Course 3: Digital encoding of the baseband signals

"Soon it's all going to be digital anyway. It's all going to be saved on a little coin somewhere."

Richard Donner
(Superman's creator)

Agenda

- Transmission types
- Digital signals
- Key terms
- Objectives of the digital encoding
- Digital encoding techniques: classification, features etc

Transmission types

- Digital data, digital signal (digital transmissions)
- Analog data, digital signal (analog data storage on digital media)
- Digital data, analog signal (e.g. "digital modulations")
- Analog data, analog signal (PSTN, broadcasted TV & radio)

A **digital signal** is a signal whose amplitude levels can take only values from a finite set of possible values. Also, if the signal comes from an analog signal, it will be discretized on the time axis too (by sampling) and afterwards quantified, the later operation leading to a discrete number of amplitude values. Those discrete amplitude values can be next converted into a sequence of numbers (represented as bits).

Unlike the digital signal, the analog signal can take any value within the range defined by its minimum and maximum value respectively. Furthermore, such a signal is defined in any time moment from the time axis and is commonly referred to as continuous time signal.

The data to be sent over a channel can be digital or analog, depending on the data source. The signal that figures that data can be analog or digital too; consequently four cases can be identified.

1. Digital data, digital signal

This happens when we have a digital signal source (e.g. a computer), that sends digital signals which represent these data. The wired computer networks make use of such a transmission, when the bits are encoded by rectangular electrical signals transmitted through the cable.

2. Analog data, digital signal

This is a more particular case, when analog signals are transformed into digital data and transmitted and/or stored. One common way to transform analog data in digital signals is the so called PCM (=Pulse Coded Modulation), when the data is sampled and the voltage levels so obtained are quantized and next encoded through an appropriate number of bits (according to a certain resolution level on the amplitude axis).

Example: the human voice (analog signal) can be captured by a microphone connected to the PC and next transmitted (e.g.: using a voice call service, such as Yahoo Messenger) or stored.

3. Digital data, analog signal

This represents a widely used method nowadays, figured by all the digital modulation techniques, such as PSK, ASK, FSK, QAM. Digital data is sent (mostly on the air interface) using modulation. Thus, the signals that are effectively transmitted are sine waveforms (thus analog signals), whose parameters (amplitude, frequency, phase) are shaped (modulated) by the sequence of bits/symbols to be transmitted. All radio transmissions make use of modulation, and all the last decades wireless technologies (GSM, UMTS, WiMAX, LTE, DAB, DVB, WiFi) make use of digital technology, i.e., the signal to be transmitted are digital.

4. Analog data, analog signal

Older, but still widely used technologies that transmit voice or video, transmit them in their original form (as analog signals), either in the baseband, or by modulation. Thus, the analog telephony (PSTN) transmit the voice as an electrical analog voltage in the 300Hz-3400 Hz band. Radio transmissions, by the other hand (e.g.: broadcasted TV and radio) use analog modulations (AM or FM) when transmitting the analog signal. Although the digital systems known a spectacular expansion in the last decades, the analog transmission still occupies an important place in our lives.

Digital signals

- A definition for discrete signal: a signal which maintains a constant level then changes to another constant level
- Represented by discrete, discontinuous voltage pulses
- Each pulse is a signal element
- Binary data (bits of 1 and 0) encoded into signal elements
- Each bit is mapped to one or more signal elements

Key terms

- Unipolar: all signal elements have same polarity (+ or -)
- Polar (bipolar): one logic state represented by positive voltage the other by negative voltage
- "Mark" stands for 1 and "Space" stands for 0 (from telegraphy)
- Differential encoding: the signal element(s) which encode a bit are a function of the previous bit representation
- E.g.: the occurrence of a bit of 1 determines a transition from the previous signal state
- Absolute encoding: the signal element(s) encoding a bit are function of only the current bit to be encoded
- Baseband transmission: the signal occupies a spectrum which is located around the "0" frequency (and NOT centered to a high frequency carrier)

Data rate

$$R = \frac{1}{T_b} \quad [bps] \quad$$ T_b being the bit duration (time taken to emit one bit)

- The modulation rate is different of the data rate (can be lower or higher)
- Defined as: rate at which the signal level changes (the reversed of the duration for which the signal level stays constant)
- Measured in Baud = signal elements per second

Whereas the data rate simply shows how many bits can be sent on a specific time unit (one second), the modulation rate is rather related to the physical signal transmitted through the medium. The later is expressed as the rate at which the signal level changes. For example, during one bit representation, the signal can stay half of the bit period on a certain voltage level (e.g. +1V) and the other half on the opposite level (-1V). This is the case in the bi-phase codes (Manchester), and the result is that, in this specific example, the modulation rate is twice the data rate.

