Detector Processor for a 5G Base Station

Cao Niu † and Dake Liu *,†

State Key Laboratory of Marine Resource Utilization in South China Sea, School of Information and Communication Engineering, Hainan University, Haikou 570228, China
* Correspondence: liu.dake@outlook.com
† These authors contributed equally to this work.

Abstract: Nonlinear soft bit detection is essential for the uplink receivers of 5G base stations, especially for users around the cell edge. However, its throughput and computing complexity are always challenges for both research and industry. A low-cost and low-power parallel implementation of a soft-output detector based on sorted QR decomposition (SQRD) and the K-best breadth-first search algorithm is thus proposed to reduce computational complexity and latency. In addition, to save area and reduce latency, two improvement methods are used: (1) reduce the computing cost by saturating and truncating large values during PED computing and (2) reduce the sorting cost by using the binary bit sorting method for a reduced sample set with finite accuracy. Furthermore, a pipelined VLSI architecture is designed using 28-nm digital CMOS technology offered by Semiconductor Manufacturing International Corporation (SMIC). It can achieve a peak throughput of 6400 Mbps while consuming 153 K gates (including all flip-flops) for SMIC’s 28-nm technology and running at 800 MHz, a 32% cost reduction compared with the published reference design.

Keywords: MIMO; OFDM; K-best; signal detection

1. Introduction

5G technology requires great bandwidth, a fast data transmission rate and very reliable transmission performance. The detector processor implemented in the past can no longer meet the needs of signal detection in 5G micro base stations. How to reduce the hardware overhead and scale while ensuring the demodulation performance so that it can be well applied in 5G micro base stations is an important problem. The performance of nonlinear detection is better than that of linear detection in multiple-input multiple-output (MIMO) systems [1]. Its computational complexity is high at the same time. A low-cost, high-throughput and near-optimal MIMO detector has always been the key to the designing a receiver [2]. The maximum likelihood (ML) algorithm can theoretically provide the best detection performance. In [3–5], the exhaustive search method was suitable for a small modulation scale, and the computational complexity was unacceptable when the modulation scale was large. In order to reduce the computational complexity, a nonlinear detection method, spherical decoding(SD) based on tree searching, is widely used. For example, the width-first (K-best) method was used in [6–10], the depth-first method was used in [11,12], and so on. By properly controlling the number of nodes at each search level, it has near-optimal performance and scalable complexity. The K-best algorithm is a forward algorithm with a fixed throughput, so it is suitable for the implementation of a pipelined architecture. However, when the size of the antenna and signal modulation are large, in order to be close to the ML algorithm, the K value often needs to be very large, so the nodes that need to be calculated in the search will be doubled, thus increasing the amount of computation. Therefore, in the design of 5G micro base station receivers, the selection of candidate nodes needs a new method to further improve.

A few implementations using field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs) have been designed based on K-best detectors. For
example, the authors of [13] used the K-best method of an adaptive radius for hardware implementation, but this requires constant backtracking to calculate the radius and control the number of candidate nodes, which is not efficient enough. In [14], although the K-best algorithm could achieve a high throughput, it was not flexible enough, and the hardware cost was too high. In [15,16], preprocessing and parallel computing were used to improve the SD algorithm, but the complexity of the SD algorithm itself would lead to a waste of hardware resources and higher computing time cost. In order to further improve the bit error rate performance, a soft output detector [17] which can generate a logarithmic likelihood ratio (LLR) is proposed. It can be seen that the demodulation performance was improved, but it brought an unnecessary hardware overhead and needs further improvement. However, in most of the references, the hardware and computing costs of the aforementioned solutions were high and hard to accept in 5G base stations. In this paper, we focus on cost reduction while maintaining near-optimum performance.

In this paper, the soft-output K-best MIMO detector based on sorted QR decomposition (SQRD) was improved and implemented, with low complexity and near-optimum performance as well support for different kinds of modulation. Improvements include data saturation for partial Euclidian distance (PED) calculation and a bit sorting method for small-sample binary data with finite accuracy. Our design is verified by its synthesized RTL design and silicon layout. The IC layout is based on the COMS technology of SMIC (28 nm). The circuit achieves a peak throughput of 6400 Mbps at 800 MHz while consuming 153 K gates, with 55 cycles for latency. It is better than the currently published solutions and can meet the requirements of 5G micro base station uplink receivers.

This paper is organized as follows. Section 2 reviews the frequency domain model of the MIMO system and K-best decoding algorithm. Section 3 discusses the modified K-best decoding algorithm and presents the performance simulation results. Section 4 describes the hardware design and implementation results of the proposed algorithm. The conclusion is given in Section 5.

