

Efficient Stochastic Detector for Large-Scale MIMO

Junmei Yang¹, Chuan Zhang¹, Shugong Xu², and Xiaohu You¹

¹National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

²Intel Collaborative Research Institutes on Mobile Networking and Computing, Intel Labs, Shanghai, China

Email: ¹{jm.yang, chzhang, xhyu}@seu.edu.cn, ²shugong.xu@intel.com

Abstract—In this paper, a low-complexity stochastic belief propagation (BP) detector for large-scale MIMO is first proposed. Its efficient hardware architecture, with parallel pipeline, is presented in detail. Thanks to the stochastic approach, all arithmetic operations of the detector are implemented with simple logic structures. Several approaches which can potentially improve the detection performance are exploited. Simulation results have demonstrated that the stochastic BP detector can achieve similar detection performance compared with deterministic one for 32×32 MIMO system with 4-quadrature amplitude modulation (4-QAM). With the increase of antenna number, the detection performance improves at the linear expense of complexity and latency. Therefore, the proposed stochastic BP detector is suitable for large-scale MIMO system applications with good balance of detection performance and implementation complexity.

Index Terms—Large-scale MIMO, factor graph (FG), belief propagation (BP), stochastic detector, parallel architecture.

I. INTRODUCTION

By transmitting multiple data streams concurrently within the same frequency band, the multiple-input multiple-output (MIMO) system successfully improves system capacity and data rate compared to the single-antenna system. The resulting higher spectral efficiency and better link reliability have made it increasingly popular in both academia and industry. Nowadays, MIMO in combination with spatial multiplexing has been adopted by the latest standards such as 3GPP LTE-Advanced [1] and IEEE 802.11n [2]. In order to fulfill the ever-increasing demands of future wireless communication, MIMO systems are required to be equipped with an order of higher magnitude of antenna arrays than the conventional ones. That is how the popular concept of large-scale MIMO comes out.

Without a doubt, large-scale MIMO has ranked one of the key technologies of 5G, with its significant improvement in spectral efficiency, link reliability, and coverage over conventional MIMO [3]. Unfortunately, its huge size hinders the use of conventional deterministic detection approaches such as *maximum likelihood* (ML) and *minimum mean square error* (MMSE), since their computational complexity grows drastically with the number of antennas.

Stochastic computation, which represents continuous values by streams of random binary bits, can implement complicated computations with simple logic gates. Therefore, it has drawn increasing attentions from the fields of channel coding [4–6] and signal processing [7]. Due to its advantages in hardware efficiency and fault tolerance, it is believed that stochastic belief propagation (BP) detection serves as an alternative solution to this problem. Among the very few existing literatures, a

fully-parallel stochastic Markov Chain Monte Carlo (MCMC) detector for a 4×4 16-QAM MIMO system was proposed in [8, 9], significantly reducing the implementation complexity while achieving high detection throughput. However, this approach turns out to be not suitable for large-scale MIMO systems due to the significantly increasing complexity. To the best knowledge of the authors, stochastic BP detection defined over real domain is first proposed for large-scale MIMO systems. Different antenna configurations are discussed. Theoretical analysis and numerical results have demonstrated its advantages in hardware efficiency and detection performance.

The remainder of the paper is organized as follows. Section II presents the symbol-based BP detector based on factor graphs (FGs), and the stochastic BP detector design for large-scale MIMO system with 4-quadrature amplitude modulation (4-QAM). The efficient hardware architecture is proposed in Section III. Numerical results and complexity comparison are given in Section IV. Section V concludes the entire paper.

II. STOCHASTIC BP DETECTOR ON FGs

In the following, the large-scale MIMO uplink model with M transmitting and N receiving antennas is considered. The deterministic scalar and its stochastic version are denoted by x and x^s , respectively. The deterministic vector (or matrix) and its stochastic version are denoted by \mathbf{x} and \mathbf{x}^s , respectively. For quantization, 1 sign-bit and k data-bits are employed to represent a deterministic value. The length of stochastic stream is then set as $L = 2^k$.

A. BP Detection in Real Domain

Assume each entry of the complex transmitted vector is mapped to one point of a rectangular complex QAM Θ with $|\Theta| = Q$. The real transmitted vector, channel matrix, noise vector, and received vector are denoted by \mathbf{x} , \mathbf{H} , \mathbf{n} , and \mathbf{r} , respectively. Employing the *real value decomposition* (RVD) scheme [10], the equivalent real system model is given by

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{r} = [r_1, r_2, \dots, r_{2N}]^T$, $\mathbf{x} = [x_1, x_2, \dots, x_{2M}]^T$, and $\mathbf{H} = \{h_{j,i}\}_{1 \leq j \leq 2N, 1 \leq i \leq 2M}$. Note that $x_i \in \Omega$, where Ω denotes the set of in-phase or quadrature parts of points in the complex constellation Θ .

