

# Grubbs and Wright Algorithms For WiMAX and LTE Systems

Wu Yin<sup>†</sup>, Pei Xiao<sup>††</sup>, Colin Cowan<sup>††</sup>, Liujun Hu<sup>†</sup>

<sup>†</sup>ZTE Corporation, Shenzhen, China.

E-mail: {yin.wu, hu.liujun}@zte.com.cn

<sup>††</sup>Institute of ECIT, Queen's University Belfast, UK

E-mail: {pei.xiao, c.cowan}@ecit.qub.ac.uk

**Abstract**—In this paper, we address the problem of interference mitigation with data pre-processing in the 4G uplink systems, and propose to use the Grubbs/Wright algorithm to detect and remove the interference contaminated data. The Malikov algorithm is also applied to correct the system errors. The pre-processed data are used for channel estimation and data detection in base stations.

## I. INTRODUCTION

There has been a rapid development in the 4th generation (4G) wireless communication systems, such as WiMAX (Worldwide Interoperability Microwave Access) and LTE (Long Term Evolution). Multiple input, multiple output (MIMO) and orthogonal frequency division multiplexing (OFDM) have become increasingly popular techniques in wireless communications due to their ability to dramatically improve the performance and efficiency of a wireless system.

The performance of a 4G system deteriorates significantly in the presence of interference and channel variations, such as inter-cell, intra-cell interference, frequency offset, Doppler shift due to the movement of the mobile users, etc. In the base station radio frequency (RF) system, the phase offset due to internal cables, RF coupling devices, printed circuit board (PCB) in the antenna amplifier unit will bring additional errors. These problems will largely reduce the reliability of the uplink transmission link and has an adverse effect on the accuracy of the channel measurement.

The interference also causes periodic or gross errors in the received data. In the current 4G uplink systems, the unprocessed data are used for channel estimation purpose. The received data might have large errors due to interference and system deficiencies. Various interference cancelation and phase/frequency offset correction techniques have been proposed in the WiMAX/LTE systems. However, most existing solutions depend on accurate channel estimation and received data. How to guarantee the accuracy of received data in the uplink physical channel remains unaddressed in the literature. The Grubbs [1] and

Wright ( $3\sigma$ ) algorithms [2], [3] have previously been used in the fields of Aeronautics, Astronautics and Medicine to solve the problem of periodic or gross errors caused by interference [4]. However, they have not yet been applied to wireless communication systems which require proper handling of the received data in order to ensure satisfactory system performance. In this paper, we investigate data pre-processing and outlier detection techniques which are needed to address the above issues. In particular, we apply the Grubbs and Wright algorithms to detect the erroneous data and replace them with the data obtained from curve fitting so as to improve the system performance. The remainder of the paper is organized as follows. The proposed schemes are introduced in Section II and the experiment results are shown in Section III. The conclusions are drawn in Section IV based on the demonstrated results.

## II. PROPOSED ALGORITHM

Fig. 1 is the block diagram for uplink data processing in the base station. Denote  $r_1(i), r_2(i), \dots, r_m(i)$  as the uplink pilots or channel sounding signals transmitted via a fast feedback (FFB) channel at the time instant  $i$ . For simplicity, we use equal gain combining, the received signal can be expressed as  $y_i = \sum_{j=1}^m r_j(i)$ , where  $m$  is the number of receive antennas. In this context, the uplink data refers to all the data in the uplink data frame excluding the one in ranging access time slots. As shown in Fig. 1, the received data is processed in the radio remote unit (RRU). We first remove the cyclic prefix (CP) and perform Fast Fourier Transform (FFT), then detect the erroneous data and perform error correction/compensation. The processed data will be used for the estimation of the wireless channels and frequency offset as well as for subsequent data demodulation and channel decoding.

