A scalable MIMO near-maximum likelihood (ML) detector

Eduardo Lopez-Estraviz and Liesbet Van der Perre, IMEC - November 30, 2008

IMEC has developed a multiple input multiple output (MIMO) near-maximum likelihood (ML) detection technique that is scalable in terms of performance and power consumption. The scalable MIMO functionality has been designed to meet the high spectral efficiency requirements in next-generation wireless standards such as 3GPP LTE. The scalable MIMO detector will be implemented on a software defined radio (SDR) platform that is based on parallel architectures. This implementation approach supports different communication standards and meets today's demand for high spectral efficiency, low power consumption, low cost and reduced time to market.

Introduction

In order to address high spectrum costs, emerging standards such as 3GPP LTE provide high spectral efficiency. These emerging standards achieve high spectral efficiency through a combination of orthogonal frequency division multiplexing (OFDM) and spatial multiplexing (SM). To fully exploit these techniques, LTE terminals require high-complexity and power-consuming detection techniques such as multiple input multiple output (MIMO) maximum likelihood (ML) detection. Furthermore, wireless user terminals require support for multiple standards in order to operate in different parts of the world. This flexibility comes at the penalty of increasing design complexity and silicon area, since more functionality has to be implemented in the baseband platform. These requirements add up to a major challenge, because LTE terminals must be developed at low cost, short time to market and low power consumption in order to be competitive.

IMEC provides a solution for the majority of these challenges with an SDR implementation of a scalable near-ML detector. This solution adaptively and efficiently scales the ML search space. Firstly, the detection technique is scalable in terms of performance and power consumption. This scalability tailors the power consumption to the required high spectral efficiency given certain channel propagation conditions. Secondly, the scalable LTE MIMO functionality has been designed to be implemented on an SDR platform based on parallel architectures. This approach supports reduced time to market and power consumption, and it supports different communication standards. Thirdly, in combination with a flexible and scalable forward error correction (FEC) architecture, it allows a cross-layer controller to tune the detector and the FEC performance/complexity to obtain minimal power consumption given a required data throughput.

In this article, the novel MIMO near-ML detection technique is presented and its scalability in both performance and power consumption is evaluated. Next, the co-optimization of the near-ML detector algorithm and implementation for parallel architectures is discussed. Finally, the benefits of these approaches are summarized in view of the challenges that today's wireless terminals impose.

Multiple antenna detection

MIMO wireless systems use spatial multiplexing (SM) as a way of increasing the data rate for 4G systems. In spatial multiplexing, a high rate signal is split into multiple lower rate streams and each
stream is transmitted from a different transmit antenna in the same frequency channel. Although SM MIMO is a very powerful technique, its maximum benefits in terms of spectral efficiency can only be reached when advanced detection techniques such as maximum likelihood (ML) are applied. ML is a well-known non-linear technique that allows an optimal detection of signals transmitted over MIMO channels. It offers a higher performance than linear detectors in SM MIMO systems.

However, ML detectors are highly complex, with their complexity depending on the size of the likelihood search space. This in turn depends on the number of antenna-parallel streams, constellation order and excess delay. The size of this search space can be significantly reduced by combining ML detection with OFDM, which efficiently handles the problems created by multi-path channels. OFDM simplifies the multi-path channel propagation into orthogonal frequency domain signal tap propagation channels. OFDM modulation and MIMO are therefore two key components for supporting most of the emerging (and on-going) wireless communication standards such as 3GPP LTE. Still, in most cases, the complexity requirements of ML detection when applied to SM MIMO OFDM prohibit its practical implementation.

The complexity of the likelihood search space can be reduced by applying so-called sub-optimal techniques. In recent years, several scalable, low-complexity sub-optimal ML likelihood detectors have been proposed. Among the proposed methods, tree search detection algorithms are the most promising ones. Most contributions in this field constrain the complexity in a uniform and sub-optimal way by performing a uniform tree search which does not take into account the channel propagation conditions and signal-to-noise ratio. Although most of the proposed solutions result in a simple implementation, they suffer from severe performance degradation and require a large number of operations for achieving near-ML performance. **Novel scalable tree-search near-ML detector**

IMEC presents a novel near-ML detector that is scalable in performance and complexity of the detector. Below, the general principles are highlighted and the performance of the proposed detector is evaluated.

