A 5.8mW 3GPP-LTE Compliant 8×8 MIMO Sphere Decoder Chip with Soft-Outputs

Chia-Hsiang Yang, Tsung-Han Yu, Dejan Marković

Electrical Engineering Department, University of California, Los Angeles, CA, USA

Abstract

A MIMO chip for 3GPP-LTE standard and beyond is described. The chip implements sphere decoding algorithm with 16-core architecture. The chip is flexible to support multiple configurations: antenna arrays from 2×2 to 8×8, modulations from BPSK to 64QAM, FFT sizes from 128 to 2048 and hard/soft outputs. The chip dissipates 5.8mW for the 3GPP-LTE standard in 3.35mm^2 area in 65nm CMOS.

Introduction

The 3GPP Long Term Evolution (LTE) is the next generation cellular standard. LTE provides up to 100Mbps and 50Mbps for downlink and uplink. Two key technologies used in the LTE standard are Orthogonal Frequency Division Multiplexing (OFDM) and Multiple Input Multiple Output (MIMO) transmission. OFDM systems transmit data over several narrow-band sub-carriers. MIMO systems leverage spatial-multiplexing and diversity for improved data rate and range, adding another layer of data dimensionality.

LTE design specs are listed in Table I. The convergence of a variety of operation modes argues for a more flexible implementation, which makes the integration under a fixed power budget very challenging. In this work, we demonstrate an 8×8 OFDM-based MIMO sphere decoder with a 128-2048 FFT block and soft-outputs in 3.35mm^2, dissipating 13.83mW in a standard-V_T 65nm CMOS technology. Operating at 160MHz, the chip provides a peak data rate of 960Mbps in the 8×8, 64QAM mode over a 20MHz channel. LTE specs are met with 5.8mW power for throughput of 480Mbps.

System Architecture

The receiver architecture considered in this work is shown in Fig. 1. The signals captured by multiple receive antennas are digitized through ADCs and then converted back to frequency domain through FFT. The modulated data streams carried over the narrow-band subcarriers are constructively combined through the MIMO decoder. Soft-outputs are generated for advanced error correction signal processing. We have demonstrated a flexible (antenna, modulation, search method) MIMO sphere decoder in [1] for one signal band. In this work, the flexibility is extended to multiple signal bands. We integrate a sphere decoder kernel, a reconfigurable FFT block, a soft-output register bank, and a pre-processing unit.

The sphere decoding algorithm is one of the most promising solutions to achieve maximum likelihood (ML) performance with practical computational complexity. It is realized by reducing the search space without discarding the ML solution. The ML solution is chosen according to the ML criterion, as shown in Fig. 1. Scalable processing element (PE) used in a multi-core architecture (Fig. 2) allows the flexibility in search method. With the capability of supporting 8×8 antenna arrays, the system outperforms 4×4 systems by 5dB in BER performance by leveraging repetition coding [2].

A reconfigurable FFT block is implemented by several small processing units (PUs), as shown in Fig. 3. The PUs are modular to support different radix representations, allowing power-area tradeoffs. The FFT block is scalable to support 128 to 2048 points by changing data-path inter-connection. Multi-path single-delay-feedback (SDF) architecture provides high utilization for varying FFT sizes. Unused PUs and delay lines are clock-gated for power saving. Twiddle (TW) factors are generated by trigonometric approximation [3] instead of fetching coefficients from ROMs for area reduction. The TW factors for FFT size 1536 are calculated through hardware reuse, which is infeasible for the ROM-based design.

The chip supports both hard-outputs (1/0) and soft-outputs. Soft-outputs are calculated by log-likelihood ratio $L(\mathbf{s}_{i,b}|\mathbf{y})$ to describe the *likelihood* of each decoded bit $s_{i,b}$ (i^{th} antenna, b^{th} bit). Soft-outputs provide a 2-3dB improvement in BER performance in conjunction with proper channel coding schemes, [4]. Fig. 4 shows the block diagram of soft-output register bank. Soft-outputs are supported through a low-power clock-gated register bank. Operated in the WRITE mode, only registers storing $s_{i,b}$ are updated, while others are clock-gated to minimize power dissipation. In the READ mode, the register contents are scanned-out through a three-dimensional (antenna-bit-data stream (DS)) delay-line based shift-register chain to minimize power dissipation. The pre-processing unit consists of eight arithmetic units to calculate $\tilde{\mathbf{y}}$ in parallel. $\mathbf{Q}$ is pre-computed and updated at a slower packet rate, so only 3 multipliers and adders are required for a complex multiplication.

