

RESEARCH

Open Access

Data processing techniques for a wireless data transmission application via mud

Qingjie Zhao*, Baojun Zhang and Wei Wang

Abstract

The data measured by well bottom sensors can be transmitted to the surface through the drilling mud during oil drilling operations. This article introduces a data processing scheme for a wireless data transmission application via mud. The detailed signal processing procedure is given, and several data processing techniques used are discussed, mainly including data encoding and signal integrating method, signal filtering, data storage and manage method, peak detection, signal recognition, and data decoding method. The article uses M pulses in N slots to encode the values of actual parameters. A two step filtering method and a dynamic data storing and managing method are proposed. A mix peak detection method is utilized to find the position of a pulse by combining threshold method and neighbor comparison method. These techniques have been successfully used in an oil well drilling operation.

Keywords: Signal processing, data encoding and decoding, data transmission

Introduction

When drilling oil wells, especially in directional drilling, it is very helpful to utilize a kind of measurement-while-drilling system to provide real-time monitoring to the direction of a bottom-hole assembly, the angle of the hole, the gamma radiation from formations, and some other physical parameters. However, it is difficult to transmit the data measured from down-hole environments with thousands of meters depth, high temperature, and high pressure. At present, transmitting the data through cables may not be a good method because this will disturb ordinary drilling operations and the cables may be eroded under the rigorous down-hole conditions.

Mud pulse telemetry [1] is one of feasible wireless methods used for oil drilling operations, mainly for the control and transmission of the data from a well bottom to the surface during drilling operations. Drilling mud is added to the wellbore to facilitate the drilling process by suspending cuttings, controlling pressure, stabilizing exposed rock, providing buoyancy, and cooling and lubricating. Transmitting the data from a well bottom to the surface is an another function of drilling mud, which can help

drilling operations but give less influence to the drilling process.

Although there are some reports [2] that introduce measurement-while-drilling tools, and enormous literatures on signal processing in other fields such as geophysics, medical imaging, vibration studies, etc., however, there are few literatures that introduce data or signal processing techniques for a measurement-while-drilling system in petroleum engineering. A measurement-while-drilling system based on a microcontroller is developed in [3]. The data come from different down-hole sensors such as three-axe accelerometers, magnetometers, gamma-ray detector, resistivity detector, and other sensors. Ledroz et al. [4] and Pecht et al. [5] use a fiber-optic-gyroscope-based inertial measurement unit in gyroscope aims. Wavelet transform in [6] is used to get rid of high-frequency noise from the contaminated data. In [7] and [8], a limited impulse response low-pass filter is used as a DC (direct current) estimator, and a band-pass filter is used to eliminate the large out-of-band noise components caused by the mud pumps, and at last a zero mean signal is acquired. In [9], we propose a two-step filtering method in which a dynamic part mean filtering algorithm is proposed to separate the direct current components and a windowed limited impulse response algorithm is used to filter out the high-frequency noise.

* Correspondence: qingjie.zhao@gmail.com
Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China

In this article, we introduce several data processing techniques used in a mud pulse telemetry system, mainly discuss signal encoding, filtering, data storing and managing, signal recognition and decoding methods. Although the scheme using M pulses in N slots to encode values is not a new idea, our contributions mostly lie in giving detailed signal flow, formulating the relation between the maximal number of codes and M/N , proposing a two step filtering method, a dynamic data storing and managing method, and a mix peak detection method.

The rest of this paper is organized as follows. 'Principle of mud pulse telemetry' section describes the principle of mud pulse telemetry. 'Down-hole data processing' section introduces the data encoding method and the combined signal components. 'Surface data processing' section discusses surface data processing techniques including filtering, sequence recognition and decoding. Finally, a brief conclusion is presented in 'Conclusion' section.

Principle of mud pulse telemetry

The system designed includes two parts: the down-hole part and the surface part. The down-hole part modulates the data from both down-hole sensors and the embedded computing module, creating mud pressure pulses to carry encoded down-hole data to the surface. At the surface, the mud pressure pulses are detected, transformed, processed, and decoded.