Digital encoding

- "What is digital encoding, anyway?"
- "Information" is an abstract concept, the "bit" concept too
- When transmitting, we must face the physical reality
- Signals are used to figure and to transmit the bits
- Digital encoding = a set of rules according to which a sequence of bits is mapped to a signal
- Depending on the transmission environment, the signal can be an electrical signal, a radio wave (modulation) or a sequence of light pulses
- Mostly, digital encoding refers to transmission through guided media (cables)

Digital encoding of the baseband signals = a certain strategy, by which a sequence of bits is associated with a certain electrical signal, with some expected properties.

Decoding

- The receiver "observes" the medium (captures a signal)
- Based on this signal, the receiver must decide on what were the transmitted bits
- The receiver must know: the bits timing (the beginning and the end of each bit) and the signal levels
- The first step is sampling, next a threshold comparison

The essential triplet

- Signal to Noise Ratio, Data Rate, Bandwidth
- By increasing the data rate, the BER increases too (E_b/N_0 degrades)
- A better SNR allows BER improvement
- More bandwidth will allow higher data rates

Objectives of digital encoding [1]

- Spectral efficiency
  - Lack of high frequencies reduces the required bandwidth
  - Concentrate power in the middle of the bandwidth
- No DC component: allows AC coupling via transformer, providing isolation
- Synchronization
  - The receiver must have exactly the same time reference as the transmitter (beginning of the bit interval, bit duration etc)
  - Sampling must be very precise, especially at high data rates
  - External clock versus sync mechanism based on the received signal

Spectral efficiency

The most common way to define the spectral efficiency is in terms of bits/s/Hz, i.e., this parameter measures how much information (data) can be sent through a channel having a specific bandwidth. Since the bandwidth is always constrained, ingenious ways to push more data through a fixed bandwidth are a big challenge and an important research field.

In their digital form, signals are represented by rectangular signal elements, that occupy a large bandwidth (Reminder: the Fourier transform of a rectangular pulse spans the whole frequency axis). In practice, it is not required (nor possible) to use an infinite band for transmitting a waveform that closely approximates the rectangular pulse. Rather, an effective bandwidth is defined, above which the energy of the signal can be neglected with minor consequences on the reception quality. Different digital encoding strategies lead to various shapes of the signals spectra, and, consequently to different values for the spectral efficiency. Concrete examples will be given on the next slides.

No DC component

The DC component occurs for any signal whose average amplitude level is different of 0. E.g.: in the uni-polar encoding, the signal is composed of voltage levels of a single polarity (either + or -) and periods of transmission break (0 voltage level). In this case, the signal will introduce a DC component. DC component in a signal is not desirable because the DC component does not pass through some components of a communication system such as a transformer. This leads to distortion of the signal and may create error at the output. The DC component also results in unwanted energy loss on the line.

Synchronization

Synchronization means that the transmitter and the receiver share the same time reference that dictates the bit interval. More concretely, the receiver knows the right boundaries of the bit (the moment when the bit representation begins). Practically, the receiver uses a local clock that gives the sampling moments, and these moment must be faithfully respected in order to achieve good synchronization. Usually, the local clock is synchronized from the received signal with the help of a special hardware known as Phase Lock Loop (PLL). However, this can be better achieved only if the received signal is self-synchronizing having frequent transitions (preferably, a minimum of one transition per bit interval) in the signal. The synchronization issue is very critical in the high speed communications, where the bit duration is very short.