We define the transmitted signals $\{x_1, x_2, \dots, x_{2M}\}$ and received signals $\{r_1, r_2, \dots, r_{2N}\}$ as “symbol nodes” and “observation nodes”, respectively. The messages from x_i to r_j and from r_j to x_i , are denoted by $\alpha_{x_i \rightarrow r_j}$ and $\beta_{r_j \rightarrow x_i}$,

respectively. The message updating procedure of BP algorithm is illustrated in Fig. 1. Details of symbol-based BP detection in real domain [11] are listed in Algorithm 1.

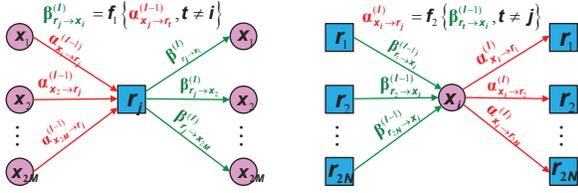


Fig. 1. Message passing between variable nodes and observation nodes.

Algorithm 1 BP Detection Based on FG

Input: $\mathbf{r} \in \mathbb{R}^{2N}$, $\mathbf{H} \in \mathbb{R}^{2N \times 2M}$, $\mathbf{s} \in \mathbb{R}^{\sqrt{Q}}$, $\text{Var}\{\mathbf{n}\} = \sigma^2$.

Iteration:

- 1: $\forall i, j, \mathbf{p}_{i,j}^{(0)} = \mathbf{p}_{x_i \rightarrow r_j}^{(0)} = (p_{i,j}^{(0)}(s_0), \dots, p_{i,j}^{(0)}(s_{\sqrt{Q}-1}))$
- 2: $p_{i,j}^{(0)}(s_k) = 1/\sqrt{M}$
- 3: **for** $l = 1 : L$ **do**
- 4: **for** $j = 1 : 2N$ **do**
- 5: $\mu_{z_j} = \sum_{k=1}^{2M} h_{j,k} \mathbf{s}^T \mathbf{p}_{j,k}$
- 6: $\sigma_{z_j}^2 = \sum_{k=1}^{2M} h_{j,k}^2 ((\mathbf{s} \odot \mathbf{s})^T \mathbf{p}_{j,k} - (\mathbf{s}^T \mathbf{p}_{j,k})^2)$
- 7: **for** $i = 1 : 2M$ **do**
- 8: $\mu_{z_j,i}^{(l)} = \mu_{z_j} - h_{j,i} \mathbf{s}^T \mathbf{p}_{j,i}$
- 9: $(\sigma_{z_j,i}^2)^{(l)} = \sigma_{z_j}^2 - h_{j,i}^2 ((\mathbf{s} \odot \mathbf{s})^T \mathbf{p}_{j,i} - (\mathbf{s}^T \mathbf{p}_{j,i})^2) + \sigma^2$
- 10: $\forall s, \beta_{j,i}^{(l)}(s) = \frac{2h_{j,i}(r_j - \mu_{z_j,i}^{(l)})(s - s_0) - h_{j,i}^2(s^2 - s_0^2)}{2(\sigma_{z_j,i}^2)^{(l)}}$.
- 11: **end for**
- 12: **end for**
- 13: **for** $i = 1 : 2M$ **do**
- 14: $\forall s, \gamma_i^{(l)}(s) = \sum_{k=1}^{2N} \beta_{k,i}^{(l)}(s)$
- 15: **for** $j = 1 : 2N$ **do**
- 16: $\forall s, \alpha_{i,j}^{(l)}(s) = \gamma_i^{(l)}(s) - \beta_{j,i}^{(l)}(s)$
- 17: $\forall s, p_{i,j}^{(l)}(x_i = s_k) = \frac{\exp(\alpha_{i,j}^{(l)}(s))}{1 + \sum_{s_m \in \Omega} \exp(\alpha_{i,j}^{(l)}(s_m))}$
- 18: **end for**
- 19: Symbol-based decision: $\gamma_i \Rightarrow \hat{x}_i = s_k$
- 20: **end for**
- 21: **end for**
- 22: **Output:** Estimated symbol $\hat{\mathbf{x}}$.

B. Stochastic Real Multiplication (SRM)

Assume 2 inputs are x^s and y^s , and the output is z^s . Their deterministic values satisfy $x, y, z \in [-1, 1]$. The SRM is implemented by an XOR gate and an AND gate in Fig. 2. The sign-bit stream of z^s is obtained by an XOR gate. An AND gate is used to generate the absolute value stream of z^s .

C. Stochastic Real Addition (SRA)

According to [12], the structure of SRA is given in Fig. 2. The variable $c(z^s(t))$, regarded as the carry-bit in a signed addition, is used to improve the accuracy. Obviously, the stochastic real subtraction (SRS) can be easily implemented with an additional bit flipping of the sign-bit stream.