2. System Model

The system in this paper actually refers to the 5G micro base station system, for which a flow chart of the 5G baseband system is shown in Figure 1.

![Figure 1. The flow chart of the 5G baseband system.](image)

First, the data are modulated on subcarriers, and the data stream is mapped to the constellation. At the same time, the RS reference signal is generated and modulates other subcarriers, which is convenient for the receiver to estimate the channel distortion. After layer mapping, resource mapping and precoding, IFFT and other steps, the modulated subcarriers in the F domain are converted into symbols in the time domain that can be transmitted by the digital front end. Together with the noise interference and channel distortion, the transmitted symbols are received at the front. After FFT and de-mapping processing, the transmitted symbols are converted into a subcarrier data signal for subsequent processing. The channel is estimated by analyzing the RS reference signal and noise, and the channel matrix H of each data subcarrier is equalized by interpolation or Wiener filtering. Finally, the received signal is detected and decoded.
2.1. Soft-Output Detection

Consider a MIMO system with M transmitting and N receiving antennas. The complex equivalent model can be expressed as [18]

\[ y = H \cdot s + n \]  

where \( s \) is the \( M \) dimensional transmit symbol vector and each element is modulated by \( M_c \) coded bits. \( H \) is an \( N \times M \) channel matrix, and \( n \) is independent and identically distributed complex Gaussian noise, while \( y \) is the received symbol vector.

Each symbol in \( s \) is drawn from the real constellation \( W \). \( H \) can be decomposed by QR decomposition into a unitary matrix \( Q \) and an upper triangular matrix \( R \) [19]. Then, the objective of MIMO detection is to find the closest possible transmitted vector \( \hat{s} \):

\[ \hat{s} = \arg \min_{s \in \Omega^M} \| \hat{y} - Rs \|_2 \]  

where \( \hat{y} = Q^H y \), with \( Q^H y \) being the conjugate transpose of \( Q \).

The soft-output MIMO detector generates a posteriori probability for each bit by calculating the extrinsic LLR values. The LLR of bit \( x_i \) is defined as [20,21]

\[ \text{LLR}(x_i | y) = \ln \frac{P(x_i = 1 | y)}{P(x_i = 0 | y)} \approx \frac{1}{2\sigma^2} \left\{ \min_{x_i^1} \| \hat{y} - Rs \| - \min_{x_i^0} \| \hat{y} - Rs \| \right\} \]  

2.2. K-Best Algorithm

K-best is an optimization algorithm based on the breadth-first tree search algorithm. For M-QAM modulation, the K-best algorithm can be described as the following operations:

1. Extension: Store the candidate child nodes of each layer in the node collection as candidate nodes of each layer;
2. Calculation: The PED is calculated based on the root node and its candidate child nodes;
3. Sorting: When the PED values of all candidate child nodes in a layer are calculated, the sorting operation is carried out, and the smallest \( K \) nodes are retained as the parents of the next layer’s candidate nodes. If this is not the last layer, then return to step 2.

According to the flow of the K-best algorithm, we can see that the whole of path expansion is a forward process. Therefore, algorithm-level parallelism can be realized. When the K-best detector needs to calculate the soft output, the reserved paths can be easily used to generate an ordered candidate list.

3. Algorithm Implementations

3.1. Algorithm Description

In this paper, the soft-output K-best MIMO detector based on SQRD is proposed which combines the advantages of the K-best algorithm and spherical decoding algorithm and can achieve a performance close to the ML algorithm, even after significantly reducing the number of search nodes.

In K-best detection, the channel matrix \( H \) should be decomposed by QR first. The \( i \)-th column of the channel matrix \( H \) represents the transmission function from the \( i \)-th transmitting antenna to multiple receiving antennas, so the column norm of the channel matrix is the strength of the transmitted signal [22,23]. Before the QR decomposition of the channel matrix \( H \), the column norm of \( H \) is calculated, and the channels are sorted from weak to strong according to the column norm order from small to large. The QRD of the
sorted H ensures that the order of signal detection is carried out according to the SNR from high to low, namely in SQRD detection.

When detecting the signals layer by layer, to maintain performance with a small K, and to minimize the candidate nodes computing in each layer of the K-best calculation as much as possible, it is necessary to limit the number of candidate nodes in each layer of the K-best algorithm [24]. In the algorithm designed in this paper, the nearest 16 nodes within a square, as the received signal, are selected as the expansion nodes by trading off performance and cost. At the same time, to improve the expansion node accuracy, for the received signal, the MMSE channel equalization is preprocessed to solve the initial estimate $S_{MMSE}$, and the nearest 16 nodes are selected as the expansion nodes of each layer to participate in the operation. The MMSE algorithm is used to limit the region of the candidate nodes, which greatly reduces the number of nodes needed in demodulation. The volume results show that the reduction in candidate nodes does not affect the performance in the case of a low SNR. Therefore, the search space reduction in this paper gives a negligible performance penalty.

