VLSI Implementation of a WiMAX/LTE Compliant Low-Complexity High-Throughput Soft-Output K-Best MIMO Detector

Dimpesh Patel[1], Vadim Smolyakov[1], Mahdi Shabany[1,2], P. Glenn Gulak[1]

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto[1]

Department of Electrical Engineering, Sharif University of Technology[2]

Abstract— **This paper presents a VLSI architecture of a novel soft-output K-Best MIMO detector. The proposed detector attains low computational complexity using three improvement ideas: relevant discarded paths selection, last stage on-demand expansion, and relaxed LLR computation. A deeply pipelined architecture for a soft-output MIMO detector is implemented for a 4x4 64-QAM MIMO system realizing a peak throughput of 655Mbps, while consuming 174K gates and 195mW in 0.13um CMOS. Synthesis results in 65nm CMOS show the potential to support a sustained throughput up to 2Gbps achieving the data rates envisioned by emerging IEEE 802.16m and LTE-Advanced wireless standards.**

I. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) systems have recently gained increasing attention in the development of competing wireless standards such as IEEE 802.16m and LTE-Advanced due to their high spectral efficiency. However, an efficient, high-throughput VLSI implementation of near-optimum 4×4 MIMO detectors has been a major design challenge for high-order Quadrature Amplitude Modulation (QAM) schemes. Near-optimum MIMO detectors offer scalable complexity, while providing comparable performance to the Maximum Likelihood (ML) detector. Depending on how MIMO detectors carry out the non-exhaustive search, they generally fall into two main categories: depth-first and breadth-first search. Among the breadth-first search detectors, the K-Best algorithm [1] guarantees an SNR-independent fixed throughput with performance close to ML. Furthermore, its feed-forward detection approach makes it particularly attractive for pipelined VLSI implementation.

Soft-output K-Best detection offers an added advantage of computing the Log-Likelihood Ratio (LLR) output required for iterative Forward Error Correction (FEC), employed by Convolutional Turbo Codes (CTC) and Low-Density Parity-Check (LDPC) codes. A Spatial Multiplexing (SM) Soft-output MIMO detector based on the K-Best algorithm has been realized for 4×4 16-QAM MIMO systems [2], which provides a peak throughput of 107Mbps and requires 97K gates core area in $0.13 \mu m$ CMOS. However, to meet the requirements of 802.16m and LTE-Advanced wireless standards, higher order constellations (64-QAM), greater number of transmit and receive antennas (4×4) and higher data rates (up to 1 Gbps) are desired. State-of-the-art 4×4 64-QAM MIMO detectors consume large silicon area ([3], [4]), achieve moderate SNR-dependent throughput (115Mbps) and show a performance loss compared to the ideal K-Best detection. One of the major design issues faced by Soft-output detection is the large number of computations required to generate the LLR output for each transmitted bit, leading to either low-throughput or large area and power requirements.

The proposed Soft-output K-Best MIMO detection scheme greatly reduces the computational complexity by implementing three innovative ideas to extend and process only relevant paths and use approximations to actual computations where appropriate. The proposed Soft K-Best detector uses a deeply pipelined and highly parallel architecture that computes LLR values for a block of 24 transmitted bits every $K = 10$ clock cycles. For 4×4 64-QAM MIMO systems, the proposed Soft K-Best detector reduces computational complexity by a factor of 5 compared to [2] and offers a 1.7dB improvement at BER = 10^{-3} compared to the conventional Soft K-Best detection scheme. In 0.13 μm CMOS, the proposed detector achieves a throughput of 655Mbps, while consuming 174K gates and 195mW. Synthesis results in 65nm CMOS show that the proposed detector reaches a throughput of 2000Mbps, while operating at 833MHz.