Objectives of digital encoding [2]

- Error detection
  - Usually the task of the upper layers, but some mechanisms can be incorporated in the digital encoding strategy too (phy layer)
- Signal interference and noise immunity
  - Some codes are better than others (in terms of BER versus SNR)
- Cost and complexity
  - Higher signaling rate lead to higher costs
  - Some codes require modulation rate higher than data rate
  - The relation between the modulation and the data rate is critical

Classification

- NRZ codes (NRZ, NRZI)
- Multilevel binary (AMI, pseudoternary)
- Biphase (Manchester, differential Manchester)
- Scrambling techniques: HDB-3, B8ZS

NRZ codes

- Rule: bits are represented with constant levels, for the whole bit duration
- No voltage for "0", constant positive voltage for "1" (unipolar)
- In practice, negative voltage is assigned to "1"
- More often, bipolar version used (+ and – pulses, zero voltage avoided)
- Used in slow speed communication interfaces (e.g. RS232) and storage media (e.g. magnetic tapes)
- NRZI is the differential version of NRZ ("1" encoded by transition from the voltage level which represents the previous bit)

NRZ are the simplest form of digital transmission, where the signal level remains constant for the whole duration of one bit. Various NRZ versions exist (absolute and differential / unipolar and bipolar). In the differential version, signal is encoded by transitions rather than by levels (e.g. the occurrence of a bit of "1" causes a transition from the previous voltage level).

NRZ pros and cons

- Pros
  - Easy to engineer
  - Make good use of <u>bandwidth</u>: can achieve 2 bps per Hz of spectral efficiency
- Cons
  - DC component (for the unipolar version)
  - Lack of synchronization capability: long runs of identical bits mean no transition in the signal

Advantages

-The simplest codes

-Most of the signal energy (the whole main lobe) is concentrated to a bandwidth that equals the data rate. This means that NRZ is efficient from the spectral point of view. Notice that at half of the bandwidth, the spectrum falls off considerably, which means that in practice we can achieve even better than 1 bit/sec/Hz, and the largely declared spectral efficiency value for NRZ is of 2bits/sec/Hz.

Disadvantages

-In its unipolar version a DC component occurs

-The most annoying drawback of NRZ codes is given by the fact that long runs of identical bits will determine constant signal levels for a long time, and the synchronization at reception may fail. This drawback is very important, and it discards NRZ from being used in the modern, high-speed systems. Yet, versions of NRZ can be combined with other strategies that alleviate the synchronization issue (see 4B/5B NRZI).

Multi-level binary codes

- Alternate Mark Inversion (AMI) and pseudo-ternary
- Three signal levels (0, +V and -V)
- AMI rule: "0" means no signal, "1" is represented by voltage pulses with alternating sign
- Pseudo-ternary is the opposite of AMI
- Used in first generation PCM equipments
- Improved versions of AMI are currently in use
- Pseudo-ternary is the inverse of AMI (1 encoded by no signal, 0 encoded by alternating pulses)

Unlike the NRZ codes, bipolar AMI codes use three voltage levels (+V, -V, 0) to encode data.

Pros and cons of multi-level binary codes

- Pros
  - No loss of sync in case a long string of ones (in AMI)
  - Zero DC component by the nature of these codes
  - Bandwidth efficiency (main amount of power concentrated in less bandwidth than in NRZ case)
  - Simple mechanism for error detection (two consecutive pulses of the same sign show an error)
- Cons:
  - synchronization issues for long runs of "0" (AMI) or of "1" (pseudo-ternary)
  - They introduce some kind of redundancy: three levels to encode two bits
  - 3dB SNR loss versus the BER performance of NRZ

Biphase codes

- Manchester and differential Manchester
- Principle: supplementary transitions introduced compared to NRZ, AMI
- Manchester rule: there is always a transition in the middle of the bit period, and sometimes at the beginning
  - E.g.: "1" is represented by a low-high transition at the middle of the bit interval, "0" is the reversed
  - Inherently, a transition will be performed at the beginning of the bit, if the current bit it's identical with the previous one
- Differential Manchester: always transition at the middle of the bit (clocking only, it doesn't meter the sense of the transition)
  - Information is carried by the transition at the beginning of the bit: "0" means transition, "1" means no transition
- Manchester is used in IEEE 802.3 Ethernet

Bi-phase codes address the main disadvantage of NRZ, i.e. the lack of transitions in case of long runs of identical bits. It is oftentimes sad that "Manchester is made of a data signal and a clock signal", because in a Manchester code a transition always occurs in the middle of the bit period (similar to the transition exhibited by a clock signal).

