Int. J. Communications, Network and System Sciences, 2016, 9, 126-134
Published Online May 2016 in SciRes. http://www.scirp.org/journal/ijcns
http://dx.doi.org/10.4236/ijcns.2016.95011

Parallel Programming Design of BPSK Signal Demodulation Based on CUDA

Yandu Liu, Baoling Zhang, Haixin Zheng

Equipment Academy, Beijing, China
Email: liuyandu1029@126.com

Received 12 April 2016; accepted 24 May 2016; published 30 May 2016

Abstract

Realizing digital signal demodulation on the general computer is an important research direction in the field of signal processing in recent years. In this paper, the algorithm of BPSK signal demodulation which has high real-time requirements is researched on the general computer. According to the characteristics of "CPU + GPU" heterogeneous computing, the parallel computation model of digital communication is put forward, and BPSK signal demodulation is realized on CUDA platform. Test results show that the computing time ratio of 1:1.7, when $E_b/N_0 = 9.6dB$ the bit error rate can be achieved 10^{-5}.

Keywords

BPSK, Demodulation, CUDA, Parallel

1. Introduction

In recent years, with the constant improvement of the general computer performance, experienced from hardware platform towards digital platform of software radio technology, the platform of digital signal processing in communication system is beginning to change the direction of development. The signal after the A/D directly complete real-time processing in pure software processing way based on general computer platform.

Digital Phase modulation, namely Phase Shift Keying (Phase Shift Keying, PSK), is a very important basic digital modulation technology, which using carrier Phase modulation technique information to express input signal. Under the condition of stability channel, phase shift keying compared with amplitude shift keying, frequency shift keying, not only has high noise resistance, but also can effectively use band, even in a phenomenon of fading and multipath channel also has a good effect [1]. Therefore, BPSK is a kind of excellent modulation method, and in medium and high speed data transmission has been widely applied. This paper is based on "CPU + GPU" heterogeneous platform, the real-time BPSK signal demodulation algorithm and the method based on CUDA parallel programs are researched. In view of the implementation, parallel programming test verify the feasibility of the system.

2. BPSK Signal Demodulation Algorithm

By multiple BPSK signal is coherent demodulation method based on phase lock loop, such as square ring me-

How to cite this paper: Liu, Y.D., Zhang, B.L. and Zheng, H.X. (2016) Parallel Programming Design of BPSK Signal Demodulation Based on CUDA. *Int. J. Communications, Network and System Sciences*, **9**, 126-134.
http://dx.doi.org/10.4236/ijcns.2016.95011

thod, decision feedback method, Costas loop method, etc. The differential demodulation method which use adjacent element phase jump is also used [2]. Although differential demodulation does not need to obtain coherent carrier, the algorithm is relatively simple, but its anti-noise performance significantly worse in the coherent demodulation. As the GPU is widely used in signal processing, coherent demodulation which has excellent performance is easy to implement. Costas loop is the most widely used suppressed carrier tracking loop in engineering, literature [3] prove its track suppress carrier signal with low SNR is the best device, its structure as shown in **Figure 1**.

The input BPSK modulation signal is [4]:

$$\begin{aligned} s(t) &= m(t)\cos\left[\omega_c t + \theta_1(t)\right] \\ &= \left[\sum_n a_n g(t - nT_s)\right]\cos\left[\omega_c t + \theta_1(t)\right] \end{aligned} \tag{1}$$

Here, $m(t)$ is digital modulation signals; $\omega_c(t)$ is carrier angular frequency. The local oscillator output respectively are:

$$\begin{aligned} v_Q(t) &= \cos\left[\omega_c t + \theta_2(t)\right] \\ v_I(t) &= \sin\left[\omega_c t + \theta_2(t)\right] \end{aligned} \tag{2}$$