II. SYSTEM MODEL

Consider a MIMO system with N_T transmit and N_R receive antennas and an equivalent baseband model of the Rayleigh fading channel described by a complex-valued $N_R \times N_T$ channel matrix $\tilde{H}$. The complex baseband equivalent model can be expressed as $\tilde{Y} = \tilde{H}\tilde{S} + \tilde{V}$, where $\tilde{S} = [\tilde{s}_1, \tilde{s}_2, \ldots, s_{\tilde{N}_T}]^T$ is the N_T-dimensional complex transmitted signal vector, $\tilde{Y} = [\tilde{y}_1, \tilde{y}_2, \ldots, y_{\tilde{N}_R}]^T$ is the received symbol vector, and $\tilde{V} = [\tilde{v}_1, \tilde{v}_2, \ldots, v_{\tilde{N}_R}]^T$ is an independent identically distributed Gaussian noise vector. We assume that the channel $\tilde{H}$ is quasi-static and is updated every four channel uses. A complex $N_R \times N_T$ MIMO system can be modeled as an equivalent $2N_R \times 2N_T$ real system using Real Value Decomposition (RVD): Y = HS + V, where the dimensions of S, Y and H are $2N_T \times 1$, $2N_R \times 1$ and $2N_R \times 2N_T$, respectively. Each symbol in S is drawn from a symmetric M-QAM constellation $\Omega = \{-\sqrt{M} + 1, \ldots, -1, 1, \ldots, +\sqrt{M} - 1\}$. Using QR-decomposition, the channel matrix H can be decomposed into a unitary matrix Q and an upper-triangular matrix R. Performing a nulling operation on the received signal by Q^H results in the updated system equation:

$$Z = Q^H Y = RS + Q^H V \qquad (1)$$

The objective of the MIMO detection system is to find the closest transmitted vector $\tilde{S}$ based on the observation Y, such that the Euclidean distance $\|Z - RS\|^2$ is minimized. Exploiting the triangular nature of R:

$$\tilde{S} = \arg \min_{s_j \in \Omega^{2N_T}} \sum_{i=1}^{2N_T} |z_i - \sum_{j=i}^{2N_T} R_{ij} s_j|^2 \qquad (2)$$

which can be thought of as a detection problem in a tree with $2N_T$ levels. The accumulated sum in (2) is referred to as Partial Euclidean Distance (PED). A soft-output detector generates a posteriori probability of the received bit being a zero or a one, expressed as a Log-Likelihood Ratio (LLR). The LLR of bit x_k is defined as:

$$\text{LLR}(x_k|y) = \ln \frac{P[x_k = 1|y]}{P[x_k = 0|y]}$$

$$\approx \min_{X_i^1} \left\{ \frac{\|y - Hs\|^2}{\sigma^2} \right\} - \min_{X_i^0} \left\{ \frac{\|y - Hs\|^2}{\sigma^2} \right\} \qquad (3)$$

$$= \text{MinPED}_i^1 - \text{MinPED}_i^0 \qquad (4)$$

where (3) is derived based on standard simplifications [5], X_i^1 and X_i^0 represent all vectors X with bit X_i being 1 and 0, respectively; the minimum PED for the ith bit in X being 1 or 0 is denoted by MinPED_i^1 and MinPED_i^0, respectively.

III. Soft K-Best Detection Scheme

The MIMO detection problem in (2) can be thought of as a detection problem in a tree with $2N_T$ levels and $\sqrt{M}$ children per parent. The K-Best algorithm starts from the last level ($2N_T$) and runs all the way up to the first level, due the upper triangular structure of matrix R. Consider an intermediate level l of the tree and assume that a set of K-Best paths at the previous level $l+1$ (denoted K_{l+1}) is known. Each K-Best path at level $l+1$ has $\sqrt{M}$ possible children, which presents $K\sqrt{M}$ candidates to the next level l of the tree. The presented Soft-output K-Best detector builds on the Hard K-Best implementation [1] that uses a novel On-Demand Expansion scheme to select the K best paths at level l (K_l) by visiting only $2K-1$ lowest PED candidates out of the total $K\sqrt{M}$ possible nodes.

The Conventional Soft K-Best Detection Scheme in [3] computes the output LLR values by exhaustively extending the K lowest PED paths at level $2N_T$-1 to all possible $K\sqrt{M}$ paths at level $2N_T$. However, for nominal values of K, this scheme offers a very small performance gain for Soft K-Best detection in comparison to Hard K-Best detection.