A classical ML detector searches through an entire vector constellation set for the transmitted vector that maximizes the probability of the received vector given certain channel propagation coefficients. This involves solving a minimization problem with a number of possible candidates involved. However, for high modulation schemes such as 64-quadrature-amplitude-modulation (64QAM) and a number of transmit antennas equal to 4, the number of possible candidates amounts to $64^4 = 16,777,216$. Due to this high value, the implementation of ML detectors becomes unaffordable from a complexity point of view.

To reduce the ML search space size, IMEC proposes to decompose the minimization in smaller minimization problems for each of the transmitted data symbols. In the novel detection method, the number of nodes involved in the tree search is derived as a function of the noise variance and multiple antenna channel coefficients. This approach results in a more efficient distribution of the complexity during the tree search. Although the tree search is not uniform, the number of nodes in each of the tree levels may be obtained deterministically, hence facilitating its implementation. In order to bring a complexity/performance tradeoff, the number of nodes is constrained according to a certain input parameter $p$ which scales the number of candidates for each of the transmitted data symbols in the tree search. This input parameter will determine, together with the noise variance and channel coefficients, the performance and complexity of the detector.
Simulation results on a cellular system demonstrate the scalability of the detector and the performance/complexity trade-off that it enables. Simulations are performed over 50 channel realizations. The channel propagation conditions are simulated according to the 3GPP channel model for MIMO simulations. The number of antennas at both transmitter and receiver sides is 4. The MIMO system is fully loaded with 4 parallel transmitted symbol streams. Sampling frequency is 7.86MHz; carrier frequency is 2GHz. OFDM modulation parameters correspond to a 5MHz transmission bandwidth.

The performance has been derived for QPSK, 16QAM and 64QAM modulation, the latter being shown in figure 2. The ML and linear zero-forcing detector performances are added for comparison. The performance (figure 2 left) and scalable complexity (expressed in terms of the number of accessed nodes during the tree search, figure 2 right) values are given for different values of p. For the 64QAM modulation the maximum number of branches in each tree level has been limited to 32 possible symbols. A large reduction in complexity is achieved compared to ML detection which would require 64^4 accessed nodes. This scalable detector is particularly interesting for the implementation of MIMO detectors in power-constrained receivers.

Near-ML detector for parallel programmable baseband architectures

Today, almost all implementations of (near-)ML detectors are delivered in application specific integrated circuits (ASICs) or field programmable gate arrays (FPGAs). However, it is advantageous to design the scalable MIMO functionality targeting an SDR platform that is based on parallel architectures such as a very-long instruction word (VLIW) digital signal processor (DSP) or IMEC's
architecture for dynamically reconfigurable embedded systems (ADRES) [1,2]. Indeed, a SDR concept reduces time to market; it allows supporting different standards with the same hardware, hence costing less silicon area; and the use of parallel architectures is more efficient in terms of power consumption.

Unfortunately, none of the existing near-ML detectors fit programmable architectures. Sphere decoders and its variants are essentially sequential and non-deterministic, making parallelization more difficult. K-Best and its variants involve modular and repetitive operations that are easily parallelized in hardware. However, they have a fundamental problem when being mapped on parallel programmable architectures: the so-called spanning-sorting-deleting process incurs irregular dataflow, non-deterministic control flow, extensive shuffling and extensive memory rearrangement. These characteristics result in very low resource utilizations on parallel programmable architectures.

Although for hardware, the architecture can be tuned to fit algorithms, for programmable platforms the algorithms must be elaborately designed to fit the given architecture, so that efficient resource utilizations can be achieved. Therefore, in order to bridge the algorithm-architecture gap, IMEC has co-optimized its LTE near-ML MIMO detector algorithm and implementation for parallel programmable baseband architectures. Key feature of the novel detection technique is the completely deterministic and regular dataflow structure [3]. Abundant vector parallelism is enabled; memory rearrangement, shuffling operations and non-deterministic dynamism are all eliminated.