Down-hole sensors include three magnetometers and three accelerometers fixed tri-axially to measure the compass direction of the bottom-hole assembly and the angle of the hole which are then used to calculate the trajectory of the well along with depth. A gamma-ray detector measures naturally occurring gamma radiation from formations encountered to estimate stratigraphic formation. A resistivity detector is to help recognize rock, oil or water. These data together with those of down-hole temperature, generator's rotate speed and battery's voltage are gathered, converted and formatted for transmission, and stored in the embedded computing module. The embedded computing module encodes data into pulses and controls the pulser operations. The pulser generates electrical power and restricts the mud flow to create pressure pulses with a valve in the stream of mud to be controlled open or close. The pressure in the pipe is caused to rise or fall respectively, and pressure waves are generated. The modulated data are then transferred to the surface with drilling mud pulses.

The surface part, which receives and decodes the data from the down-hole, includes a mud pressure sensor, an interface box, computers, and displays for drilling operators or technicians. At the surface, the pressure sensor measures the pressure pulses in the drilling mud column. The pressure pulse signals are pre-processed and then passed to a computer. The signals received by the

computer are transformed into digital ones, and then filtered, processed and decoded, and some important information and parameters, such as the data of inclination, azimuth, tool-face orientation, temperature, pressure, generator's rotate speed, battery's voltage, gamma radiation, and resistivity, are acquired. These data can be preserved, displayed, printed, or transferred to a long distance computer via the Internet.

The signal flow of mud pulse telemetry is shown as Figure 1.

Down-hole data processing

Down-hole data measured by different sensors

The down-hole sensors used include: three magnetometers and three accelerometers fixed tri-axially, one temperature sensor, a counter used to get the rotate speed of the generator, and a sensor to measure batteries' voltage. The data from these sensors are compensated and processed to acquire the values of down-hole physical parameters, such as inclination angle, azimuth angle, gravity tool-face angle, magnetic tool-face angle, total gravity, total magnetic field, temperature, rotate speed of the generator and batteries' voltage. These parameters together with resistivity and Gamma ray data are encoded and transformed, according to a predetermined form, into a data string.

Encoding method

Data encoding

A datum is encoded by using a sequence with M pulses in N sequential time slots. A signal pattern is used to express a value. There are two possible signal states in a period of slot. We use $T = 0$ to represent that there is no pulse in this slot T , and $T = 1$ to indicate there is a pulse in this slot. Different signal patterns are used to represent different data values.

The encoding rules are described as following:

- There are at least two T s with state 0 between two T s with state 1;
- The last two T s must be state 0 in the sequence with N T s;
- A code is a pattern with M pulses in N T s.

Figure 2 is an example, where a code is a pattern with 3 pulses in 17 T s, and the last two T s is 0. "↑" means there is a pulse in that T , that is, the state is 1 in the T . For the first 9 patterns (pattern 0 to 8), the state of the first and the forth T is 1. For pattern 0, the state of the seventh T is 1, that is, in 17 T s, there is a pulse respectively in the first, forth and seventh T . After pattern 8, the third state 1 moves forward one position. For pattern 9 to 16, the second state 1 is in the fifth T , and the third state 1 starts from the eighth T and moves forward until the fifteenth