The Modified K-Best Schnorr-Euchner (MKSE) scheme in [2] improves BER performance by utilizing information contained in the discarded paths at intermediate tree levels to compute the LLR values. The MKSE scheme retains ($K\sqrt{M}$ - K) discarded paths at selected tree levels, augments them to full length using ZF augmentation [2] (a simple estimation method that rounds the received symbol to the nearest constellation point) and computes the LLR output based on updated PED values resulting from ZF-augmentation. The increase in the number of vectors used for LLR computation causes a significant improvement in LLR quality, leading to a considerable gain in BER performance. However, this improvement comes at the expense of larger computational complexity. For example, for a 4×4 64-QAM MIMO system with $K = 10$ and $L = 6$, the MKSE scheme has to process a total of 420 intermediate paths.

The Proposed Soft K-Best Detection Scheme implements three improvement ideas that significantly reduce computational complexity without a major sacrifice in BER performance. Based on the MSKE concept of utilizing discarded paths from intermediate tree levels, the proposed Soft K-Best detection scheme achieves better BER performance in comparison to the Conventional Soft K-Best detection. Table I presents the proposed Soft K-Best detection algorithm for a $2N_R \times 2N_T$ real-valued MIMO system.

Improvement 1: Relevant Discarded Paths Selection. Computational complexity is greatly reduced by detecting and eliminating irrelevant discarded paths, as follows: First, a Bit Occurrences Table (which keeps track of bit value and bit position for partial paths) is populated using the K-Best paths at the current level and the accumulated chosen discarded paths. The discarded paths at the current level are then examined, in order of increasing PED, and selected (tagged) for further processing only if they fill at least one empty entry in the Bit Occurrences table. Using this improvement idea, the Soft K-Best detector ZF-augments, computes and compares PEDs for a maximum of $(2N_T-1)(\log_2(M)/2)$ - (K-1) relevant discarded paths out of the total $(2N_T-2)K(\sqrt{M}-1)$ discarded paths, thereby greatly reducing computational complexity.

Improvement 2: Last Stage On-Demand Expansion. This improvement reduces the number of path extensions and PED computations from $K\sqrt{M}$ to $2K-1$ and completely eliminates $K\sqrt{M}$ PED comparisons, as follows: First, the K-Best paths at level $2N_T$-1 are extended to exactly K paths at level $2N_T$ using ZF augmentation. Next, the extended K paths are used to fill a MinPED table for the first $(2N_T-1)(\log_2(M)/2)$ bits, since they always yield the smallest PED values for the first $2N_T$-1 symbols (first children of K-Best paths by definition contribute the smallest incremental Euclidean distance). For the last $(\log_2(M)/2)$ bits: first, the lowest PED ZF-augmented path at level $2N_T$ is used to fill exactly half of the MinPED table; then, an on-demand extension [1] uses at most $2K-1$ sorted paths to

TABLE I
PROPOSED SOFT K-BEST DETECTION SCHEME

1) Find the K-Best children of level 1. ($\mathcal{K}_1$).

2) For $l = 2 : 1 : 2N_T - L - 1$
  2.1) Find the K-Best paths at level l ($\mathcal{K}_l$),
    using on-demand expansion [1].
End

3) For $l = 2N_T - L : 1 : 2N_T - 1$
  3.1) Find the K-Best paths at level l ($\mathcal{K}_l$).
  3.2) Populate a Bit Occurrences table for
    levels l -> 1 using $\mathcal{K}_l$.
  3.3) Update Bit Occurrences table using the
    discarded paths accumulated from levels
    $l - 1$ -> 1 ($\mathcal{D}_{l-1 \to 1}$). Copy $\mathcal{D}_{l-1 \to 1}$ to $\mathcal{D}_{l \to 1}$.
  3.4) Examine discarded paths at level l ($\mathcal{D}_l$)
    and select it (add to $\mathcal{D}_{l \to 1}$) if it fills
    a void entry in Bit Occurrences table.
  3.5) Perform ZF augmentation for each path
    in $\mathcal{D}_{l \to 1}$ to level $l + 1$.
End