Here, $\omega_c(t)$ is variable frequency signal produced by the local oscillator, $\theta_1(t)$ and $\theta_2(t)$ are reference phase. After under orthogonal frequency conversion, the output is:

$$\begin{aligned} z_q(t) &= K_{p1}\left[\sum_n a_n g(t - nT_s)\right]\sin\left[\omega_c t + \theta_1(t)\right]\cos\left[\omega_c t + \theta_2(t)\right] \\ z_i(t) &= K_{p2}\left[\sum_n a_n g(t - nT_s)\right]\sin\left[\omega_c t + \theta_1(t)\right]\sin\left[\omega_c t + \theta_2(t)\right] \end{aligned} \tag{3}$$

make $\theta_e(t) = \theta_1(t) - \theta_2(t)$, then K_{p1}, K_{p2} is multiplication coefficient, after low pass filtering:

$$\begin{aligned} y_q(t) &= \frac{1}{2}K_{p1}K_{11}\left[\sum_n a_n g(t - nT_s)\right]\sin\left[\theta_e(t)\right] \\ y_i(t) &= \frac{1}{2}K_{p2}K_{12}\left[\sum_n a_n g(t - nT_s)\right]\cos\left[\theta_e(t)\right] \end{aligned} \tag{4}$$

Here, K_{11}, K_{12} is low pass filter coefficient. The result after filtering, and the in-phase and orthogonal branch phase discrimination is:

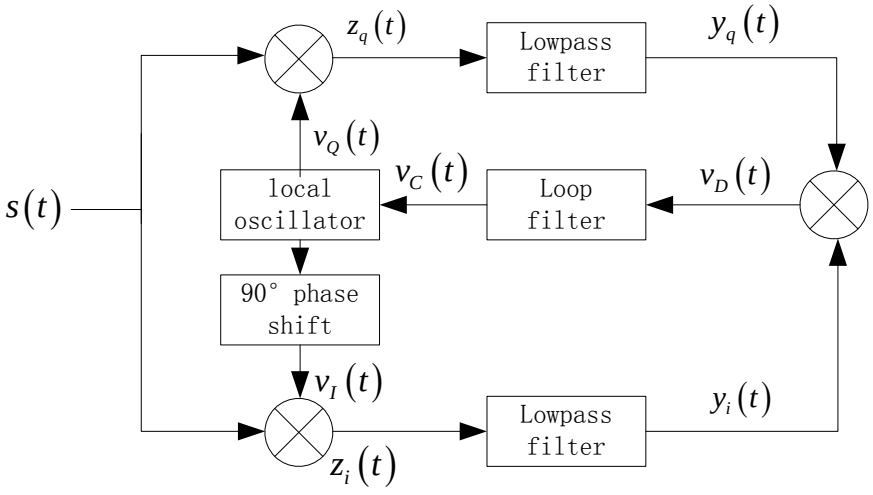

Figure 1. Costas loop structure.

$$v_c(t) = \frac{1}{8} K_p K_{p1} K_{p2} K_{11} K_{12} \sin\left[2\theta_e(t)\right] = K_d \sin\left[2\theta_e(t)\right] \quad (5)$$

K_p is gain of phase discrimination, K_d is loop gain, the output of loop filter is error signal for tracking $\theta_e(t)$.

According to the principle of coherent demodulation, extracted coherent carrier multiplied by the input of the modulated signal directly, and filtering the output, baseband signal waveform can be got (**Figure 2**).

And it can follow 25 KHZ dynamic Doppler (**Figure 3**).

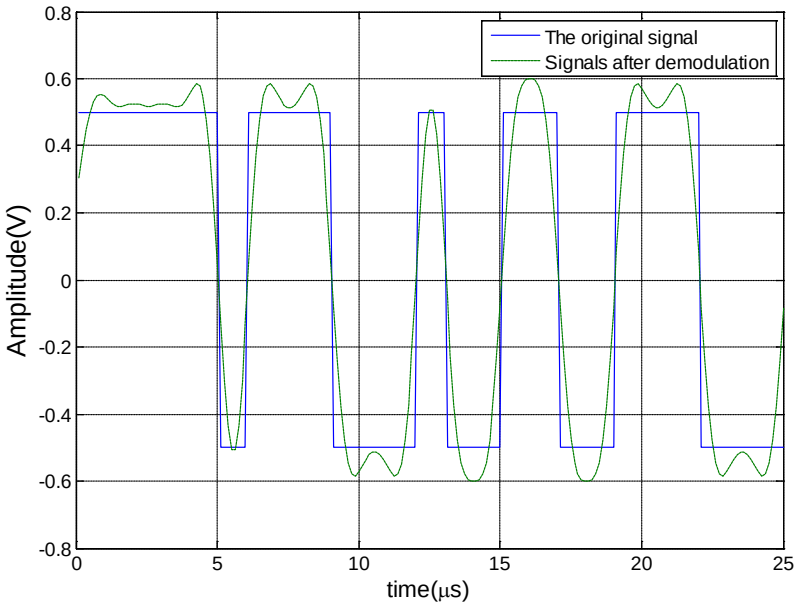

Figure 2. BPSK signals after demodulation.