Within the range of candidate nodes limited by the SD and MMSE, using the K-best algorithm can greatly reduce the number of nodes to be calculated, reduce the computing time, and ensure the BER performance in 64-QAM and 256-QAM high-order constellation demodulation. The flow chart of the system algorithm designed in this paper is shown in Figure 2.

![Flow chart of the proposed algorithm in this paper.](image)

There are M transmitting antennas and N receiving antennas. The transmitter sends the signal $s_{M \times l}$, the receiver receives the signal $y_{N \times l}$, and the channel estimation matrix is $H_{N \times M}$. For the received signal $y_{N \times l}$ and channel estimation matrix $H_{N \times M}$, the following steps are employed:

**Step 1.** Deal with the H matrix: The column norm of the H matrix is solved and rearranged from small to large according to the norm value, and the $H'$ and permutation matrix $P$ are obtained.

**Step 2.** QR decomposition: The Q matrix and R matrix are obtained by QR decomposition of $H'$.
Step 3. Calculate the MMSE channel equalization matrix [25] as follows:

\[ G_{\text{MMSE}}(N \times M) = \left( H^H H + \sigma^2 I \right)^{-1} H^H \]  

(4)

Step 4. Deal with the received signal \( y \):

\[ Y_{(M \times 1)} = Q_{(M \times N)}^{-1} Y_{(N \times 1)} \]  

(5)

Step 5. Calculate the initial estimated signal:

\[ S_{\text{MMSE}}(N \times 1) = G_{\text{MMSE}}(N \times M) Y_{(M \times 1)} \]  

(6)

Step 6. Extended candidate node: With \( S_{\text{MMSE}}(N \times 1) \) as the center, a maximum of 16 constellation points is extended as candidate child nodes. The N layer presets up to 16 child nodes per layer.

Step 7. Enter the K-best algorithm: The PED calculation formula for each layer is

\[ \Delta_i = \left| y_i - \sum_{j=i+1}^{N} R_{ij}S_j - R_{ii}S_i \right| + \Delta_{i+1} \]  

(7)

Step 8. Hard output: After step 7, the path value of each layer corresponding to the shortest PED value is directly output, which is the hard output.

Step 9. Soft output: Calculate the LLR value of each bit and output, where the formula is as follows:

\[ \text{LLR}(x_i | y) \approx \frac{1}{2\sigma^2} \left\{ \min_{X^1_i} \| \hat{y} - R_s \|^2 - \min_{X^0_i} \| \hat{y} - R_s \|^2 \right\} \]  

(8)

3.2. Algorithm Simulation Results

A MIMO detection system suitable for a 5G micro base station was built on the Matlab platform, and the antenna size was set to four to transmit and four to receive. To generate source data on the transmitter side, the binary code stream was randomly generated as the signal, the ZC sequence was used as the reference signal, and the reference signal subcarriers were modulated as BPSK. To build a channel, the channel environment was the Rayleigh channel with steady fading, and the corresponding multipath effect parameters were set to simulate the addition of noise interference in the channel. At the receiving side, the channel was estimated based on the pilot subcarriers, and interpolation and channel equalization were performed between subcarriers at the non-pilot subcarriers in the frequency domain.

The related verification of different detection algorithms was carried out in the built system. For the verification of the improved algorithm designed in this paper, several test functions were written, which were the traditional K-best algorithm function, the QRD preprocessing function of the H matrix, the SQRD preprocessing function of the H matrix, the K-best algorithm function of a simplified search space, the hard output function, the soft LLR calculation function and so on.

The above functions were tested, and the relevant bit error rate was calculated to verify the correctness of the behavior model. Compared with the ML algorithm, the performance reduction of the algorithm we used was between 0.5 and about 1 dB, which could ensure the effectiveness of the algorithm. After finalizing the nonlinear soft bit K-best detector with variant configurations, we thus defined that the verified detector in MATLAB was the golden model of our design. Figure 3 proves and compares the BER performance of 16-QAM and 64-QAM demodulation with different K values, and Figure 4 shows the BER performance comparison of QPSK and 16-QAM demodulation with the same K values in the case of QRD and SQRD.
Looking at Figures 3 and 4, we can see that when the SNR was between 15 and 35 dB, the BER of the larger K value was 1–2 dB higher than that of the low K value. When the SNR of 16QAM demodulation was higher than 20 dB, the influence of different K values on the BER became smaller. Therefore, in the proposed algorithm, the BER performance will gradually improve with the increase in the K value in the case of a low SNR, while in the case of a high SNR, the influence of different K values will be lessened, and the BER performance will be relatively stable.