4) At level $l = 2N_T$
  4.1) Sort the paths in $\mathcal{D}_{2N_T-1 \to 1}$ in the
    ascending order of PED.
  4.2) ZF Augment K-Best paths $\mathcal{K}_{2N_T-1}$ to
    exactly K paths at level $2N_T$.
  4.3) Use these K paths to fill MinPED table
    for first $(2N_T-1)(\log_2(M)/2)$ bits.
  4.4) For the last $(\log_2(M)/2)$ bits:
    4.4.1) Use the lowest PED ZF augmented path
      to fill half of the MinPED table.
    4.4.2) Perform On-Demand extension of $\mathcal{K}_{2N_T-1}$
      and use at most 2K-1 paths, to fill
      the rest half of the MinPED table.

5) Use the sorted $2N_T \times 1$ paths from $\mathcal{D}_{2N_T-1 \to 1}$
  to update the MinPED table using the
  Relaxed LLR Computation scheme.

6) Compute LLR values using the MinPED data
  for the $2N_T(\log_2(M)/2)$ transmitted bits.

attempt to fill the remaining half of the MinPED table.

Improvement 3: Relaxed LLR Computation Scheme. The improved scheme reduces computational complexity by making an assumption that MinPED values attained from the K-Best paths are always smaller compared to those attained from the discarded paths, thus eliminating a large number of PED comparisons, as follows: First, the MinPED table is filled for all $N_T\log_2(M)$ transmitted bits using the last level extension of K-Best paths. Next, for each discarded path, instead of comparing its PED with the existing MinPED, an entry in the MinPED table is filled only if the entry is still void. Finally, the MinPED data is used to compute LLR output according to (4) for $N_T\log_2(M)$ bits. This results in significant computational savings at the expense of approximating LLR computation, the impact of which is discussed in Section V.

IV. Soft K-Best VLSI Implementation

The proposed architecture for a 4×4 64-QAM Soft-output K-Best MIMO detector with $K = 10$ is shown in Figure 1. This architecture implements the detection algorithm in Table I. The detector consists of $2N_T = 8$ stages, denoted L8 through L1, corresponding to an 8-level detection tree. The first level of the tree represents the last row of equation (2). The Level I block expands all $\sqrt{M} = 8$ possible children in Ω and calculates the updated PEDs for each path, forming K_1 at the output. For each of the paths in K_1, a child with the lowest PED, called the First Child (FC), is found in Level II. The Sorter block sorts all 8 FCs from each path in just 4 clock cycles and transfers the sorted FCs from L7 to the next level
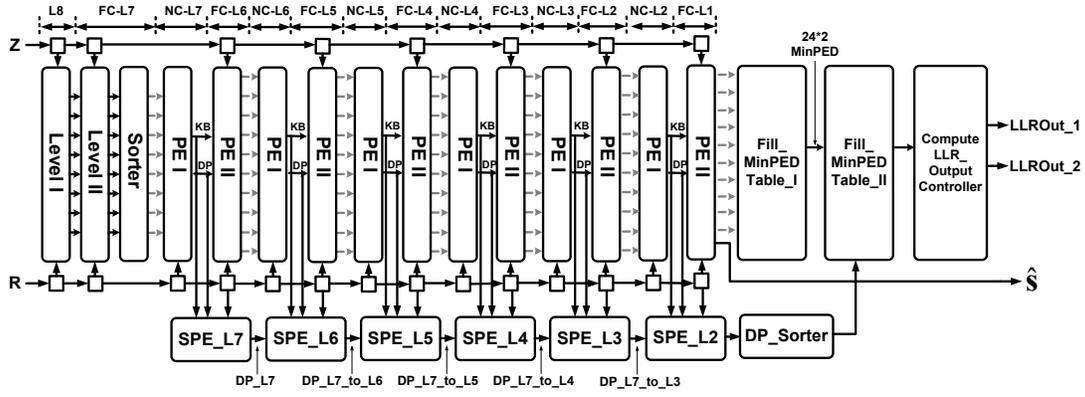
Fig. 1. The Proposed VLSI architecture of a 4×4 64-QAM soft-output K-Best MIMO detector with $K = 10$.