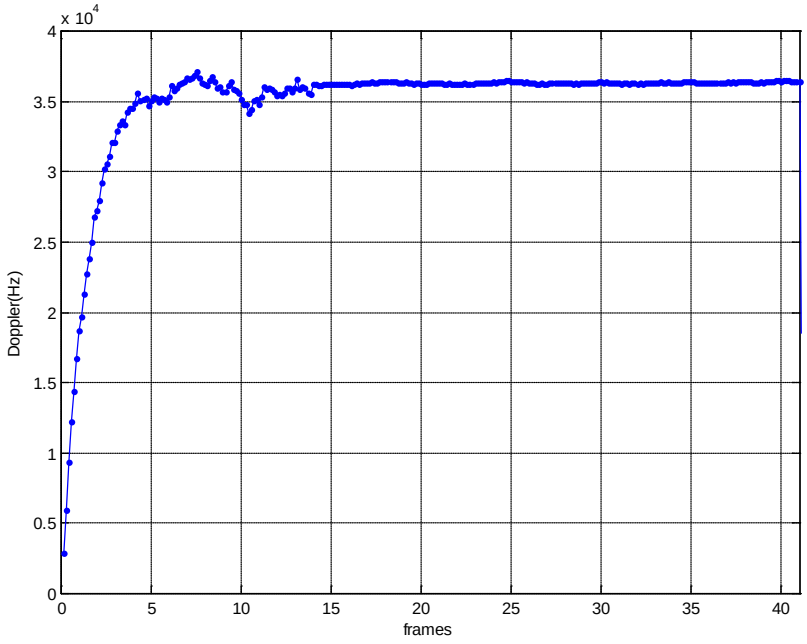

Figure 3. Follow dynamic doppler.

3. The Parallel Computing Model Based on CUDA

3.1. CUDA

Launched by NVIDIA, CUDA is a kind of general parallel computing architecture, initial designed to speed up image real-time processing which run on the GPU development platform and full use of GPU's high memory bandwidth and very large scale of floating point calculation unit. It can handle large parallel problems, especially large-scale floating point data computing [5]. CUDA hardware architecture as shown in **Figure 4**.

GPU is specially designed for the intensive and high parallelism computation, so calculation of the design will therefore more transistors used in data processing rather than data caching and flow control. In particular, the GPU is very suitable for processing the same program on multiple data parallel execution problem, so in CUDA platform is more suited for digital signal processing.

3.2. Parallel Computing Model

Parallel computing is treated with multiple core to solve the problem at the same time. For digital signal processing which has a high requirement of real-time, parallel computing is the effective way to improve the real-time performance. Currently, the most widely used parallel computing model is a layered model which consists of three layers [6] [7].

Parallel Algorithm Design Layer: abstracted the calculation parameters of from different parallel computers, parallel algorithm design model is established, this layer mainly oriented algorithm researchers.

Parallel Programming Design Layer: according software and hardware interface, using parallel programming language programming to achieve specific parallel algorithm, this layer is mainly oriented program designers.

Parallel Program Execution Layer: under the system supports parallel machine compiler running target code, and the actual performance of the optimization procedure (**Figure 5**).

According to the GPU hardware design characteristics, CUDA in layer parallel algorithm design has made a more detailed. Model assumes that the CUDA thread in physically separate GPUs execute, GPU as host coprocessor, adopt heterogeneous parallel mode, parallel computing program execute on GPU kernel, and the rest of the program execute on the CPU. And the research category of parallel program execution is a compiler, therefore, this paper mainly studies the parallel programming problem.

