A 1Gbps LTE-Advanced Turbo-Decoder ASIC in 65nm CMOS

Sandro Belfanti, Christoph Roth, Michael Gautschi, Christian Benkeser, and Qiuting Huang
Integrated Systems Laboratory, ETH Zurich, 8092 Zurich, Switzerland
e-mail: {belfanti,rothc,gautschi,benkeser,huang}@iis.ee.ethz.ch

Abstract

This paper presents a turbo-decoder ASIC for 3GPP LTE-Advanced supporting all specified code rates and block sizes. The highly parallelized architecture employs 16 SISO decoders with an optimized state-metric initialization scheme that reduces SISO-decoder latency, which is key for achieving very-high throughput. A novel CRC implementation for parallel turbo decoding prevents the decoder from performing redundant turbo iterations. The 65nm ASIC achieves a record data throughput of 1.013Gbps at 5.5 iterations with unprecedented energy efficiency of 0.17nJ/bit/iter.

Keywords: LTE-Advanced, mobile communications, turbo decoder, ASIC implementation, CRC, early termination

Introduction

The exploding demand for data-centric applications on mobile devices has led to ever-increasing data rates in mobile communication standards. The most recent 3GPP standard evolution, LTE-Advanced (LTE-A), specifies peak data rates of up to 3Gbps [1] by adding 8x8 MIMO and increasing the radio bandwidth to 100MHz, which is almost 10x the maximum throughput of the LTE standard currently deployed around the world. In fact, LTE-A is the first true 4G standard meeting all requirements imposed by ITU [2], which includes providing at least 1Gbps peak throughput. These high throughput requirements combined with the flexibility mandatory to cope with a wide range of code rates and block sizes [3] render the turbo decoder one of the key challenges in the design of LTE-A user equipment.

An efficient way to realize high-throughput turbo decoders is to use multiple parallel soft-input soft-output (SISO) decoders that share the same input and intermediate memory [4]. However, with increasing degree of parallelism, the internal latency of the SISO decoders becomes an increasingly large fraction of each turbo half-iteration. Due to the iterative nature of the decoding algorithm, this strongly affects the decoding throughput. This effect is especially pronounced for high code rates, where large window sizes are required to achieve best decoder performance, which makes the implementation of high-throughput turbo decoders especially challenging for high code rates. It is therefore essential to specially address the problem of the SISO-decoder latency for an efficient and flexible high-throughput implementation.

Building on our previous work [4], we present the first fabricated LTE-A turbo-decoder ASIC that achieves a measured throughput of 1Gbps for all specified code rates.

ASIC Architecture

The architectural diagram of the implemented decoder is shown in Fig. 1. It comprises \( p = 16 \) windowed radix-4 SISO decoders, 4 input RAMs storing systematic and parity log-likelihood ratios (LLRs), 2 intermediate RAMs for the extrinsic LLRs, address generators with interleaving networks as proposed in [5], and a CRC-computation unit for early termination.

Our previous windowed radix-4 SISO-decoder architecture [4] has been complemented by a carefully optimized combination of state-metric propagation and acquisition runs to initialize the window boundaries. Compared to conventional acquisition-based SISO-decoder implementations, our novel initialization approach reduces the required window size (and thus SISO-decoder latency) significantly for all code rates, which is the key to obtain a flexible and efficient high-throughput decoder. Note that the achieved reduction in required window size comes without any loss in BER performance.

A. Overlapping Turbo Half-Iterations

We further limit the effect of SISO-decoder latency on throughput by performing overlapping half-iterations as described next. Conventional turbo decoders process consecutive half-iterations strictly sequentially, which requires all computation stages in the SISO decoder to be flushed at the
end of each half-iteration. In contrast, consecutive half-iterations can be processed in our implementation in an overlapping fashion, where all computation stages in the SISO decoders are always fully utilized. This novel approach eliminates the impact of SISO-decoder latency on the iterative decoding algorithm and, thus, leads to a considerable throughput gain. It can even be realized without additional silicon complexity or address conflicts. However, the overlapping processing generates data dependencies across consecutive half-iterations, which may degrade the BER-performance depending on the fraction of LLRs processed in the overlap. The analysis of the issue has shown that the resulting BER loss is negligible (roughly 0.03dB) for lower code rates, where small window sizes are required (as illustrated in Fig. 2). In this regime of operation, performing overlapping half-iterations is a highly efficient approach to further boost throughput.

B. CRC-Based Early Termination

In order to increase the energy efficiency of the decoder, our realization avoids redundant turbo iterations by identifying correctly decoded code blocks using the 24-bit CRC checksum defined in the standard. In each non-interleaved half-iteration, the CRC unit computes the CRC of the current code block and triggers the end of decoding when the check passes. The fact that a parallel radix-4 architecture discloses $p$ non-contiguous pairs of bits per clock cycle prevents regular, straightforward CRC implementation by a linear feedback shift register (LFSR) to be used. We will now present a novel architecture which solves this problem efficiently for all block sizes and $p$. Relying on the distributive property of the CRC, our novel approach calculates in each clock cycle the checksum of the code sub-block represented by the $2p$ disclosed bits and then accumulates these individual CRCs to obtain the final result. The computation of the individual checksums is divided into two steps (see Fig. 1) in order to make it as efficient as possible. In the first step, $p$ CRC cores use look-up tables (LUTs) to calculate checksums of the code sub-blocks that contain only the two output bits of the corresponding radix-4 SISO decoders. In the second step, the $p$ checksums are combined and finally added to the content of the accumulation register. This modular approach combined with carefully tuned LUTs enables the CRC unit to support all block sizes and degrees of parallelism.

**ASIC Measurement Results**

The LTE-A-compliant [3] turbo-decoder ASIC has been realized in 65nm CMOS. The corresponding micrograph is shown in Fig. 3. In order to achieve best throughput for all block sizes and code rates, the decoder window-size can be tuned between 14 and 30. The implementation loss in terms of BER performance is only 0.14dB for all code rates, as indicated in Fig. 2. TABLE I summarizes the measurement results of the turbo decoder and provides a comparison to other published LTE turbo decoders. At the maximum clock frequency of 410MHz, the highest block size of 6144 is processed with a throughput of 1013Mb/s at a code rate of 0.95 using a window size of 30 at 5.5 iterations. At low code rates, optimum decoder performance can be achieved with a window size of only 14, and the throughput additionally boosted to 1161Mb/s with overlapping half-iterations. At maximum throughput the power consumption is 966mW, which corresponds to an unprecedented energy efficiency of 0.17nJ/bit/iter. In strong contrast to the other state-of-the-art implementations in TABLE I, our decoder has been optimized for the entire range of code rates, which enables to maintain high throughput even for highest code rates without any concessions in BER performance.

**References**