L6. Notice that every level contains two Processing Elements (PEs): PE I and PE II. The PE I block takes the FCs of each level and uses a PED sorter to generate the K-Best (KB) paths of the current level one-by-one and retains $K - 1$ discarded paths (DP). The PE II block receives the KB paths of the previous level one-by-one, computes their First Child (FC) and sorts these FC augmented K-Best paths as they arrive.

The Soft-output Processing Element (SPE) and DP_Sorter blocks together create the DP datapath that retains only the relevant DPs from each tree level and performs ZF augmentation, PED computation and PED sorting. The *SPE* block, shown in Figure 2, implements *Improvement 1* (steps 3.2-3.5 in Table I). The NBO sub-block observes the KB paths and accumulated chosen DPs to populate the Bit Occurrences Table at the current level. The TDP sub-block then uses the Bit Occurrences Table to select or tag only those DPs that contribute to LLR computation at the current level. The FC sub-block calculates ZF augmentation for selected DPs at the current level, the accumulated chosen DPs, and updates their PED values. Figure 3 shows the detailed architecture of DP_TDP, a sub-block of the TDP block, for a single bit. Logical XNOR operations with '0' and '1' were used to classify the current input bit. A special reset mechanism was implemented using logic OR and AND, to reset the sub-block in the same clock cycle in which the first KB is processed due to a lack of idle cycles for register reset.

At the end of the DP datapath, the DP_Sorter block sorts all of the accumulated selected DPs in order of ascending PEDs in preparation for LLR computation. The Fill_MinPEDTable_I and Fill_MinPEDTable_II blocks implement *Improvements 2 and 3*, respectively, and together with the ComputeLLR_OutputController block compute soft LLR output using the accumulated selected DPs and extended KB paths at the last tree level. LLR computation is performed in two steps. First, the selected DPs and extended KB paths are observed and their PEDs are compared to fill the MinPED table for each transmitted bit. Secondly, the MinPED table is used to compute the LLR for each bit by subtracting the MinPED values according to (4). The Fill_MinPEDTable_I block performs the task of initializing and
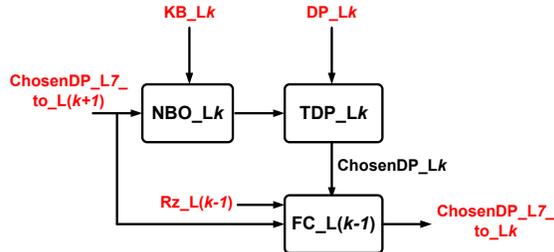
Fig. 2. Overall architecture of the SPE block.
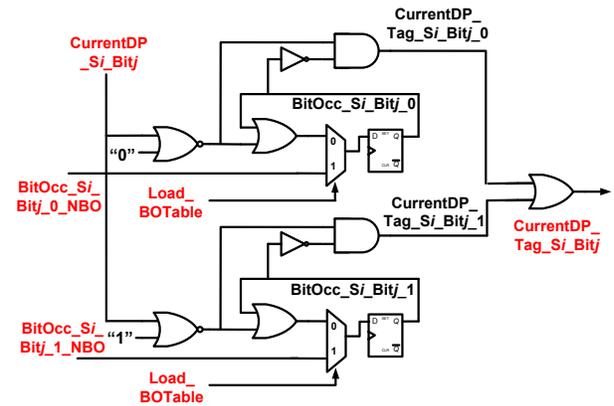
Fig. 3. Architecture of the DP_TDP sub-block.

filling the MinPED table with the $2N_T \times 1$ paths generated by the *Last Stage On-Demand Expansion* scheme (steps 4.2-4.4 in Table I). Figure 4 shows the architecture used to fill in the MinPED table for the jth bit of a real-valued symbol S1, which corresponds to the last $\log_2(M)/2$ transmitted bits. The architecture uses both the KB paths and FC augmented paths to fill the MinPED table and compute the corresponding MinPED tags that will be used in LLR computation. The Fill_MinPEDTable_II block uses PEDs of the selected DPs to update the MinPED table received from the previous stage. Finally, the ComputeLLR_OutputController block computes LLR values for a block of 24 transmitted bits and outputs them in parallel using LLROut_1 and LLROut_2 ports every 10 clock cycles.