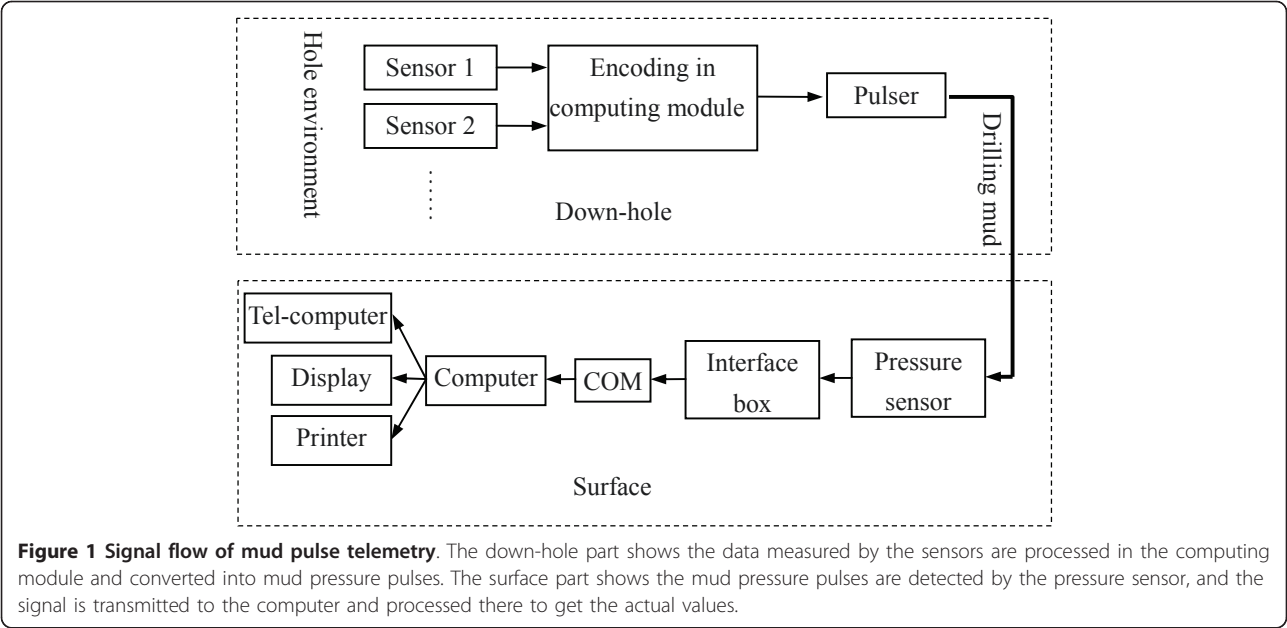


Figure 1 Signal flow of mud pulse telemetry. The down-hole part shows the data measured by the sensors are processed in the computing module and converted into mud pressure pulses. The surface part shows the mud pressure pulses are detected by the pressure sensor, and the signal is transmitted to the computer and processed there to get the actual values.

T , which represents the pattern 16. And then if we continue to move the states, we can get other patterns.

M and N are determined by the number of binary digital bits. When M and N are determined, the maximal number of codes available can be also determined. The relation between the maximal number of codes (N_{\max}) and M, N is described as follows (Figure 3):

(a) if $M = 1, N > 3$, then $N_{\max} = N - 2$;
(b) if $3 * M = N$, then $N_{\max} = 1$;

(c) for other cases, $N_{\max}(M, N) = N_{\max}(M - 1, N - 3) + N_{\max}(M, N - 1)$

According to the rules, if we know M, N and the pattern number, we can know the corresponding code. For example, if $M = 1, N = 4$, the maximum number of codes is 2, the two pattern numbers are 0 and 1, and the two codes are 1000 and 0100. Vice versa, if we know a code, we can know M, N and the pattern number.

Pattern	T series																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	↑			↑			↑									-	-
1	↑			↑				↑								-	-
...
8	↑			↑											↑	-	-
9	↑				↑			↑								-	-
10	↑				↑				↑							-	-
...	-	-
16	↑				↑										↑	-	-
17	↑					↑			↑							-	-
18	↑					↑				↑						-	-

Figure 2 A data encoding example with 3 pulses in 17 Ts. The last two Ts are state 0, and there are at least two Ts with state 0 between two Ts with state 1.

Number of bits	M	N	N_{\max}	Excess
1	1	4	2	$2-2^1 = 0$
2	1	6	4	$4-2^2 = 0$
3	2	9	10	$10-2^3 = 2$
4	2	11	21	$21-2^4 = 5$
5	2	13	36	$36-2^5 = 4$
6	3	15	84	$84-2^6 = 20$
7	3	17	165	$165-2^7 = 37$
8	3	19	286	$286-2^8 = 30$

Figure 3 The relation between Num and M, N. Using this pulse encoding method can get more pattern cases than using the binary encoding method.