Figure 3. BER of 16-QAM and 64-QAM demodulation with different K values.

Looking at Figures 5 and 6, we can see that under the condition of the same K value, the QRD algorithm and the SQRD algorithm based on the H-column permutation of the channel matrix were simulated and compared. Because the SQRD algorithm minimized the impact of error propagation on subsequent detection, the overall performance of the algorithm was about 1–1.5 dB of gain compared with the traditional QR decomposition algorithm. The results show that the correctness and performances are sufficient for further hardware implementation.

Figure 4. BER of 256-QAM demodulation with different K values.
Further BER comparison between the soft and hard bits showed that the soft bit gain of high-order modulation with good channel conditions was 2–3 dB, so both the soft output and hard output were realized in this paper.

The authors of [17] used the K-best algorithm based on SQRD, which is similar to that in this paper, under the same SNR, and the overall performance of the algorithm was about 1.5–2 dB of gain compared with that in [17]. In addition, compared with other published reference algorithm designs, the performance in this paper was very similar to that of the ML detection algorithm and met the requirements of the system.

4. Hardware Architecture and Implementation

4.1. Algorithm Implementation

The architecture simulator designed in this paper ensured that the low-cost architecture design was consistent with golden model compliance after optimization. The modules in this paper were the sample selector, PED calculator, sorter and LLR calculator. In the selector, the candidate node was selected by configuring the modulation mode and the
addressing function. In the PED calculator and sorter, the output was adjusted according to the amount of data to be processed. The LLR calculator was only used to calculate the soft bit information value.

4.1.1. Sample Selector

The function of the module is divided into two parts: first, the data region of the candidate node is divided according to the modulation mode M and the received signal y, and then the candidate nodes are chosen in the selected area in the full constellation table S. When the modulation mode was QPSK or 16QAM, all the constellation points in the full constellation table S were directly selected without going through the data region division, and when the modulation mode was 64QAM or 256QAM, the data region division was needed first, and then point selection was carried out. It should be noted that in this module, the region division of the candidate nodes is based on the estimated received signal \( H' \) obtained by the preprocessing of MMSE equalization. Table 1 shows the number of nodes that need to be searched by the ML, K-best and proposed algorithm under different modulation modes (with a subcarrier in the case of demodulation with a receiving antenna of four as an example). It can be seen that after the square region division of the candidate nodes, the number of nodes that 64QAM and 256QAM needed to search and calculate for the K-best algorithm was greatly reduced, which saved computing time and hardware overhead.

| Mode    | ML                | K-Best               | K-Best in This Work |
|---------|-------------------|----------------------|---------------------|
| QPSK    | \( 4 + 4^2 + 4^3 + 4^4 \) | \( 4 \times (3n + 1) \) | \( 4 \times (3n + 1) \) |
| 16QAM   | \( 16 + 16^2 + 16^3 + 16^4 \) | \( 16 \times (3n + 1) \) | \( 16 \times (3n + 1) \) |
| 64QAM   | \( 64 + 64^2 + 64^3 + 64^4 \) | \( 64 \times (3n + 1) \) | \( 16 \times (3n + 1) \) |
| 256QAM  | \( 256 + 256^2 + 256^3 + 256^4 \) | \( 256 \times (3n + 1) \) | \( 16 \times (3n + 1) \) |

The flow chart of the module is shown in Figures 7 and 8, and the pseudocode of the algorithm of the sample selector is as follows (Algorithm 1).

![Figure 7. MMSE equalization preprocessing process before selecting points in sample selector.](image-url)
Figure 8. Flowchart of candidate node selection (taking layer 4 as an example).

Algorithm 1 Sample selector

Input: M: Modulation scale; s: Initial estimate s
Output: Candidate node: Set of each layer candidate node
1. if $M = 4$ (QPSK) or 16 (16QAM) then
2. Base address = 0;
3. Select all constellation points as Candidate node;
4. else
5. Calculate Base address based on s value;
6. Add the corresponding full constellation S table according to M;
7. Find the candidate node value of the corresponding sequence number in the S table and store it in candidate node;
8. end if
9. return candidate node

4.1.2. PED Calculator

The PED value of each layer was calculated according to Equation (7). In order to achieve the accuracy sorting of the simplest data in the minimum time, after the PED value of each candidate node was calculated, it was necessary to carry out large value saturation processing. We needed to find the smallest K PED values (the K value is usually small), and there were many candidate PED values for each layer, so we directly set the upper limit of the PED value, quantified the value within the upper limit (8-bit) and then discarded the larger PED values. This can not only reduce the number of samples but also reduce the overhead of sorting. The number of samples after the large value saturation treatment was 10% lower, and the performance was basically unchanged.