V. PERFORMANCE AND COMPLEXITY COMPARISON

The proposed Soft-output K-Best detector offers a significant reduction in the number of operations required for LLR computation as shown in Table II. By processing only the relevant paths and using approximations when appropriate, the number of PED computations, comparisons and ZF augmentations required is reduced by approximately a factor of 5 compared to the MKSE scheme.

The reduction in computation comes at the expense of BER performance, as shown in Figure 5. The BER plot was obtained by integrating Soft K-Best MATLAB models with CTC Encoder and Decoder models configured for a 4×4 64-QAM MIMO system with $K = 10$, $L = 6$, Code-Rate = 1/2, 600 bytes/block, and 8 decoder iterations. The first improvement idea, *Relevant Discarded Paths Selection*, results in BER degradation of 0.4 dB at BER = 10^{-3}, compared to the MKSE scheme. The loss is caused by the fact that PEDs of unselected DPs are ignored. Thus, an error occurs when an unselected DP yields a smaller final Euclidean distance compared to a selected DP, which was found to occur in 7.75%
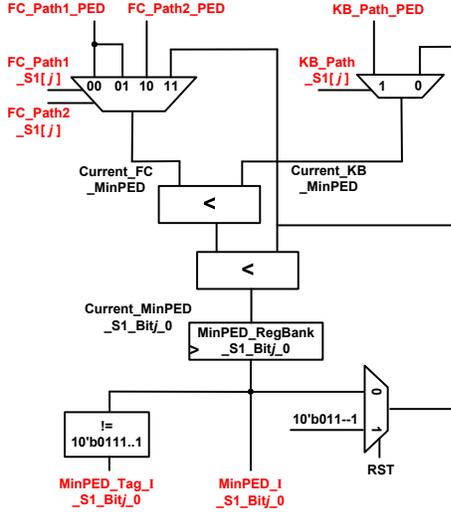
Fig. 4. Architecture of the `Fill_MinPEDTable_I_Part2` sub-block.

of the cases through extensive simulations. However, the other two improvement ideas, *Last Stage On-Demand Expansion* and *Relaxed LLR Computation*, incur a loss of 0.1 dB and 0.04 dB at BER = 10^{-3}, respectively. Overall, the proposed Soft K-Best detector results in 0.54 dB loss compared to the MKSE scheme, while reducing computational complexity by a factor of 5. Also, the proposed detector improves BER by $\approx$ 1.7 dB and 2.9 dB at BER = 10^{-3}, compared to the conventional Soft K-Best [3] and Hard K-Best detection schemes, respectively.

Comparing this work with other Soft K-Best implementations (Table III) shows a significant reduction in area and increase in throughput. The proposed detector also attains lower power and lower decoding energy per bit, since only a tiny fraction of all children are expanded in the proposed architecture. Note that [6] offers a much smaller power consumption. However, the attained throughput is only 115Mbps. Also, the detection method in [6] uses approximate sorting and bidirectional tree search that introduce BER loss. The proposed Soft K-Best detector achieves BER improvement of 2.6 dB at BER = 10^{-3} compared to the Hard K-Best detector in [1], while maintaining high throughput of 655Mbps. Table III shows synthesis results of the proposed Soft K-Best detector in 65nm CMOS with typical LVT cells at 1.3V and 25°C. The proposed K-Best detector attains peak data throughput of 2Gbps, while requiring 0.57 mm^2 silicon area and 280mW of power, thus fulfilling the data rate requirements of the emerging IEEE 802.16m and LTE-Advanced wireless standards.