Power-Area Minimization

Area and energy minimization is achieved by combining signal processing, architecture, and circuit techniques. High energy efficiency is reached through aggressive V_{DD} scaling. The supply voltages of most blocks are scaled down to around 0.4V to balance energy-delay tradeoff [1]. The design has 5 V_{DD} domains. The supply voltages are adjusted according to the frequency specs to achieve minimal power consumption (see Fig. 6). Hard-output MMO sphere decoder kernel from [2] is redesigned to accommodate increased number of subcarriers as required by LTE channelization. The FFT block is optimized for minimal Power-Area product. Given a fixed FFT size, there are several possible arithmetic factorizations, which are mapped to different configurations of PUs in Fig. 3. The optimal configuration is chosen with the minimum power-area product (solid-fill markers in Fig. 5(a)). At the architecture level, the optimal memory partition is decided according to the power-area product. Fig. 5(b) shows possible memory partition schemes and the optimal partition for length 512 and 1024. At the circuit level, register-file (RF) (6T/bitcell) is adopted for the delay line with length larger than 512, as shown in Fig. 5(c). RF-based design is estimated to consume only 29% power with an 85% silicon area compared to the SRAM-based counterpart. DFF-based delay lines are optimal in terms of power and area for length less than 512. Level-shifters are properly inserted between the low-V_{DD} and high-V_{DD} domains.

Measurement Results

Chip testing is performed with the use of a custom FPGA board for pattern generation and data analysis. FPGA feeds the test vectors into the chip and captures the outputs through an integrated Simulink/Matlab control interface. Each block was tested separately for individual voltage scaling. Power consumption for different operation modes over 20MHz band in the 64QAM mode is shown in Fig. 6. The chip dissipates 5.8mW for the LTE standard, 13.83mW for full 8×8 array with soft-outputs. The hard-output sphere decoder kernel achieves E/bit of 15pJ/bit (18.7 GOPS/mW) and outperforms prior work [1][5] (Fig. 7, Table). The chip micrograph and summary are shown in Fig. 7. The chip fully supports LTE and has added flexibility for systems with larger array size or cooperative MIMO processing on smaller sub-arrays.

Acknowledgments

The authors acknowledge the support of the Focus Center for Circuit & System Solutions (C2S2), one of five research centers funded under the Focus Center Research Programs, a Semiconductor Research Corporation Program; Cadence and Synopsys software; H. Chen for help with testing.

References

[1] C.-H. Yang and D. Marković, *Proc. ESSCIRC,* Sep. 2009.
[2] C.-H. Yang and D. Marković, *IEEE TCAS-I*, Oct. 2009.
[3] T. D Chiueh and P. Y. Tsai, John Wiley and Sons, 2007.
[4] B. M. Hochwald et. al., *IEEE Trans. Communications*. Mar. 2003.
[5] M. Shabany et. al., *ISSCC Dig. Tech. Papers*, Feb. 2009.

TABLE I. 3GPP-LTE DESIGN SPECIFICATIONS.

Parameters	Configurations
Bandwidth (MHz)	1.25, 2.5, 5, 10, 15, 20
FFT size	128, 256, 512, 1024, 1536, 2048
Antenna array	1×1, 2×2, 3×2, 4×2
Modulation	QPSK, 16QAM, 64QAM