When compared with the binary encoding method, this pulse encoding method has obvious advantages. For an 8-bit binary number, it can represent $2^8 = 256$ cases. When using this pulse encoding method, $M = 3$, $N = 19$, then 286 cases can be represented, which are 30 more than that the binary encoding method can provide, and each case only three pulses need to be activated.

Combined signal

The signal to the ground is a sequence of pulses and consists of synchronization pulses, mode pulses, status pulses and data pulses.

The section of synchronization pulses is used to keep the surface software to synchronize with the down-hole equipments. It is allocated at the beginning of each combined signal, and has its own special format with 3 pulses in 11 T_s . Each pulse lasts 1.5 T_s , and the interval is 3.5 T_s between two pulses, as shown in Figure 4.

The section of mode pulses is used to illustrate the components and characteristics of a data set. The signal format is 3 pulses in 14 T_s as shown in Figure 5. The first pulse lasts 2.5 T_s and the other two last 1.5 T_s . The first pulse is a flag pulse to mark the beginning of the mode section. One combination of the other two pulses determines one of the nine data modes used. One mode corresponds to one predetermined data components. For example, Mode 1 corresponds to the data set: 2 tool face angles, 3 gravity data, 3 magnetic data, generator's rotate speed, and temperature.

The section of status pulses is used to tell the working status of down-hole equipments such as a resistivity detector and a gamma-ray detector. In our software, the status section is used only in mode 9. The signal format is 1 pulse with width 1.5 T_s in 6 T_s . There are four cases (Figure 6) to represent whether the resistivity detector or the gamma-ray detector is valid or not.

The section of data pulses includes more than one parameters measured by down-hole sensors. The binary source codes of these parameters are first acquired, and then they are converted into pulse signal patterns. Data encoding rules used have been described in the above. We use 4 pulses in 25 T_s to encode the data of hole inclination angle and azimuth angle respectively, 5 pulses in 26 T_s to encode the data from three magnetometers and three accelerometers, respectively, 3 pulses in 17 T_s to encode the data of tool-face angle, total gravity, total magnetic field, magnetic inclination, temperature, rotate speed of the generator and batteries' voltage, respectively, and 3 pulses in 19 T_s to encode the data of resistivity and Gamma ray, respectively.

For example, the tool-face angle is in the range of 0° to 360° . We use 7 binary bits to denote the values and the range is 0×00 to $0 \times 7F$. Since $360/2^7 = 2.8125$, the binary code 0×01 corresponds to 2.8125° . When using 3 pulses in 17 T_s to encode the value 2.8125, the equivalent pulse pattern code is 10010001000000000, and the pattern number is 1.

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
↑				↑			↑			

Figure 4 Synchronization pulses. The format is 3 pulses with width 1.5 T_s in 11 T_s , and the interval is 3.5 T_s between two pulses.

Mode	T series													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	↑					↑			↑					
2	↑					↑				↑				
3	↑					↑					↑			
4	↑					↑						↑		
5	↑						↑			↑				
6	↑						↑				↑			
7	↑						↑					↑		
8	↑							↑			↑			
9	↑							↑				↑		

Figure 5 Mode pulses. The format is 3 pulses in 14 Ts, where the first flag pulse lasts 2.5 Ts and the other two last 1.5 Ts which determine the data mode or data components.

In a word, the combined signal is in the order of synchronization section, mode section, status section and data set, where the status section only used in mode 9.

Afterwards the combined signal is magnified and used to control the pulser, and the signal is converted to a series of drilling mud pressure pulses. The pressure sensor fixed in the riser pipe converts the pressure pulses to 4 to 20 mA electric current signal to overcome the problems of disturbance and voltage reduction for the long transmission distance between the sensor and the interface box. In the interface box, the signal is processed and converted to a voltage sequence, and is transmitted to the surface computer by a serial port.