In essence, the novel implementation method for the tree based near-ML search consists of a distributed and greedy algorithm, similar to the dynamic programming. The ML search is performed starting from the transmitted symbol level $i = N_t$. The detection spans each node at level $i+1$ to $m_i$ nodes at level $i$. The spanned nodes are never deleted. If the node at level $i = N_t+1$ has the associated partial symbol vector being $s^{i+1} = [s_{i+1},... ,s_{N_t}]$, the spanning is to select a set of $s' = [s_i, s_{i+1},... ,s_{N_t}]$ in the way that $||e_i(s)||^2$, the increment of the partial Euclidean distance (PED) of the partial symbol vector $s_i$, is minimized. The proposed method is a distributed and greedy algorithm because it minimizes $||e_i(s')||^2$ at each level of the tree; it is distributed because the minimization of $||e_i(s')||^2$ is local for each node at level $i+1$ when spanning this node to $m_i$ nodes at level $i$. This distributed and greedy algorithmic structure enables very simple and efficient heuristics to minimize $||e_i(s')||^2$ and allows a very efficient implementation of the near-ML search in parallel architectures.

Since the proposed dataflow is completely deterministic and regular, it is simple to parallelize it for parallel programmable architectures. There are at least two options. Firstly, we can observe that the spanning operations of different nodes can be parallelized. Secondly, we can search multiple trees simultaneously. The second scheme is preferred because of the following advantages: (1) it brings
abundant vector parallelism, which can be easily mapped on VLIW, single-instruction multiple-data (SIMD) or vector architectures; (2) the parallelization is scalable, the number of parallel trees can be determined according to the supported parallelism on the given architecture; (3) this scheme perfectly fits OFDM systems, where the detection is essentially parallel for blocks of MIMO symbols.

![Graphs showing BER comparisons for different environments](image)

*Figure 4. Bit-error-rate (BER) comparisons, demonstrating the quality/cost scalability in the novel detector (referred to as SSFE).*

The near-ML detector has been evaluated for coded OFDM transmissions over 802.11n channels and 3GPP channels. Under the same performance constraints, its complexity is significantly lower than
the K-Best, the most popular detector implemented in hardware. More importantly, it can be easily parallelized and efficiently mapped on programmable baseband architectures. As an illustration, with TI TMS320C6416T (a real-life commercial fixed-point VLIW DSP), the detector delivers 37.4 - 125.3Mbps throughput for 4x4 64QAM transmissions. To the best of our knowledge, this is the first reported near-ML MIMO detector explicitly designed for parallel programmable architectures and demonstrated on a real-life platform. **FEC decoding**

**Summarizing the benefits**

With IMEC's novel MIMO near-ML detection approach, several of the major challenges for tomorrow's wireless terminals have been tackled. Firstly, the scalable MIMO near-ML detector offers the required high spectral efficiency presupposed by emerging standards such as 3GPP LTE. Secondly, the scalable LTE MIMO functionality has been designed to be implemented on an SDR platform based on a parallel architecture. This is in favor of a reduced time to market, small silicon area and minimal power consumption, and allows supporting different communication standards. Thirdly, the scalability provided in the FEC decoder and the scalable near-ML detectors to trade off performance versus power can be exploited by an intelligent controller in order to save power consumption at the system level. The key observation is that SDRs are typically facing very dynamic operating conditions, in terms of propagation conditions and radio environment, on the one hand, and in terms of link utilization resulting from application requirements on the other hand. By carefully adapting to these dynamics at runtime, building on the scalability enabled by the baseband engine and the FEC decoder, much energy can be saved with respect to conventional worst case design.

**Conclusion**

A novel MIMO near-ML detection technique is proposed that is scalable in both performance and complexity. Simulation results on a cellular system demonstrate the offered scalability and the reduced complexity compared to a conventional near-ML detection technique. In addition, the scalable MIMO near-ML detector has been designed in order to be easily parallelized and mapped on parallel architectures, as demonstrated for LTE functionality on a real-life platform. The needs for and ongoing research of a flexible advanced FEC decoder are highlighted. The scalability provided in the baseband and FEC algorithms will eventually be exploited by an intelligent cross-layer controller to minimize the overall power consumption of the SDR system.

With the presented approach, several major needs for today's wireless terminals – being high spectral efficiency; low production complexity, power consumption, production cost; fast time to market; support of multiple standards and reduced silicon area - have been answered.

**References**