4.1.3. Sorter

In the K-best detection algorithm, a very important step is to sort the calculated PED values. In the proposed algorithm in this paper, the binary data with limited precision and
small samples need to be sorted. Most of the traditional sorting methods are bubble sorting and comparison sorting, suffering from large resource overheads and large computing time costs. Therefore, this paper designed a small sample limited precision binary data bit sorting method to complete the sorting function.

To minimize the sorting time, the 8-bit sorting method in this paper used the method of parallel filtering to process the data in clusters at the same time, saving the needed data from small to large. This method only processed the data that met the conditions and discarded the data that did not meet the conditions by constantly updating the conditions, which could reduce the number of times the data was processed. Shown in Figure 9 is a sorter flow chart of a plurality of 8-bit data.

Figure 9. Flowchart of multiple 8-bit data sorting.

Table 2 compares the common bubble sorting method with the bit sorting method presented in this paper. It can be seen that although the bubble sorting method and the bit sorting method in this paper had considerable stability, the average time required by the bubble sorting method increased exponentially with the increase in the number of samples, while the average time of the bit sorting method used in this paper was stable, and the latter had obvious advantages when the number of samples was large. In addition, in the hardware implementation, the bubble sorting method used subtractors many times when comparing adjacent data, and the bit sorting rule of this paper used multiple AND gates in parallel instead of subtractors to complete the comparison function, which not only reduced the time delay but also further reduced much of the hardware overhead.
Table 2. Comparison of bubble sort and bit sort.

| Mode                  | Average Time Complexity | Space Complexity | Stability |
|-----------------------|-------------------------|------------------|-----------|
| Bubble sort           | $o(n^2)$                | $o(1)$           | Stable    |
| Bit sort in this paper| $o(1)$                  | $o(n)$           | Stable    |

The pseudocode of the algorithm of the sorter module is as follows (Algorithm 2).

**Algorithm 2 Sorter**

**Input:** PED: PED sample values that need to be sorted.
Path: Candidate path corresponding to the sample.
K: The number of samples to be retained after sorting.

**Output:** remPED: Reserved PED value.
remPath: Corresponding candidate path.

1. Function First sorting(PED, Path)
2. PED is divided into $2^n$ heaps according to the characteristics of pre-$\frac{n}{2}$ bits, and the number of each heap $num1(i)$ is calculated.
3. for $i = 1$ to $2^n$ do
4. if $num1(i) > K$ then
5. The $num1$ data (as PED2) and the paths (as Path2) are sorted for the second time;
6. else if $num1(i) < K$ then
7. The $num1$ data and the paths as output, $i = i + 1$;
8. else
9. The $num1$ data and the paths as output;
10. end if
11. NEXT i
12. end for
13. Function second sorting(PED2, Path2)
14. PED2 is divided into $2^n$ heaps according to the characteristics of rest-$\frac{n}{2}$ bits, and the number of each heap $num2(i)$ is calculated.
15. for $i = 1$ to $2^n$ do
16. if $num2(i) < K$ then
17. The $num2$ data and the paths as output, $i = i + 1$;
18. else
19. The $num2$ data and the paths as output;
20. end if
21. NEXT i
22. end for
23. return

4.2. Hardware Architecture Design

The hardware architecture includes a control path and data path, which are integrated into the same processor module [26]. Starting from the data path, the data from the register are sent to the computing cell array through the distribution logic, the output data of the computing cell array are stored in the memory or buffer, and the reconfigurable parameter memory sends the reconfigurable parameters into the computing cell array. Starting with the control path, the control module inputs the activation signal into the calculation function unit, activates the corresponding calculation module and outputs the activation signal of the next module to the control module, which is transmitted by the control module and activates the next calculation module. The proposed hardware top-level architecture of detector processors is shown in Figure 10.
According to the different functional steps in the algorithm, the calculation function module is divided into several smaller modules. The pipeline structure is implemented by a multi-stage processing unit, and each level performs the function of a module [26]. After calculation, the data as input are forwarded to the next level’s unit, and the activation signal of the next module is transmitted to the control module at the same time. The application of this pipeline can improve data throughput and hardware utilization. The K-best detection algorithm has no backtracking operation and is suitable for pipelining.