Surface data processing

The data processing at the surface is shown as Figure 7. The surface computer receives, memorizes, and processes raw signals to get filtered data sequences. Then a dynamic storing and managing container is used to hold

and manage the filtered data sequences. Real-time decoding is used to get the values of various parameters. The software is capable to provide a graphical and numerical view of the raw, filtered and decoded data.

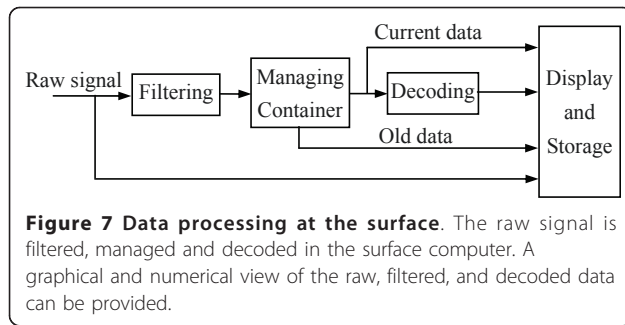
Signal filtering

While transmitted from down-hole to the surface, the combined signal is inevitably contaminated by various kinds of noise, which may have much bigger amplitudes or much higher frequencies than that of the encoded signal, so the received signal should be processed to pick out useful components.

Based on the analysis to the signal, the received signal can be roughly divided into three parts: strong direct current part, weak low-frequency part (0.5 to 1.2 Hz) and high-frequency noise. The direct current component corresponds to the drilling fluid pressure at the measure point, which is much stronger (1400 to 4000 mV) than the low-frequency component (10 to 200 mV) that

Status	T series						Resistivity detector	Gamma-ray detector
	1	2	3	4	5	6		
1	↑						Good	Good
2		↑					Good	Bad
3			↑				Bad	Good
4				↑			Bad	Bad

Figure 6 Status pulses and their significations. The signal format is 1 pulse with width 1.5 Ts in 6 Ts to tell the working status of down-hole equipments.



comes from the down-hole combined signal from which important parameters would be gained, and the high-frequency noise is more complicated. The pump noise is the main source of the noise during the signal transmission. In addition, it is worth noting that the low-frequency component is composed of positive pulses.

From the encoded signals we hope to acquire true down-hole conditions or accurate physical parameters. Therefore, before decoding the signal, the direct current component and the high frequency noise should be firstly separated or filtered out.

In view of the characteristics of the combined signal and the encoding and decoding method, we propose a two step filtering method before decoding the encoded signal. Firstly, a dynamic part mean filtering algorithm is proposed to separate the direct current components, and then a limited impulse response filtering algorithm is deployed to filter out the high-frequency noise. We have provided detailed implementations of these algorithms in [9].

Signal storage

During the well drilling process, the mud pump usually works a long time before shut down. The measurement-while-drilling system will produce a huge quantity of data, which bring a great challenge to the data storage and manage technology. On the one hand, the system should keep real time data so that accurate physical parameters could be acquired in time and correct decisions could be made as early as possible. On the other hand, the system should be capable to preserve all of the data so that technicians can access and refer to the old data when needed. Apparently a database technique can provide the function of data accessing and storage, but the huge quantity of data may result in slowness when the software is started and the data are accessed.

To solve the above problem this paper proposes a feasible scheme. A dynamic vector container is created in the memory to hold the current data, and the old data from the container are saved in files. At the beginning the filtered data (a sample per 50 ms) are allowed to get into the container. When the amount (samples) of data in the

container exceed a threshold (5 samples), the data are allowed to simultaneously flow out of the container in first-in-first-out order and saved into a file in the hard disk. When a pause operation is needed, no data are allowed to flow out of the container. When the container is almost full, the outflow of data from the container is controlled to be faster than the inflow in order to keep the data in the container newest. To make the play-back operation rapidly, each file will not exceed the size of 1 megabyte. When a file reaches 1 megabyte, a new file is created. In addition, in routine operation the container can not be empty to keep the output data continuous. With these skills, the current data can be decoded and displayed in time, and the old data are saved perfectly and can be selectively played back at any time.