We propose a novel data pre-processing method to mitigate the effect of the interference and system errors. The proposed procedure includes: 1) retrieving data transmitted in the uplink system; 2) processing data and removing the corrupted data using the Grubbs

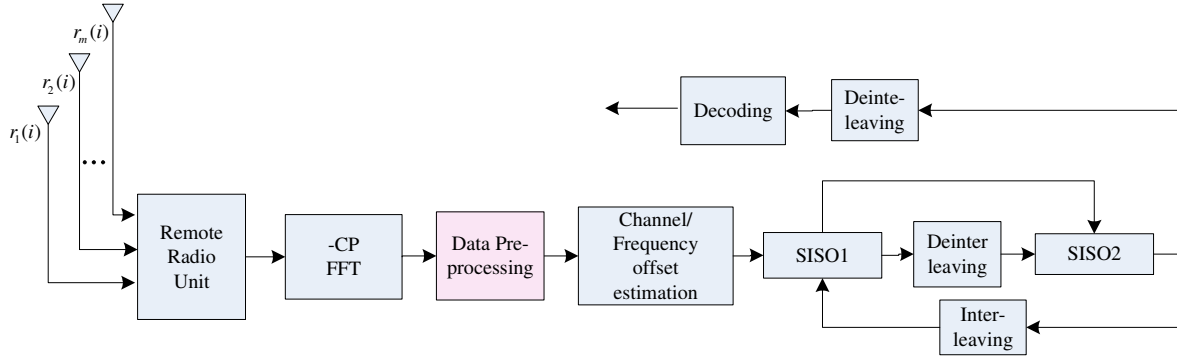


Fig. 1. Uplink data processing in base station.

or Wright algorithm; 3) performing curve fitting with the processed data and replacing the corrupted data with the fitted data.

The following operations need to be performed at least once during the 3rd step: calculating the residual error and standard deviation as well as their ratio from the measured data; searching for threshold value in the fiducial probability table based on the ordered sample index. If the ratio between the residual error and standard deviation is greater than the threshold, the data is regarded as an outlier and will be subsequently removed. The data pre-processing is performed on the array of data consisting of the same pilot subcarriers over several time slots.

In order to correct system errors, we use the Malikov algorithm to replace the contaminated data with the fitted data and process the pilot and data subcarriers within one or several OFDM symbols, and compensate/correct the measured data if system errors occur. We divide the data sequence into multiple segments during the system initialization stage and perform data fitting for each individual segment.

We use WiMAX as an example to explain this procedure in details. In the WiMAX uplink system, each data frame contains 5-7 time slots, which are equivalent to 15-21 OFDM symbols. Each OFDM symbol in one time slot has a 6-tile structure, each tile contains 4 partial usage sub-channels (PUSC) OFDM pilot subcarriers for the estimation of channel state information (CSI). The pilot subcarriers in each tile can be divided into two parts based on their positions, each contains two subcarriers. Therefore, each uplink OFDM symbol has 24 pilot subcarriers for CSI estimation. The uplink channel can be measured by an auto-correlator (conjugate multiplication and linear interpolation) to correlate two pilot subcarriers from the same uplink tile, then the frequency offset

is estimated, and the base station will be instructed to compensate the frequency offset accordingly based on the measured offset value.

The Grubbs and Wright ( $3\sigma$ ) algorithms provide us effective tools to detect the erroneous data with limited measurements. The basic principle is to use the statistic information to iteratively estimate the data which follow a normal distribution due to the central limit theorem for a sufficiently large number of independent samples, and to effectively detect the outliers from the measured data using Bayesian and Least Mean Square algorithms. The Wright scheme is mainly used for detecting errors from a large amount of measured data; whereas the Grubbs' scheme is suitable for processing data of small or medium sizes. These two methods can be combined with 4G communication technologies in different operation environments.

Fig. 2 shows the flow chart of the proposed data pre-processing using the Grubbs/Wright algorithm. First, we form a data array consisting of the pilot subcarriers at the same position of each tile from one or several time slots, then calculate the expectation value, variance, and standard deviation of the measured data. The elements in the data array are properly ordered such that  $y_1 < y_2 < \dots < y_n$ . The residual error that reflects the level of deviation will also be calculated. We then set a fiducial probability and fiducial region, and the Grubbs' threshold value  $N(\sigma, i)$  or the Wright's standard variance  $3\sigma$ , based on which all the outliers which do not follow the normal distribution will be detected and removed. If the ratio between the residual error and standard deviation is larger than the threshold, the data will be removed. The above procedure can be repeated several times.