4.3. Hardware Implementation Results

In the hardware system verification, we performed the following two actions of work. On the one hand, we used the C model to simulate each sub-module at the hardware level and then write the RTL code, compile it in QuartusII, check the correctness of the RTL and finally write the test files, add incentives and test the standard output results so as to verify the correctness of the function. On the other hand, under the condition of ensuring the performance of the algorithm, we used the SMIC 28-nm digital CMOS technology to carry out the hardware comprehensive compilation and layout. After the hardware synthesis and layout, we found the area cost, circuit speed and power consumption report, which were used for comparison.

4.3.1. IC Hardware Layout

The IC hardware layout is shown in Figure 11.
4.3.2. Hardware Overhead and Comparison

The previously presented K-best MIMO detector was fabricated in SMIC 28-nm technology. The area of the chip was 159.13 $\mu$m $\times$ 95.48 $\mu$m. At the normal 1.2-V core power supply, the detector could work at an 800-MHz clock rate to support a 6400-Mb/s throughput. Table 3 shows the comparison of this paper’s results with other published K-best detectors. The detection throughput in Table 3 is given by [26]

$$\text{Throughput} = f_c \times \frac{\log_2 M \times N_t}{T}$$ (9)

where $M$ is the real constellation size, $N_t$ is the antenna size, $f_c$ is the clock frequency and $T$ is the number of clocks required for each set of data output.

| Reference | [12] | [15] | [27] | [28] | [29] | This Work |
|-----------|------|------|------|------|------|-----------|
| Modulation Mode | 64QAM | 64QAM | 64QAM | 64QAM | 64QAM | BPSK~256QAM |
| Technology  | 65 nm | 65 nm | 65 nm | 65 nm | 40 nm | SMIC 28 nm |
| Antennas   | $4 \times 4$ | $16 \times 16$ | $16 \times 16$ | $4 \times 4$ | $8 \times 8$ | $4 \times 4$ |
| Pre-progress | SQR  | SQR  | SQR  | SQR  | QR   | SQR      |
| Performance | Near-ML | Near-ML | Near-ML | Near-ML | Near-ML | Near-ML |
| Max Freq. (MHz) | 450 | 588 | 625 | 500 | 641 | 800 |
| Throughput (Mbps) | 300 | 3528 | 3910 | 368 | 3846 | 6400 |
| Area       | 184 K | 3720 K | 5891 K | 1055 K | 3873 K | 153 K |
| Power (mW) | 138.2 | 1831 | 548 | 307 | 1555 | 136 |
| NAE        | 1.63 | 15.9 | 10.56 | 0.34 | 2.44 | 18.01 |

Normalized area efficiency (NAE) [13,26]: $\text{NAE} = \frac{\text{Throughput}}{N_t^2} \times \frac{S}{S^2}$. All designs are scaled to 65 nm, where $S = 65$ nm. All designs are scaled to $4 \times 4 : \frac{N_t^2}{S^2} = \frac{S^2}{S^2}$.

Specifically, the normalized silicon overhead was used as a parameter to compare the capabilities of different detector systems, and this parameter can eliminate the differences between different designs as a unified parameter of measurement. The specific calculation formula is given in Table 3.

Through comparison, we can see that the K-best soft-output signal detector based on SQRD designed in this paper is superior to other designs in the following aspects:

1. In 64QAM and higher-order modulation, the target algorithm reduces the number of search nodes to 16, which significantly reduces the computing overhead, area and power consumption. At the same time, the MMSE algorithm is used to estimate the initial nodes, which ensures a certain demodulation efficiency while reducing the search space.

2. The pipeline design of the detector and the prediction of candidate nodes greatly improve the demodulation efficiency. Therefore, the throughput of decoding can be significantly improved.
(3) In the hardware implementation, the design of the detector adopts parallel processing and a pipeline structure at the RTL level so as to achieve the purpose of reducing power consumption. At the same time, the design of the bit sorting method for binary data with limited precision for small samples reduces the use of comparators and reduces the cost of the logic hardware.

In fact, base station chip industries need better designs to achieve both low costs and high performance. However, in the existing designs, some pay more attention to the improvement of performance, which leads to excessive hardware overhead. Some prefer a low hardware overhead, inducing low data throughput. Through the comparison at the end of the article, we can see that the design in this paper has great advantages when using the normalized silicon cost as the comparison standard, which also means that we found better results, which can be used for all base station IC companies. In addition, the design of the detector can be used in the modulation systems of BPSK, QPSK, 16QAM, 64QAM and 256QAM. The detector can choose the demodulation mode and output mode (hard or soft bits) according to the set parameters.