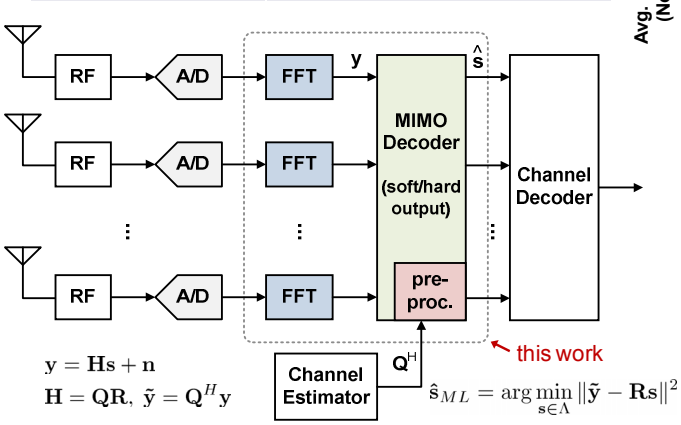


Fig. 1. Receiver architecture for the proposed MIMO decoder.

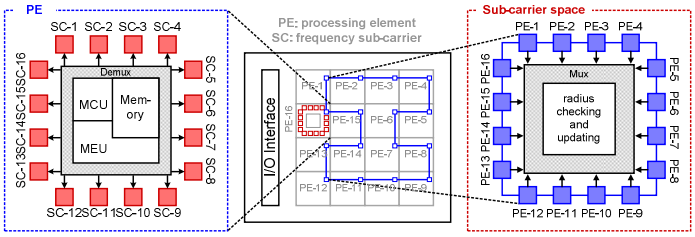
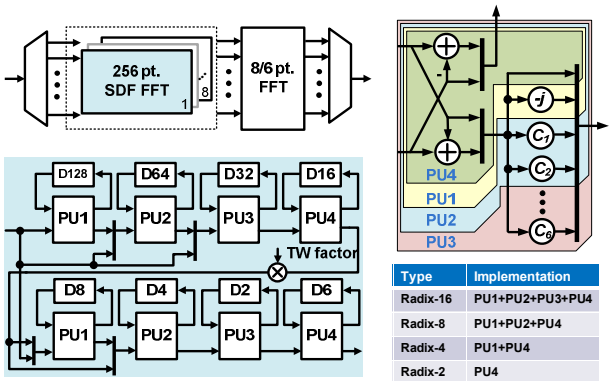


Fig. 2. Hard-output MIMO sphere-decoding kernel.



Type	Implementation
Radix-16	PU1+PU2+PU3+PU4
Radix-8	PU1+PU2+PU4
Radix-4	PU1+PU4
Radix-2	PU4

Fig. 3. Reconfigurable FFT architecture.

$$L(s_{i,b}|\mathbf{y}) \approx \min_{s_{i,b}=0} \Lambda(\mathbf{s},\mathbf{y}) - \min_{s_{i,b}=1} \Lambda(\mathbf{s},\mathbf{y}),\ \Lambda(\mathbf{s},\mathbf{y}) = \frac{1}{N_0}\|\tilde{\mathbf{y}} - \mathbf{Rs}\|^2$$

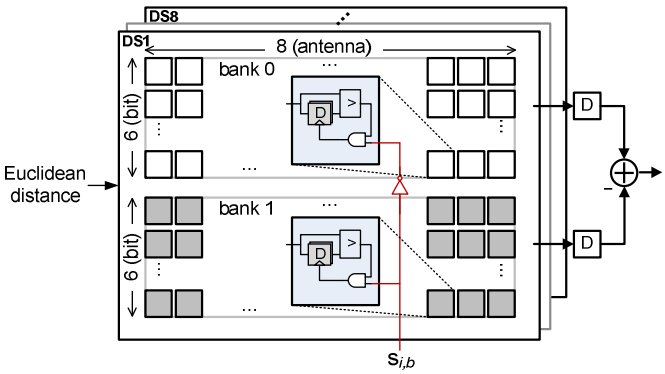
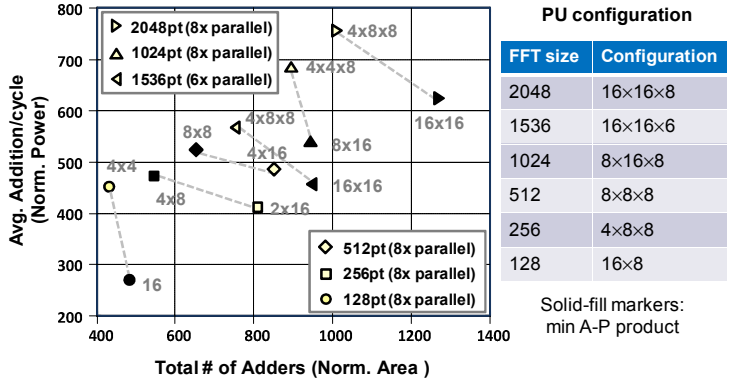