Biphase codes pros and cons

- Pros
  - Predictable transition at the middle of each bit, which can be used for synchronization ("self-clocking" codes)
  - Error detection possible due to the same feature
  - No DC component
- Cons
  - Bandwidth efficiency issues (half the bandwidth efficiency of NRZ)
  - Modulation rate can be two times the data rate

Clearly, Manchester encoded signal has a larger number of transitions than NRZ or AMI signals, thus line synchronization is easier. But, in the same time, the larger number of transitions leads to high frequency components of important energy, meaning that the spectral efficiency of the Manchester encoded signal is lower than in the above mentioned cases (notice the wider PSD curve from slide 26). Practically, Manchester coding is a good choice in those applications where the couple transmission throughput – physical available bandwidth guaranteed that the signal is not distorted by the transmission environment.

Example: in the 10Mbps Ethernet (let'it be on coaxial or twisted pairs cable), the bandwidth of the physical medium is large enough such as to allow Manchester coding. By the other hand, this bandwidth becomes insufficient in 100Mbps Fast Ethernet, where a different strategy is adopted.

Modulation versus data rate

$$D = \frac{R}{b}$$

D = modulation rate [Baud]

R = data rate [bps]

b =number of bits per signal element

- Good <u>efficiency</u> if R>=D

20
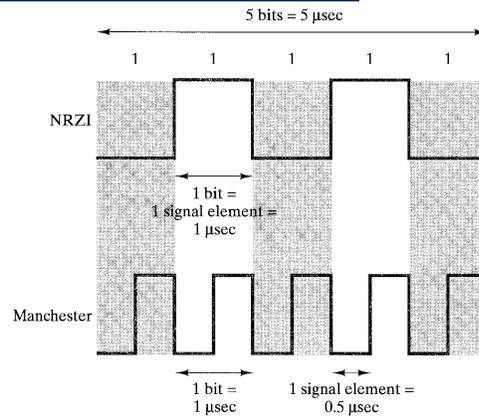
Scrambling techniques

- Scrambling: sequences that would produce constant voltage level are replaced by a signal with some transitions
- "Filling" sequence
    - Must produce enough transitions to sync
    - Must be recognized by receiver and replaced with original
    - Same length as original (no efficiency degradation)
- No DC component
- No long sequences of constant line signal
- The data rate is preserved (no overhead, redundancy)
- Error detection capability
- Most widely used are HDB-3 and B8ZS, both based on AMI

Note that the word "scrambling" may have a different meaning in data communications than its meaning in the present context. In this course, by "scrambling" we simply mean a way to introduce transitions into a signal that, by regular encoding strategies (e.g.: AMI), would produce not enough transitions. The codes presented on the next slides are simply improved versions of AMI, that, while respecting the general rule from AMI, make use of some "tricks" in order to introduce transitions into the signal transmitted through the line.

B8ZS

- Bipolar With 8 Zeros Substitution
- Can be considered as an improvement of the bipolar-AMI
- Rule: 8 consecutive zeros are NOT encoded with no signal for eight bit periods, a signal which has 4 transitions being used instead
- Whenever an all-zero octet occurs:
  - encode as 000+-0-+ if last non-zero voltage pulse was positive
  - encode as 000-+0+- if the last non zero pulse was a "-"
- AMI code rules are broken twice by "polarity violation": once inside of the eight-zeros group, once between the first non-zero pulse preceding the group and its correspondent within the group
- Unlikely to occur as a result of noise
- Receiver detects and interprets as octet of all zeros

Both B8ZS and HDB-3 may be seen as improved versions of AMI. They are generally used in long-distance communications.

HDB3

- High-density Bipolar, order 3
- Another improvement of the bipolar-AMI
- Rule: 4 consecutive zeros are NOT encoded with no signal for four bit periods, but with a signal which has at least one transition
- When a group of 4 zeros occurs:
  - encode as 000V if the number of bits of "1" after the last polarity violation is odd
  - B00V if the number of bits of "1" after the last polarity violation is even
  - V means the same polarity as the previous one, B respects AMI pattern
- This rule targets the elimination of the DC component (the polarity violation always changes its sign)

Scrambling pros and cons

- Pros:
  - Classical drawbacks of AMI are eliminated
  - No redundancy, therefore the same rate is maintained
  - Error detection capabilities
  - Good spectral efficiency
- Cons:
  - Extra-processing needed at transmitter and, especially at receiver side
- B8ZS employed in 1.544 Mbps-T1, HDB-3 in 2.048 Mbps-E1