5. Conclusions

In this paper, an SQRD-based K-best MIMO detector for 5G micro base stations is proposed which can be configured to complete the detection of high-order modulation (256QAM and 64QAM) and be backward compatible with low-order modulation (16QAM, QPSK, and BPSK) signal detection. At the algorithm level, the MMSE preprocessing method is used to set a $4 \times 4$ square range limit for the candidate nodes to achieve lower computational complexity, while the performance loss is negligible. At the hardware level, two improvement methods are used: reducing the computing cost by saturating and truncating large values during PED computing and reducing the sorting cost by using the binary bit sorting method for a reduced sample set with finite accuracy. The pipelined structure we proposed can output one subcarrier’s soft information in each clock cycle. In the case of 256QAM modulation, a maximum of 8 soft bits of information can be achieved in a cycle (here, the resolution of the soft bits was 8-bit), which further improves the detection efficiency. Compared with the existing detectors, the area is reduced to 153 K, and the power consumption is only 136 mW while offering a 6400-Mbps throughput when the clock cycle is 800 MHz. This result also shows that under the premise of ensuring near-ideal performance, this design has great advantages over those in the recent literature, which verifies the feasibility of the proposed algorithm and structure.

Author Contributions: Data curation, C.N.; Formal analysis, C.N.; Funding acquisition, D.L.; Supervision, D.L.; Writing—original draft, C.N.; Writing—review & editing, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by maritime communication technology in the South China Sea of ZDKJ2021022, a major project of Hainan Province, grant number ZDKJ2021022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Hochwald, B.M.; Brink, S.T. Achieving near-capacity on a multiple antenna channel. IEEE Trans. Commun. 2003, 51, 389–399. [CrossRef]
2. Rusek, F.; Persson, D.; Lau, B.K.; Larsson, E.G.; Marzetta, T.L.; Edfors, O.; Tufvesson, F. Scaling up MIMO: Opportunities and challenges with very large arrays. IEEE Signal Process. Mag. 2013, 30, 40–60. [CrossRef]
3. Van Etten, W. Maximum likelihood receiver for multiple channel transmission systems. IEEE Trans. Commun. 1976, 24, 276–283. [CrossRef]
1. Kisialiou, M.; Luo, X.; Luo, Z.Q. Efficient implementation of quasi maximum-likelihood detection based on semidefinite relaxation. *IEEE Trans. Signal Process.* 2009, 57, 4811–4822. [CrossRef]

2. Liu, Y.-X.; Dai, Y.-X.; Ueng, Y.-L. Post-Processing of K-best MIMO Detection for High-Order Modulations. In Proceedings of the 2021 IEEE Asia Pacific Conference on Circuit and Systems (APCCAS), Penang, Malaysia, 22–26 November 2021; p. 1.

3. Ivanov, A.; Osinsky, A.; Lakontsev, D.; Yarotsky, D. High Performance Interference Suppression in Multi-User Massive MIMO Detector. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; Volume 52, pp. 1–5.

4. Mahdavi, M.; Shabany, M. Novel MIMO Detection Algorithm for High-Order Constellations in the Complex Domain. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 2013, 21, 834–847. [CrossRef]

5. Neshatpour, K.; Shabany, M.; Gulak, G. A High-Throughput VLSI Architecture for Hard and Soft SC-FDMA MIMO Detectors. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2015, 62, 761–770. [CrossRef]

6. Hu, X.; Liu, Z.; Ling, Y.; Yu, Z. Improved Sphere Decoding Detection Algorithm Based on Lattice-Reduction-Aided. In Proceedings of the 2019 28th Wireless and Optical Communications Conference (WOCC), Beijing, China, 9–10 May 2019; pp. 1–6.

7. Studer, C.; Bölcskei, H.; Yin, S.; Wei, S. Soft–Input Soft–Output Single Tree-Search Sphere Decoding. *IEEE Trans. Inf. Theory* 2010, 56, 4827–4842. [CrossRef]

8. Liu, D. *Efficient MIMO Preprocessor with Sorting-Relaxed QR Decomposition and Modified Greedy LLL Algorithm.* *IEEE Trans. Circuits Syst. I Regul. Pap.* 2013, 60, 2531–2542. [CrossRef]

9. Wang, J.; Liu, X.; Li, Z.; Si, J. Soft-output MIMO detector for MMSE receiver with channel estimation error. In *Proceedings of the 2011 IEEE 22nd International Conference on Microelectronics (ICM),* Singapore, 9–11 November 2011; pp. 29–32.