Table I shows an example of the Grubbs' fiducial probability and fiducial regions,  $i$  and  $p$  represents the sample index and fiducial probability, respectively. The

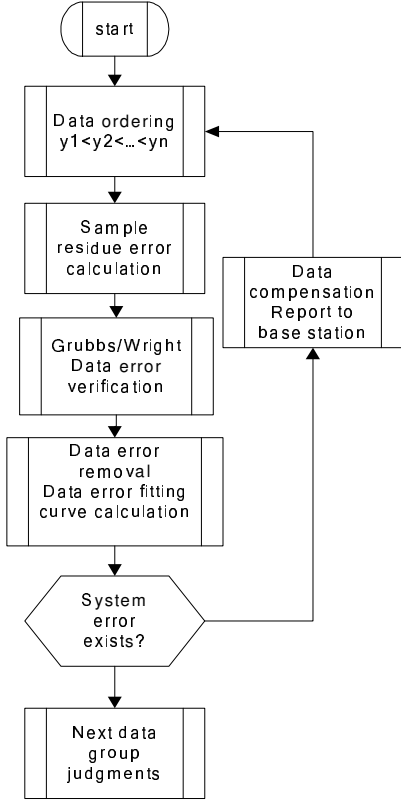


Fig. 2. Flow chart of the data pre-processing.

TABLE I  
GRUBBS FIDUCIAL PROBABILITY.

| i | $p = 99\%$ | i  | $p = 99\%$ |
|---|------------|----|------------|
| 1 | 1.17       | 10 | 2.58       |
| 2 | 1.51       | 11 | 2.64       |
| 3 | 1.77       | 12 | 2.67       |
| 4 | 1.96       | 13 | 2.74       |
| 5 | 2.12       | 14 | 2.79       |
| 6 | 2.25       | 15 | 2.81       |
| 7 | 2.35       | 16 | 2.84       |
| 8 | 2.44       | 17 | 2.87       |
| 9 | 2.51       | 18 | 2.91       |

fiducial probability is set to be  $p = 99\%$  here taking into consideration the accuracy and complexity of a real-time application. The fiducial probability can be formulated according to the Bayesian algorithm as [3]

$$p = \frac{y_i - \bar{y}_{i-1}}{y_i - \bar{y}_2}; \quad i \geq 3. \quad (1)$$

The standard deviation of the measured samples can be derived according to the Bessel formula as

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2)$$

TABLE II  
TESTING PARAMETERS.

| Parameters                   | Value     |
|------------------------------|-----------|
| Direction                    | uplink    |
| Bandwidth                    | 5 MHz     |
| FFT size                     | 512       |
| Modulation                   | PUSC      |
| Number of uplink subcarriers | 35        |
| Uplink burst symbols         | 3         |
| Uplink time slots            | 105       |
| BS receive antennas          | 4         |
| MS transmit antennas         | 1         |
| Wireless channel             | ITU Ped B |

where  $n$  is the number of measured samples. When  $n > 20$ , the measured data approach a normal distribution. The expectation value of the measured samples can be calculated as  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . Let us denote  $N(\sigma, a)$  as the ratio between the residual error and standard deviation, i.e.,

$$N(\sigma, a) = \frac{\max\{|y_i - \bar{y}|\}}{s}. \quad (3)$$

According to (3), if the measured data sample  $y_i$  is greater than the fiducial degree in the fiducial probability and the threshold  $N(\sigma, a)$  in the fiducial region, this data sample is corrupted by interference, and will be replaced by the data obtained from curve fitting.

For a large amount of measured data, the Wright criterion can be expressed as  $|y_m - \bar{y}| \leq 3\sigma$ . The Wright scheme proceeds as follows: we first calculate the expectation value of the measured data, then keep the data whose fiducial probability is within 3 times of the absolute expectation value, remove the data otherwise.

### III. EXPERIMENT RESULTS

The parameters for the testing environment are summarized in Table II. In this work, all the experiments are conducted with the Azimuth MIMO channel emulator (shown in Fig. 3) in various channel scenarios such as the international telecommunication union (ITU) [6] and spatial channel modelling extended (SCME) [7] channel models. When channel sounding technique is used, the uplink and downlink channels are calibrated in advance to guarantee the channel reciprocity.

Fig. 4 shows the received uplink frame data in 5 time slots, where the first slot is used for the base station ranging access; the other 4 time slots (12 OFDM symbols or 120 data samples) are used for channel estimation.