VI. Conclusions

A low-complexity high-throughput VLSI implementation of a Soft-output K-Best Detector was presented for a 4×4 64-QAM MIMO

TABLE II
BASIC OPERATIONS COUNT FOR VARIOUS SOFT K-BEST SCHEMES

Scheme	Metric	Conv. Soft K-Best	MKSE [2]	**Proposed Soft K-Best**
4×4 64-QAM (K=10)	Shift	240	1158	**261**
	Additions	480	1884	**426**
	Multiplications	80	134	**31**
	Comparisons	80	2672	**19**
8×8 256-QAM (K=15)	Shift	720	2148	**869**
	Additions	1440	3624	**1370**
	Multiplications	240	324	**75**
	Comparisons	240	5616	**29**
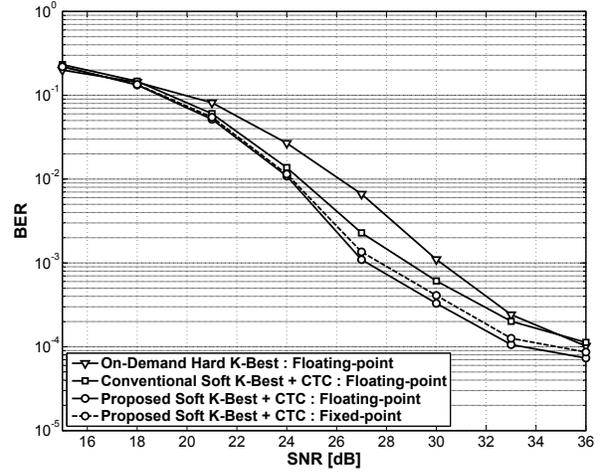
Fig. 5. BER Performance of various K-Best schemes (4x4, 64-QAM, K=10)

TABLE III
COMPARISON OF THE CURRENT ASIC IMPLEMENTATIONS OF 4×4
SOFT-OUTPUT MIMO DETECTORS.

Reference	[3]	[6]	This work $0.13\mu m$	This work 65nm
Modulation	64-QAM	64-QAM	64-QAM	64-QAM
Method	K-Best	K-Best	K-Best	K-Best
K-value	64	64	10	10
Process	0.13 μm	65 nm	0.13 μm	65 nm
Core Area	5270 KG	174 KG	174 KG	298 KG
Max Freq.	270 MHz	200 MHz	270 MHz	833 MHz
Throughput	100 Mbps	115 Mbps	655 Mbps	2000 Mbps
Latency	N/A	N/A	0.7 μs	0.23 μs
Power	847 mW	11 mW	195 mW	280 mW
Energy/bit	8470pJ/b	96pJ/b	298pJ/b	140pJ/b
Soft/Hard	Soft	Soft	Soft	Soft
SNR Dep.	Yes	Yes	No	No
Domain	Complex	Complex	Real	Real

system. Using a novel Soft K-Best detection scheme, the proposed detector achieved 5x reduction in computational complexity and a 1.7 dB gain at 10^{-3} BER performance improvement compared to the Conventional Soft K-Best detection scheme. The detector provided a sustained throughput of 655Mbps, while requiring 174K gates and 195mW in $0.13\mu m$ CMOS. Characterization of the proposed detector in the $65nm$ CMOS showed a potential to support the high data rate applications driven by IEEE 802.16m and LTE-Advanced wireless standards.

REFERENCES

[1] M. Shabany and P. G. Gulak, "Scalable VLSI Architecture for K-Best Lattice Decoders," *in Proc. IEEE ISCAS 2008*, pp. 940–943, May 2008.
[2] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-Best sphere decoding for MIMO detection," *IEEE Journal on Sel. Areas in Comm.*, vol. 24, no. 3, pp. 491–503, Mar. 2006.
[3] S. Chen, T. Zhang, and Y. Xin, "Relaxed K-best MIMO signal detector design and VLSI implementation," *IEEE Trans. on VLSI Systems*, vol. 15, no. 3, pp. 328–337, Mar. 2007.
[4] H.-L. Lin, R. C. Chang, and H. Chan, "A high-speed SDM-MIMO decoder using efficient candidate searching for wireless communication," *IEEE Trans. on Circuits, Syst. II*, vol. 55, no. 3, pp. 289–293, Mar. 2008.
[5] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Comm*, vol. 51, pp. 389–399, Mar. 2003.
[6] S. Chen and T. Zhang, "Low power soft-output signal detector design for wireless MIMO communication systems," *in Proc. International Symp. on Low Power Electronics and Design*, pp. 232–237, 2007.