The scrambling techniques have some basic error detection capabilities. E.g.: when in HDB3 two consecutive polarity violations have the same sign, this could be only caused by a transmission error.
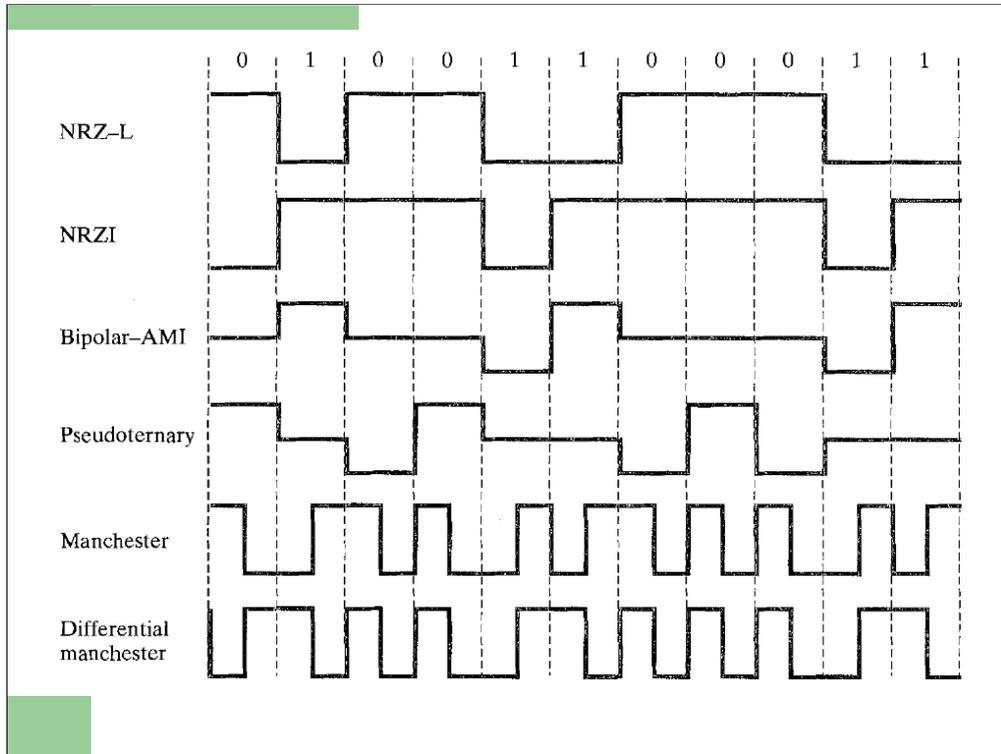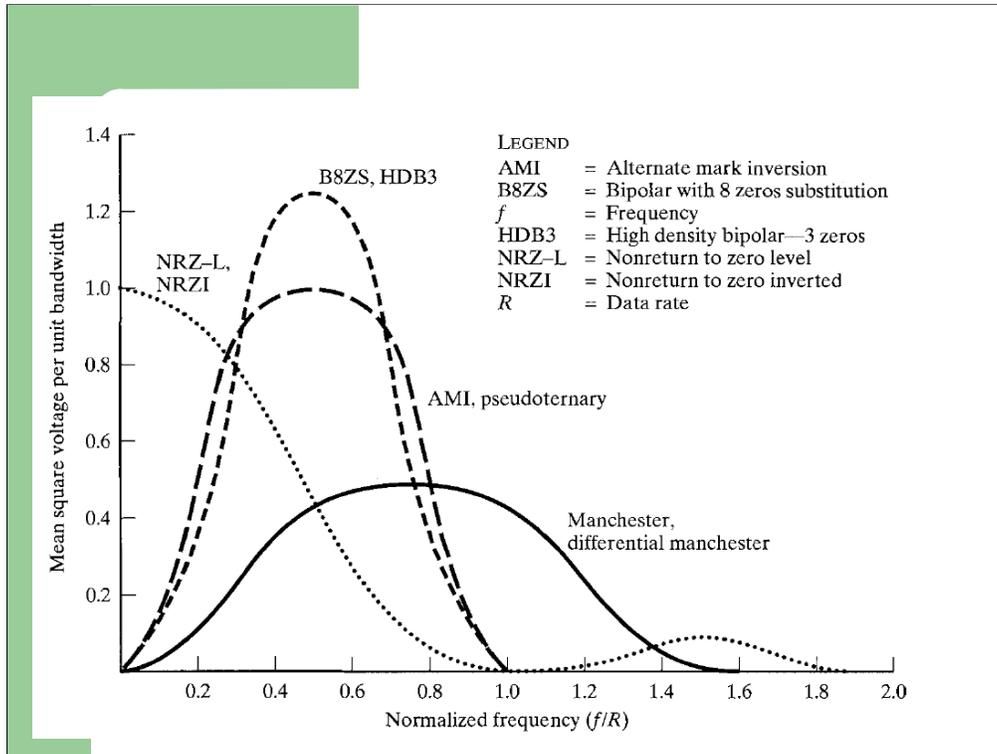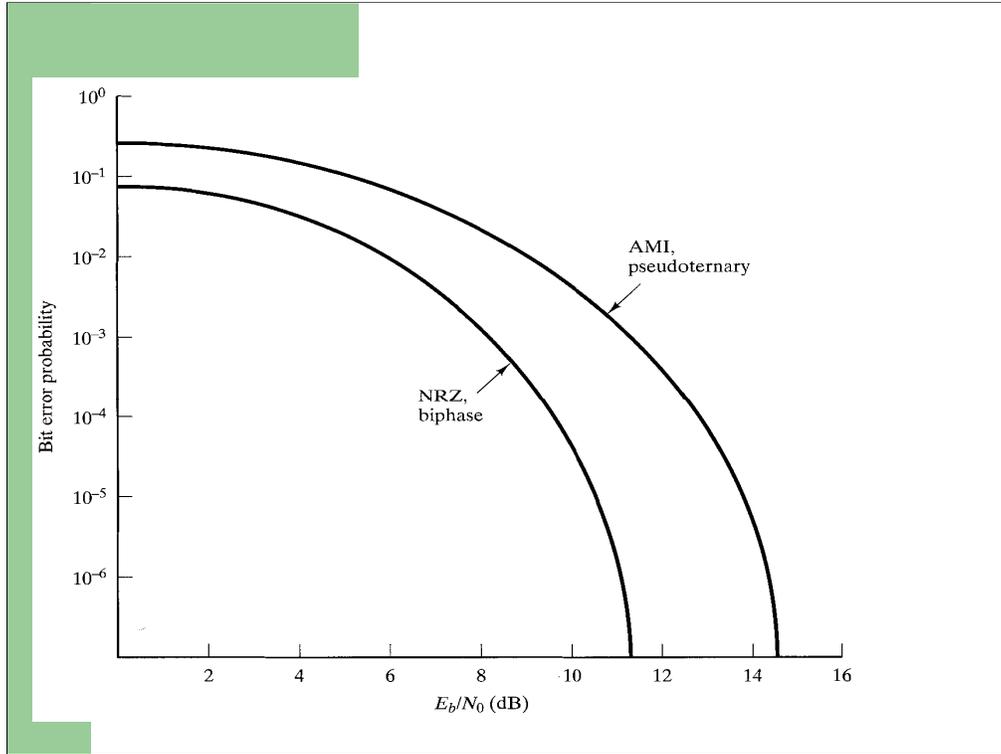
The figure above should be interpreted into the following manner: at a constrained BER (e.g.: 10^{-6}), NRZ shows approximately 3.5 dB of gain versus AMI, that is, it needs poorer channel conditions (3.5 dB less for E_b/N_0) to achieve the same performance.

TABLE 4.3 Normalized signal transition rate of various digital signal encoding rates.

	Minimum	101010. . .	Maximum
NRZ-L	0 (all 0's or 1's)	1.0	1.0
NRZI	0 (all 0's)	0.5	1.0 (all 1's)
Binary-AMI	0 (all 0's)	1.0	1.0
Pseudoternary	0 (all 1's)	1.0	1.0
Manchester	1.0 (1010 . . .)	1.0	2.0 (all 0's or 1's)
Differential Manchester	1.0 (all 1's)	1.5	2.0 (all 0's)

The table above shows the ratio between the modulation rate and the data rate. As stated previously, one should keep this ratio as low as possible, in order to achieve good spectral efficiency.

Bit value   1   1   0   0   0   0   0   0   0   0   1   1   0   0   0   0   0   1   0

Bipolar–AMI

B8ZS   0   0   0   V   B   0   V   B

HDB3   0   0   0   V   B   0   0   V   B   0   0   V

B = Valid bipolar signal
V = Bipolar violation

29