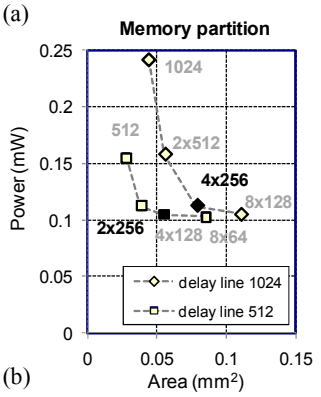
Fig. 4. Low-power clock-gated register bank for soft-outputs.



PU configuration

FFT size	Configuration
2048	16×16×8
1536	16×16×6
1024	8×16×8
512	8×8×8
256	4×8×8
128	16×8

Solid-fill markers: min A-P product



Memory element

512x32 Delay line	
SRAM	RF
0.042mm²	**0.037mm²**
0.82mW	**0.21mW**

1024x32 Delay line	
SRAM	RF
0.052mm²	**0.044mm²**
0.84mW	**0.24mW**

(c) Solid-fill markers: min A-P product

Fig. 5. Optimal FFT configuration and delay-line implementation.

(V_{DD}, f_{clk})	8x8 w/ soft-outputs	4x4 w/o soft-outputs
FFT core (parallelx8)	6.20 (0.45V, 20MHz)	2.83 (0.43V, 10MHz)
RF bank (32kb)	2.35 (1V, 40-160MHz)	1.18 (1V, 20-80MHz)
Hard-output SD kernel (16-core)	0.97 (0.42V, 10MHz)	0.45 (0.36V, 5MHz)
Pre-processing unit	4.06 (0.82V, 160MHz)	1.34 (0.64V, 80MHz)
Soft-output bank (parallelx8)	0.25 (0.42V, 20MHz)	NA
Total power	**13.83mW**	**5.8mW**

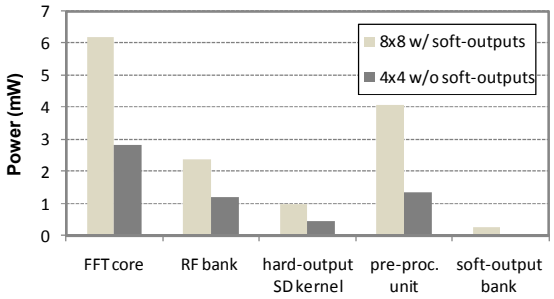
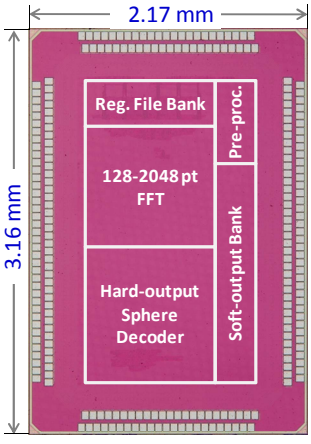


Fig. 6. Power breakdown and measurement results.



Technology	1P9M 65nm Std-V_T CMOS
Max. BW	20 MHz
FFT size	128, 256, 512, 1024, 1536, 2048
Array size	2×2 – 8×8
Modulation	BPSK, QPSK, 16QAM, 64QAM
Outputs	Hard/Soft
Mode	Complex valued
Core area	2.39×1.40 mm²
Gate count	2,946K
Power	13.83mW (5.8mW for LTE)
Energy/bit	15pJ (21pJ [1], 100pJ [5])*

* normalized to 65nm tech.

Fig. 7. Chip micrograph and summary.