Signal decoding

In this stage the software separates value sequences of down-hole parameters and converts the data back to their original binary values, from which the real physical values can be easily acquired.

Peak detection

Exact peak detection is important because the system recognizes synchronization and mode sections and confirms a value pattern according to the combination of pulses.

The simplest way to detect pulses is using a threshold. However, because of the disuniform in the pulse shapes, it is impossible to find a reasonable threshold used to find the right peaks. Another way is comparing the value of a current point with its n neighbors. If the current value is bigger than those of its n neighbors, then the current point is a peak position. However, the peaks of noise signals may also be included when using this method.

Therefore, this article utilizes a method that combines the above two ways to detect peaks. Only if the value of a current point is bigger than a predetermined threshold and bigger than those of its n neighbors (here $n = 4$), the current point will be considered as a peak position. The peaks are detected after a whole signal sequence has been transmitted over the time of $N T_s$.

Signal recognition

From the whole signal sequence, the synchronization pulses should be found firstly, and then the mode and status pulses are recognized. The rest is the data string. According to the pre-determined formats of parameters, the value of every parameter can be acquired.

(1) Synchronization section recognition

From Figure 4, we know the synchronization section uses a special format, with 3 pulses in 11 T_s and the interval between two adjacent pulses is 3.5 T_s . This is

different from those of mode pulses, status pulse and data pulses.

(2) Mode section recognition

After successfully capturing synchronization pulses, the software begins to recognize the mode pulses. In the mode section, the first pulse is a wide pulse so the both states of T1 and T2 are all 1, which is unique in the whole sequence. Other cases of two adjacent locations with state 1 at the same time are illegal. When recognizing, we can also consider this wide pulse as a part of the synchronization section. The other two pulses in the mode determine the current data mode.

Figure 8 gives a part of combined signal, where the first 3 pulses are recognized as synchronization pulses, the forth pulse is the beginning of the mode section, and the last 2 pulses reveal that the current data mode is mode 1.

(3) Status section recognition

Status section is only used in mode 9. It is easily to be recognized because it has only one pulse with width 1.5 Ts in 6 Ts and after the mode pulses.

(4) Data recognition and decoding

The system continues to go into the stage of data recognition. The data string is after the mode section or the status section. Because the mode has determined the data's components, then the pulse pattern of every parameter will be easily extracted and the pattern number will be known accordingly.

Decoding is a reverse process of encoding. Having got the current code pattern with M pulses in N Ts, the next is to convert the pattern to its equivalent binary source code and the actual value can also be required. In fact, the simplest way is to multiply the pattern number by the value represented by one binary bit. For

example, a tool face angle (0 to 360°) is represented by using a 7 bits binary source code. Here $360^\circ/2^7 = 2.8125^\circ/\text{bit}$. If the surface computer has known the pattern number of a tool-face angle is 2, then the actual value is $2 * 2.8125^\circ = 5.625^\circ$. The following is one decoding result of a combined signal with mode 1.

```
2009-05-24,09:03:41 SYN //synchronization
2009-05-24,09:03:47 Mode 1 //mode
2009-05-24,09:03:53 ATF: 2.8° //tool-face angle
2009-05-24,09:03:59 ATF: 2.8° //tool-face angle
2009-05-24,09:04:08 GX: 0.0024 //gravity-x
2009-05-24,09:04:17 GY: 0.0465 //gravity-y
2009-05-24,09:04:25 GZ: 0.9995 //gravity-z
2009-05-24,09:04:34 BX: 23.43 //magnetic-x
2009-05-24,09:04:43 BY: -7.59 //magnetic-y
2009-05-24,09:04:52 BZ: 68.00 //magnetic-x
2009-05-24,09:04:58 RPM: 0.0 rpm //rotate speed
2009-05-24,09:05:04 TMP: 26.3°C //temperature
degree centigrade
```

This section mainly introduces the surface computer processing techniques to the signal from the down-hole measurement system. Especially a two step filtering method and a dynamic data storing and managing method are proposed. A mix peak detection method is utilized to find the position of a pulse by combining threshold method and neighbor comparison method.

