code rate 5/6	5/6	4.9
10	QPSK code rate 8/9	8/9	5.9
216	QPSK code rate 11/45	11/45	-1.46
218	QPSK code rate 4/15	4/15	-2.24
220	QPSK code rate 14/45	14/45	-1.46
222	QPSK code rate 7/15	7/15	0.60
224	QPSK code rate 8/15	8/15	4.71
226	QPSK code rate 32/45	32/45	7.54
12	8PSK code rate 3/5	3/5	5.2
13	8PSK code rate 2/3	2/3	6.3
14	8PSK code rate ¾	¾	6.7
15	8PSK code rate 5/6	5/6	7.7
16	8PSK code rate 8/9	8/9	10.4
228	8PSK code rate 7/15	7/15	3.83
230	8PSK code rate 8/15	8/15	6.93
232	8PSK code rate 26/45	26/45	7.66
234	8PSK code rate 32/45	32/45	9.81

When we look at a chart of these MODCODs, we can see the effect of modulation and coding. We get about 12dB of range just using QPSK and 8PSK. We haven’t yet listed the VL-SNR codes that can bring the Es/N0 down to -10dB. They require some additional care and work to implement.



We need to select enough different MODCODs to give the performance we want, but not so many that we have a situation where the algorithm is flailing about making unnecessary changes. The starter list of MODCODs is the following. This gives a MODCOD at about every 2-3dB.

QPSK 4/15 (identification number 218) -2.24
QPSK 2/5 (identification number 3) 0
QPSK 2/3 (identification number 6) 2.8
QPSK 4/5 (identification number 8) 4.38
8PSK 5/6 (identification number 15) 7.7
8PSK 8/9 (identification number 16) 10.4

All measurements have error. There are multiple sources of error and noise. The set of target Es/N0 (or Eb/N0) numbers need to be far enough apart to where link performance instead of noise is the main trigger of an ACM decision.

If three MODCODs turn out to be the best match, then it means we use three MODCODs. If we can use more, then we use more.

Once the MODCODs are selected, hysteresis is applied, and the metrics are monitored, then the choice of which MODCOD to apply to which frame can be usefully made.

While the underlying mechanism is straightforward, there are many problems to solve. Flow control and what type of quality of service needs to be decided. The DVB implementation guidelines give a great start for ACM and describe ways to set up Generic Stream Encapsulation (GSE) to help implement ACM.