Fig. 5 shows the Grubb's parameter at the first iteration, i.e., the ratio between the residual error and standard deviation. The ratio for the 41st sample is beyond the Grubbs' threshold value, it is therefore



Fig. 3. The Azimuth MIMO channel emulator used in the experiments.

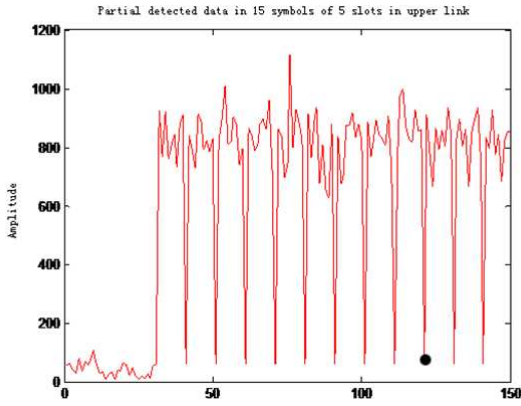


Fig. 4. Uplink data in all time slots.

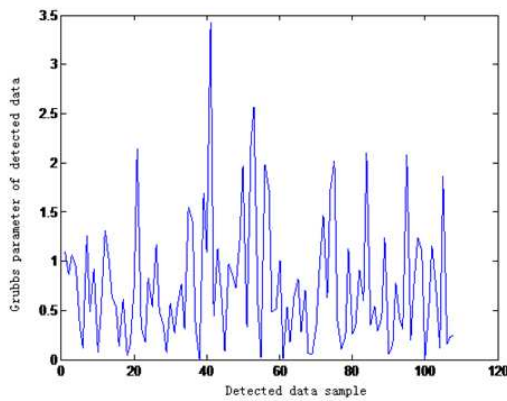


Fig. 5. The Grubbs parameter obtained at the first iteration of data pre-processing.

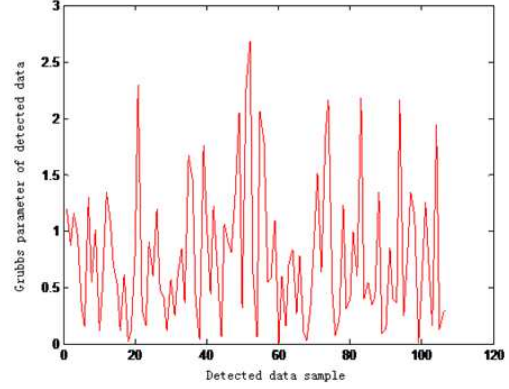


Fig. 6. The Grubbs parameter obtained at the second iteration of data pre-processing.

regarded as an outlier and will be removed. Fig. 6 shows the Grubb's parameter at the second iteration. The data after outlier removal tend to be more normally distributed, the ratio between the measured data and the residual error satisfies the Grubb's criterion. It is suggested that the error detection and removal be conducted only once in mild interference environments in order to reduce the processing load and latency. However, the procedure can be repeated several times until all the outliers are removed in the severe interference case.

We can also detect the error caused by phase offset in the base station's RF device. The procedure involves first dividing the measured data into two groups after putting them in ascending order and calculating their expectation value; then using the following equation to determine the error caused by phase offset [3]

$$t = \sum_{i=n/2+1}^n (y_i - \bar{y})^2 - \sum_{i=1}^{n/2} (y_i - \bar{y})^2. \quad (4)$$

If  $t$  approaches zero, there is no system error caused by phase offset. Otherwise, we need to correct the phase and report to the base station's MAC layer which in turn instructs the RF unit to detect and correct the phase error in the baseband signal.