3.3. Digital Communication Parallel Computing Model

Parallel algorithm is the core issue of parallel programming, and algorithm belongs to numerical parallel algorithm of digital communication system. Its design method is generally has two kinds: 1) direct parallelization of serial algorithm. Fully exploiting and utilization of the existing serial algorithm of parallelism, directly to the serial algorithm for parallel algorithm; 2) based on calculation and numerical calculation principle, does not take into account the corresponding serial algorithm, redesigned to parallel algorithm [8].

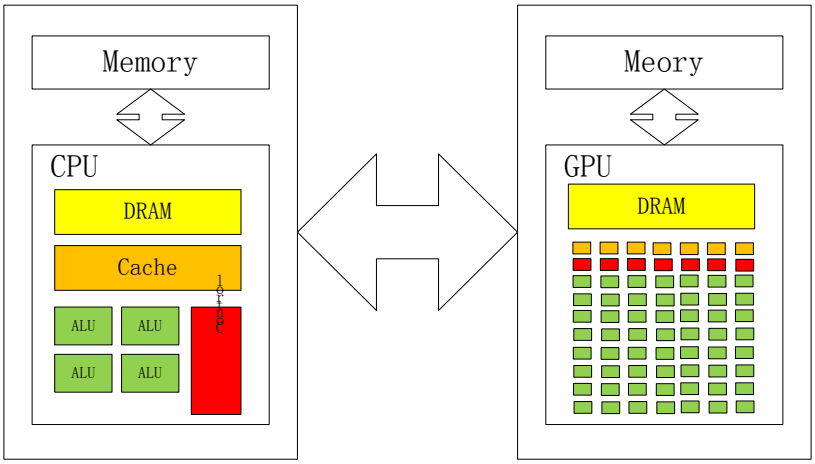

Figure 4. CUDA Architecture.

Digital communication system has a high modular degree and large amount of calculation which typical structure as shown in **Figure 6**.

Because GPU device cannot display, data needs to be interacted between memory and memory by PCIe bus. And restricted to general computer speed limit, in the large-scale numerical calculation, the data transmission time occupy most of the program execution time. **Figure 7** shows under different scale of data parallel computation time, the data size is small, transferred time almost occupied more than 99% of the program total execution time. Therefore, only when calculating the larger scale, to reflect the advantage of GPU computation.

According to this characteristic of CUDA platform, parallel computing model of digital communication system should try to reduce the data transmission, give full play to the GPU high-performance computing ability.

At the program beginning, the data need to be deal with should all transfer to memory of GPU. All the mass calculation performed by GPU. CPU and GPU in the process of program execution, only a small amount of data transmission, the CPU only run small calculation and data monitoring and display function (**Figure 8**).

4. BPSK Signal Demodulation Parallel Programming

4.1. BPSK Signal Demodulation Algorithm Structure

According to Section 3.3 of the parallel computing model and Costas loop demodulation structure, parallel BPSK demodulation algorithm are shown in **Figure 9**.

Intermediate frequency sampling data read and transferred by CPU to the GPU, completed the functions of digital orthogonal frequency conversion, low-pass filtering, bit synchronization, phase detector, loop filter and decoding in the GPU. The phase error signal filtered by loop filter transfer back to CPU to compute doppler

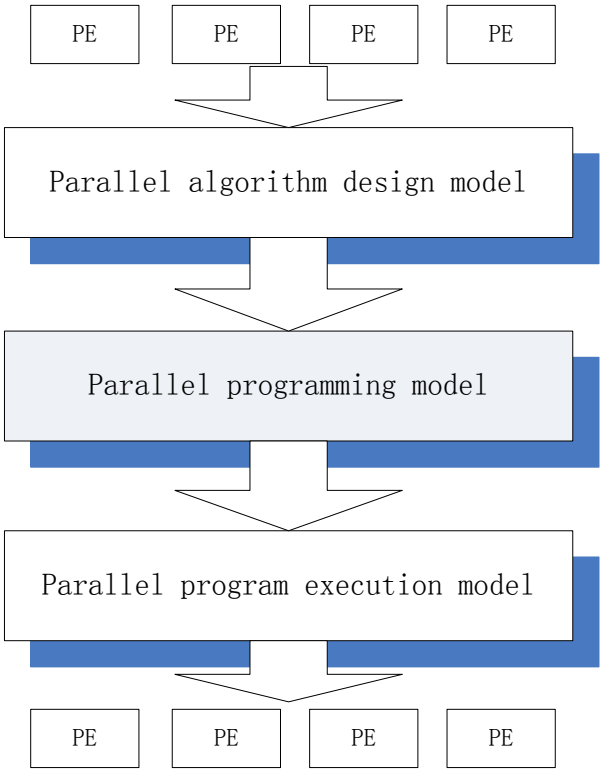

Figure 5. Parallel computing model.