10. Ren, Y.; He, G.; Ma, J. High-throughput sorted MMSE QR decomposition for MIMO detection. In *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS),* Seoul, Korea, 23–26 May 2012; pp. 2845–2848. [CrossRef]

11. Liu, L.; Ye, F.; Ma, X.; Zhang, T.; Ren, J. A 1.1-Gb/s 115-pJ/bit Configurable MIMO Detector Using 0.13-µm CMOS Technology. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2010, 57, 2476–2486. [CrossRef]

12. Shen, C.-A.; Eltawil, A.M. A Radius Adaptive K-Best Decoder With Early Termination: Algorithm and VLSI Architecture. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2008, 55, 290–301. [CrossRef]

13. Liu, L.; Ye, F.; Ma, X.; Zhang, T.; Ren, J. A 1.1-Gb/s 115-pJ/bit Configurable MIMO Detector Using 0.13-µm CMOS Technology. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2010, 57, 2476–2486. [CrossRef]

14. Peng, G.; Liu, L.; Zhou, S.; Xue, Y.; Yin, S.; Wei, S. Algorithm and Architecture of a Low-Complexity and High-Parallelism Preprocessing-Based K -Best Detector for Large-Scale MIMO Systems. *IEEE Trans. Signal Process.* 2018, 66, 1860–1875. [CrossRef]

15. Ju, C.; Ma, J.; Tian, C.; He, G. VLSI implementation of an 855 Mbps high performance soft-output K-Best MIMO detector. In *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS),* Seoul, Korea, 20–23 May 2010; pp. 2849–2852.

16. Castaneda, O.; Goldstein, T.; Studer, C. Data detection in large multi-antenna wireless systems via approximate semidefinite relaxation. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2016, 63, 2334–2346. [CrossRef]

17. Huang, Z.Y.; Tsai, P.Y. Efficient implementation of QR decompo-sition for gigabit MIMO-OFDM systems. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2011, 58, 2531–2542. [CrossRef]

18. Wang, Z.; Tsai, P.Y. Optimized configurable architectures for scalable soft-input soft-output MIMO detectors with 256-QAM. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2015, 63, 4969–4984. [CrossRef]

19. Sarieddeen, H.; Mansour, M.M.; Chehab, A. Efficient near-optimal 8 × 8 MIMO detector. In *Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Doha, Qatar,* 3–6 April 2016; pp. 1–6.

20. Ren, Y.; He, G.; Ma, J. High-throughput sorted MMSE QR decomposition for MIMO detection. In *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS),* Seoul, Korea, 23–25 May 2012; pp. 2845–2848. [CrossRef]

21. Jin, Y.; Zheng, Z.; Wang, X.; Zhang, Y. Discussion of 3GPP LTE MIMO Signal Detecting Algorithms Based on QR Decomposition. In *Proceedings of the 2009 IITA International Conference on Control, Automation and Systems Engineering, Zhangjiajie, China,* 25–28 May 2009; Volume 1, pp. 191–194. [CrossRef]

22. Hou, W.; Fujino, T.; Kojima, T. Adaptive candidate selection scheme in QRM-MLD algorithm for MIMO detection. In *Proceedings of the 2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Barcelona, Spain,* 8–10 October 2012; pp. 502–506. [CrossRef]

23. Wang, J.; Liu, X.; Li, Z.; Si, J. Soft-output MIMO detector for MMSE receiver with channel estimation error. In *Proceedings of the 2016 IEEE/CIC International Conference on Communications in China (ICCC Workshops),* Chengdu, China, 27–29 July 2016; pp. 1–5. [CrossRef]

24. Liu, D. *Embedded DSP Processor Design: Application Specific Instruction Set Processors,* 1st ed.; Morgan Kaufmann: Burlingon, MA, USA, 2008.

25. Chen, L.; Xing, Z.; Li, Y.; Qu, S. Efficient MIMO preprocessor with sorting-relaxed QR decomposition and modified greedy LLL algorithm. *IEEE Access* 2020, 8, 54085–54099. [CrossRef]

26. Zhang, C.; Liu, L.; Markovic, D.; Owall, V. A heterogeneous reconfigurable cell array for MIMO signal processing. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2015, 62, 733–742. [CrossRef]

27. Liu, C.; Cui, C.; Chen, H.-B.; He, W.; Sheng, W.; Jing, N.; Mao, Z.; He, G. A 3.85-Gb/s 8 × 8 Soft-Output MIMO Detector With Lattice-Reduction-Aided Channel Preprocessing. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* 2021, 29, 307–320. [CrossRef]