Fig. 7 shows the comparison between the original data and data obtained by cubic curve fitting after outliers are removed. For accurate interpolation results, we adopt the segmented curve fitting approach, which can be expressed mathematically as  $y = \sum_{i=1}^k a_i x^i$ . The coefficients  $\{a_i\}$  are calculated based on the pilot subcarriers in the base station. The value of  $k$  is usually chosen to be less than or equal to 3 taking into account the tradeoffs between fitting accuracy and system complexity. The value of  $k$  is set to be 3 in Fig. 7. We save the curve fitting results, e.g., in a RAM

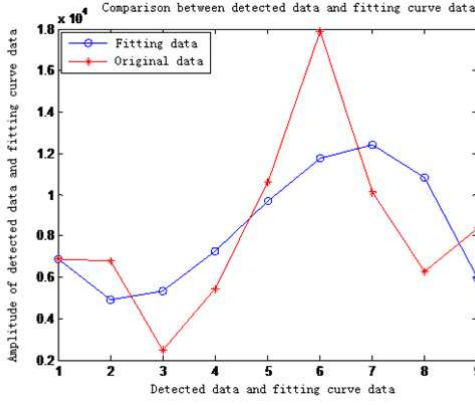


Fig. 7. Comparison of the original data with curve fitting results.

for 6 OFDM subcarriers in each tile. The saved data can be used to replace the outliers at a later stage.

The hardware complexity with DSP chips will be analyzed in the sequel. The squared root operation can be approximated by Taylor series (the precision is  $10e-04$ ) for fixed-point DSP chips, i.e.,

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{x}{2} + x^2 \cdot \frac{1 \cdot 3}{2 \cdot 4} - x^3 \cdot \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} + x^4 \cdot \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8} - x^5 \cdot \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdot 9}{2 \cdot 4 \cdot 6 \cdot 8 \cdot 10}, \quad (5)$$

which requires 24 multiplication, 3 addition and 2 subtraction operations. For the Grubbs' algorithm, if we have  $N$  measured samples (OFDM subcarriers) in the uplink data frame, then we need  $N+2$  multiplications,  $N-1$  addition,  $N$  subtraction operations to calculate the residual errors. The multiplication and addition operations in Eq. (5) can be performed in a parallel fashion on Texas Instrument (TI) DSP chips. The high efficiency of DSP chips in the multiplication and addition operations guarantees real-time data processing for a high data rate communication system.

In WiMAX and LTE systems, the uplink feedback signal, such as sounding signal is very sensitive to the effect of interference. In low signal to interference and noise ratio (SINR) environment, when the conventional average to peak ratio scheme is applied, the system error rate can be up to 10% which prevents the system from functioning properly. This problem can be tackled by the proposed data pre-processing in conjunction with the tradition detection scheme.

The Grubbs scheme gives more accurate results compared to the Wright scheme when applied to WiMAX or LTE uplink systems with moderate amount of measured data. However, its performance degrades when the amount of measured data increases. Therefore it is more suitable for fixed wireless access or low data rate environments where there are no needs

for significant measurement and data processing. The Wright scheme can perform computation for a large amount of data at time, however with a lower accuracy. It is suitable for high speed communication with a lower requirement on the detection accuracy. The choice of different schemes is application dependent.

#### IV. CONCLUSIONS

In this paper, we propose to use the Grubbs or Wright algorithm to pre-process the data in the WiMAX and LTE uplink systems. With the proposed scheme, we can robustify the 4G systems against interference and system errors, and enhance the reliability of the uplink transmission with limited measurements and computational complexity. By effectively detecting the corrupted data and replacing them with the ones obtained by curve fitting, we obtain more accurate estimate of channel state information and frequency offset, which will in turn improve the performance of data detection.

#### REFERENCES

- [1] F. Grubbs. "Procedures for detecting outlying observations in samples", *Technometrics*, vol. 11, no. 1, pp. 1-21, 1969.
- [2] R. Maronna, D. Martin, V. Yohai. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics, ISBN: 978-0-470-01092-1, June 2006.
- [3] L. Hui. "Electronic instrument and measurement", *Publisher of Chinese science and technology*, China University of Technology, 1992.
- [4] N. Petrov, O. Bustorina. "Reliability analysis of ventilation systems", *Journal of Mining Science*, vol. 22, no. 6, pp. 491-496, Nov. 2006.
- [5] G. Kitagawa, H. Akaike. "A quasi Bayesian approach to outlier detection", *Annals of the Institute of Statistical Mathematics*, vol. 34, no. 1, pp. 389-398, 1982.
- [6] "IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems," IEEE Std 802.16e-2005.
- [7] D. Baum, J. Hansen, J. Salo. "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)", *Proc. IEEE VTC'05*, pp. 3132-3136, May 2005.