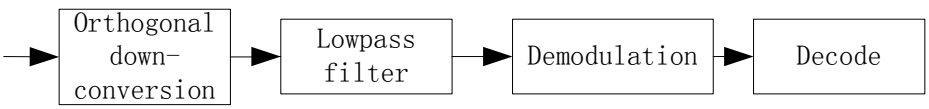

Figure 6. Digital communication model.

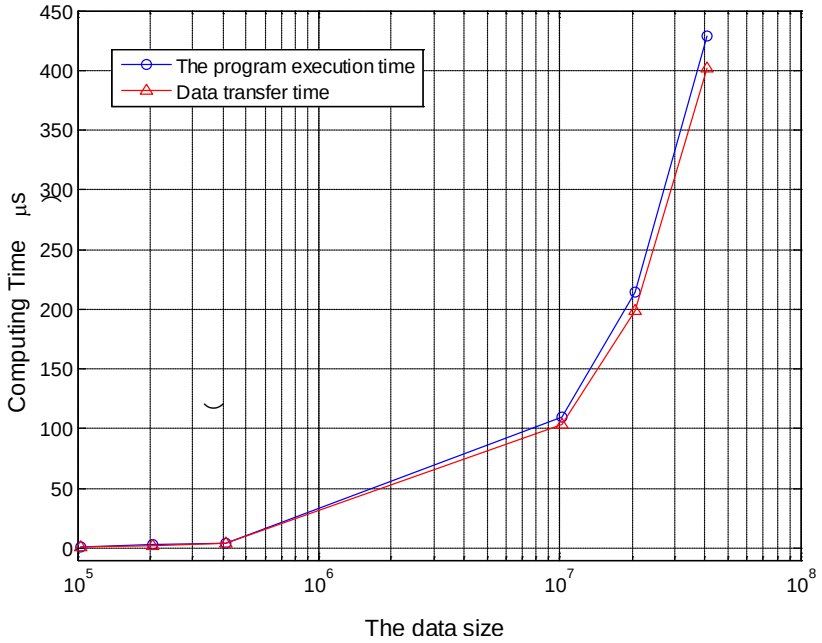

Figure 7. Computing time compared with transmission time.

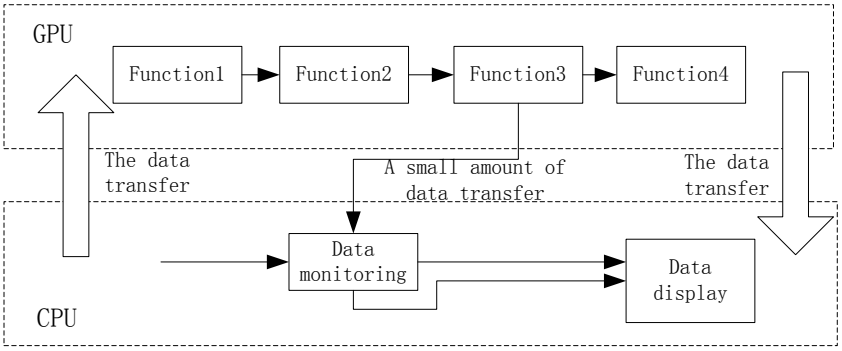

Figure 8. Parallel computing model of digital communication.