Conclusion

This article introduces the data processing techniques for a wireless data communication via mud, which includes the down-hole part and the surface part. As for the down-hole data processing techniques, data encoding and signal integrating method are mainly introduced. A pattern of M pulses in N slots is used to express a value. The data of multi-parameters are encoded and integrated with synchronization, mode and status signals to produce a sequence of mud pulses, which is transferred to the surface computer. With regard to the surface data processing techniques, signal filtering, storage and manage method, peak detection, sequence recognition, and data decoding are discussed. Although the software is capable to provide a graphical and numerical view of the raw, filtered, and decoded data, in this article, we principally discuss the signal or data processing techniques instead of the view and the interface.

Abbreviations

DC: direct current.

Competing interests

The authors declare that they have no competing interests.

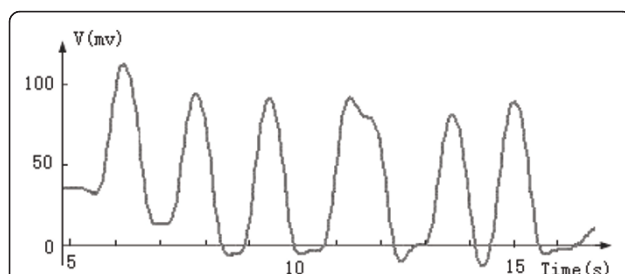


Figure 8 An example of synchronization and mode pulses. The first 3 pulses are recognized as synchronization pulses, the forth pulse is the beginning of the mode section, and the last 2 pulses reveal that the current data mode is mode 1.

Received: 11 April 2011 Accepted: 23 August 2011
Published: 23 August 2011

References

1. P Tubel, C Bergeron, S Bell, Mud pulser telemetry system for down hole measurement-while-drilling, in *Proceedings of the 9th IEEE Instrumentation and Measurement Technology Conference*, 219–222 (1992)
2. JH Cohen, G Deskins, W Motion, J Martin, Development of a mud-pulse high-temperature measurement-while-drilling (MWD) system, <http://www.osti.gov/bridge/servlets/purl/828406-9vCPaN/native/828406.pdf>
3. LL Kan, BK Gao, XF Wang, HW Liang, Development of a high precision MWD system, in *Proceedings of the 8th IEEE International Conference on Electronic Measurement and Instruments* **1**, 951–954 (2007)
4. AG Ledroz, E Pecht, D Cramer, MP Mintchev, FOG-based navigation in down-hole environment during horizontal drilling utilizing a complete inertial measurement unit: directional measurement-while-drilling surveying. *IEEE Trans Instrum Meas.* **54**(5), 1997–2006 (2005). doi:10.1109/TIM.2005.853562
5. E Pecht, MP Mintchev, Observability analysis for INS alignment in horizontal drilling. *IEEE Trans Instrum Meas.* **56**(5), 1935–1945 (2007)
6. CW Li, DJ Mu, AZ Li, Drilling mud signal processing based on wavelet, in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition*, 1545–1549 (2007)
7. RW Tennent, WJ Fitzgerald, Passband complex fractionally-spaced equalization of MSK signals over the mud pulse telemetry channel, in *Proceedings of the 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, 5–8 (1997)
8. JH Zhao, LY Wang, F Li, YL Liu, An effective approach for the noise removal of mud pulse telemetry system, in *Proceedings of the 8th International Conference on Electronic Measurement and Instruments*, 971–974 (2007)
9. QJ Zhao, BJ Zhang, HS Hu, Novel two-step filtering scheme for a logging-while-drilling system. *Comput Phys Commun.* **180**, 1566–1571 (2009). doi:10.1016/j.cpc.2009.04.011

doi:10.1186/1687-6180-2011-45

Cite this article as: Zhao et al.: Data processing techniques for a wireless data transmission application via mud. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:45.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com