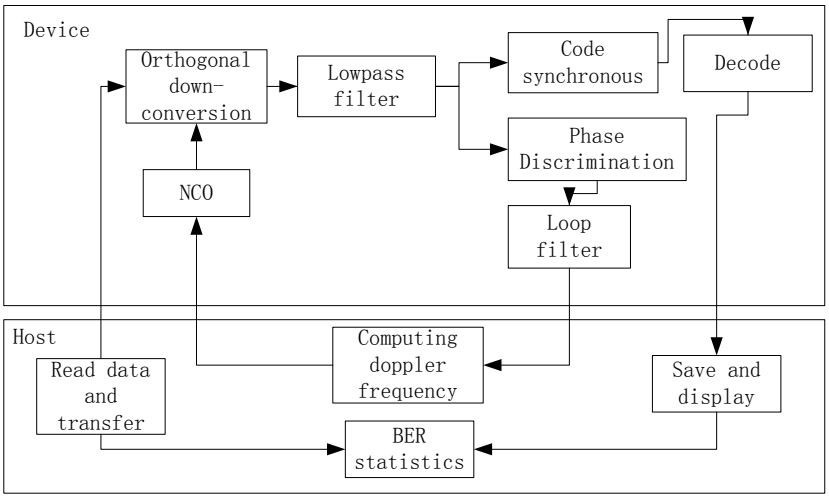

Figure 9. Parallel computing model of BPSK single demodulation.

frequency shift. Then the doppler frequency shift transfer to the GPU again to correct the output sine and cosine waveform produced by NCO. Lastly, the data decoded by GPU transfer back to the host and statistical BER.

4.2. The Mixer Design

Mixer convert the signal from the intermediate frequency to fundamental frequency, which is the core of the software defined radio. Numerical control oscillator (NCO) is usually used to produce local hardware digital carrier for mixing. When programming parallel mixing programme, the corresponding data points with the corresponding phase of the sine and cosine waveform sampling points to do multiplication, application pseudo code is as follows:

```
__global__ void DownfreqKernel()
Begin
   int tid = blockIdx.x * blockDim.x + threadIdx.x;
   if (tid<size)
          downfreq[tid].x = chan-
data[tid]*cos(2*PI*(fb+phasefd));
          downfreq[tid].y = chan-
data[tid]*sin(2*PI*(fb+phasefd));
End
```

4.3. The FIR Filter Design

FIR filter is widely used for its good group delay in the digital communication system, it can ensure any amplitude frequency characteristics of strict linear phase frequency characteristics at the same time. It has a finite impulse response at the same time. Finite length for M FIR filter transfer function for:

$$H(z)=\sum_{k=0}^{M}h(k)z^{-k} \tag{6}$$

In the time domain, the limited impulse to the corresponding input and output

$$y(n)=\sum_{i=0}^{M}h(k)x(n-k) \tag{7}$$

The parallel filter application pseudo code is as follows:

```
__global__ void  GPUFilterKernel()
Begin
          __shared__ float2 cache[];
          int tid = threadIdx.x + blockIdx.x * blockDim.x;
          cache[threadIdx.x].x=signal[tid].x;
          cache[threadIdx.x].y=signal[tid].y;
          __syncthreads();
          float sumx=(cache[threadIdx.x].x+
          ……+cache[threadIdx.x+16].x)*f.a;
     float sumx=(cache[threadIdx.y].y+
          ……+cache[threadIdx.x+16].x)*f.a;
          result[tid].x=sumx;
          result[tid].y=sumy;
End
```

4.4. The Phase Discriminator Design

Phase discriminator is mainly done to identify the function of the input signal is differ, is the key to the phase lock loop (PLL), in parallel programming, rely mainly on solving the sample point difference before and after,

application pseudo code is as follows:

```
__global__ void ComputeSubphaseKernel()
Begin
        int tid = threadIdx.x + blockIdx.x * blockDim.x;
    data[tid] = atan2(-src[tid].y,src[tid].x);
        data[tid] = src[tid+1]-src[sid];
        if (data[tid]>PI)
                data[tid] -= 2*PI;
        else if (data[tid]<-PI)
                data[tid] += 2*PI;
End
```

5. Conclusion

Test hardware platform selected Tesla K20 graphics. The size of input data is 1 ms analog data. The test computation time is within 1.7 ms, as shown in **Figure 10**, the program can correct demodulation of the original data, BER statistics as shown in **Figure 11**.

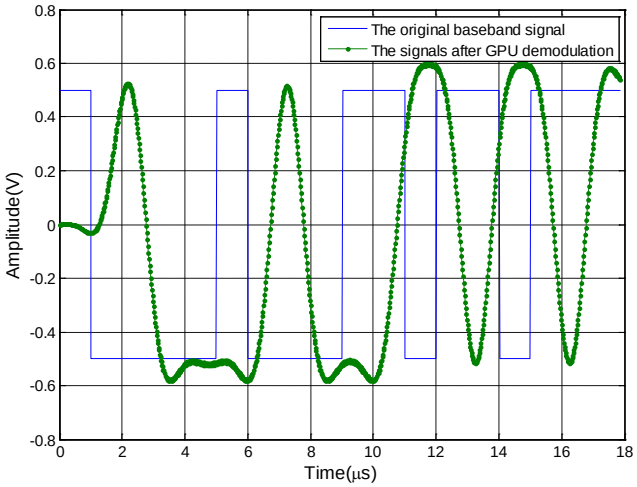

Figure 10. Signals after GPU demodulation.

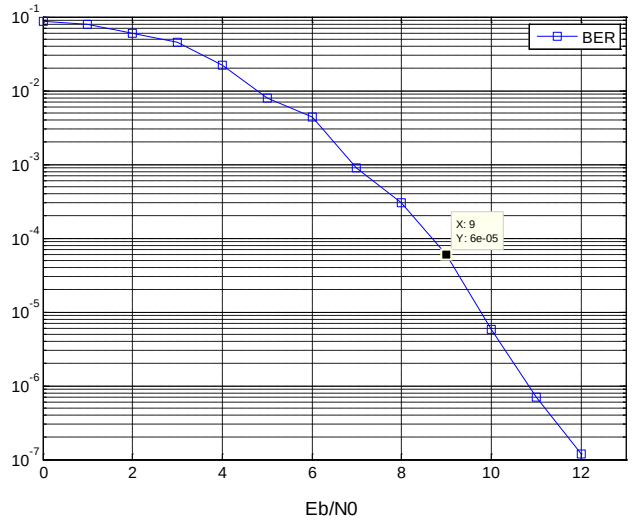

Figure 11. BER statistics.

Realizing BPSK signal demodulation on general computer platform, reducing the difficulty of system design, development and cost. And the software of processing way increasing the flexibility of the system by loading different software can realize more functions. Through hardware upgrades and reorganization, the system performance can be further improved [9]-[11]. Based on general computer platform, especially the digital signal processing based on CUDA is an important development direction of the signal processing, but also a new trend of computer application and new research areas.

References

[1] Riter, S. (1969) An Optimum Phase Reference Detector for Fully Modulated Phase Shift Keyed Signal. *IEEE AES*-5, **4**.

[2] Core, M.T. and Tan, H.H. (2002) BER for Optical Heterodyne DPSK Receivers Using Delay Demodulation and Integration Detection. *IEEE Transactions on Communications*, **50**.

[3] LI., G.X., An, Z.Q. and Yuan, S.J. (2008) Study on Software Demodulation of DQPSK Signal Based on Digital Phase Measurement. *Journal of Spacecraft TT&C Technology*, **27**.

[4] Mitra, S.K. (2001) Digital Signal Processing, A Computer-Based Approach. 2nd Edition, MeGraw-Hill Companies, Inc.

[5] NVIDIA CUDA Programming Guide 5.0. http://www.nvidia.com/object/cuda_Develop.html

[6] Sankaralingam, K., Keckler, S.W., Mark, W.R. and Burger, D. Universal Mechanisms for Data-Parallel Architectures. 36*th Annual International Symposium on Microarchitecture*.

[7] Chen, G.L., Sun, G.Z., Xu, Y. and Lu, M. (2008) Methodology of Research on Parallel Algorithms. *Chinese Journal of Computers*, **31**.

[8] Chen, Y., Wang, Y.Q. and Liu, Y. (2011) Research on the Technology of Software Space TTC System Based on Computer Platform. *The Measurement and Control Technology*, **30**.

[9] Bose, V.G. (1999) Design and Implementation of Software Radios Using a General Purpose Processor. Ph.D. Thesis, Massachusetts Institute of Technology.

[10] Bose, V.G. and Morris, R. (2001) Dynamic Physical Layers for Wireless Networks Using Software Radio. *International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT May 2001. http://dx.doi.org/10.1109/icassp.2001.940393

[11] Vaudtabatgabrn, P.P. (1990) Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications. *Proceedings of the IEEE*, **78**, 56-93