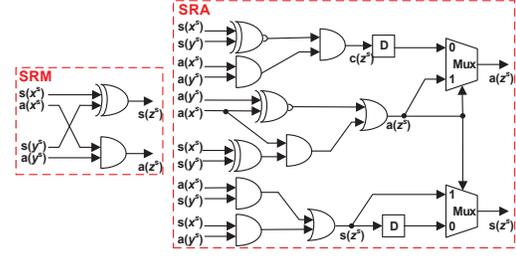


Fig. 2. Hardware architectures of SRM and SRA.

D. Message Updating of Observation Nodes

According to Algorithm 1, a signed division operation is required by the deterministic message updating of observation nodes. Since $\sigma_{z_j,i}^2 > 0$, the sign-bit of numerator is the same as $\beta_{j,i}^{(l)}(+1)$. Therefore, we only need to calculate the absolute value of $\beta_{j,i}^{(l)}(+1)$. In order to implement the stochastic division with a simple JK flip-flop, a scaled factor $\varepsilon/\sqrt{2}$ is introduced to make sure $|\varepsilon \times h_{j,i}(r_j - \mu_{z_j,i})| \ll \sigma_{z_j,i}^2$. Now the stochastic message updating is expressed as follows:

$$\beta_{j,i}^s = \frac{\varepsilon^s h_{j,i}^s (r_j^s - \mu_{z_j,i}^s)}{(\sigma_{z_j,i}^2)^s} \approx \frac{\varepsilon^s h_{j,i}^s (r_j^s - \mu_{z_j,i}^s)}{s^s h_{j,i}^s (r_j^s - \mu_{z_j,i}^s) + (\sigma_{z_j,i}^2)^s}, \quad (2)$$

where $\varepsilon = 1/32$ is the empirical value.

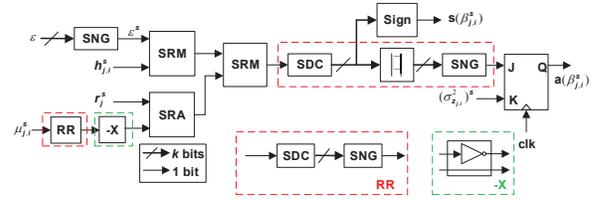


Fig. 3. Architecture for message updating of observation nodes.

For stochastic computing, the gradual loss of randomness for bit streams will result in the *latching* problem. Therefore, re-randomization operation, which estimates the deterministic value and regenerates new uncorrelated stochastic stream, is usually employed. In Eq. (2), the re-randomizations on $\mu_{z_j,i}^s$ and $\varepsilon^s \cdot h_{j,i}^s (r_j^s - \mu_{z_j,i}^s)$ are required. The overall hardware architecture for $\beta_{j,i}^s$ updating is illustrated in Fig. 3, where RR and -X denote the re-randomization unit and bit flipping unit, respectively. 2 -X units and SRA comprise SRS. Stochastic number generator (SNG) in [8] is employed to generate signed stochastic stream. The stochastic to deterministic convertor (SDC) converts signed stochastic stream to deterministic value.

E. Message Updating of Symbol Nodes and Output Decision

In 4-QAM MIMO system, if $|\beta_{j,i}^{(l)}(+1)| \ll 1$, we have $\alpha_{i,j}^{(l)}(+1) \in [-1, 1]$. As a result, $\gamma_i^{(l)}(+1)$ can be efficiently computed by the feedback architecture with only one SRA, shown in Fig. 4. Here T_s denotes the system clock. To reduce the complexity, a look-up table (LUT) is employed to compute $p_{i,j}^{(l)}(+1)$, which will be discussed in Section III in detail.

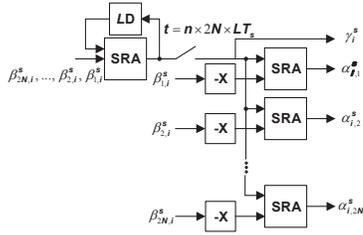


Fig. 4. Architecture for stochastic message updating of symbol nodes.

The final decision is made based on the sign of γ_i after I iterations, which is obtained from the stochastic stream γ_i^s with a controllable up-down counter illustrated in Fig. 5.

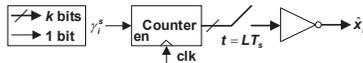


Fig. 5. Architecture for stochastic output decision.

III. HARDWARE ARCHITECTURE

In this section, architecture for the 4-QAM stochastic BP large-scale MIMO detector is proposed. Here, i.i.d. *Rayleigh* fading channel and no channel coding are assumed.

A. Empirical Selection of L

For implementation, the selection of bit stream length ranks the key consideration. First, the normalized deterministic detector with all signals normalized to $[-1, 1]$ is obtained. Second, fixed-point simulation is conducted to determine the fractional width k . Finally, L is determined by $L = 2^k$.

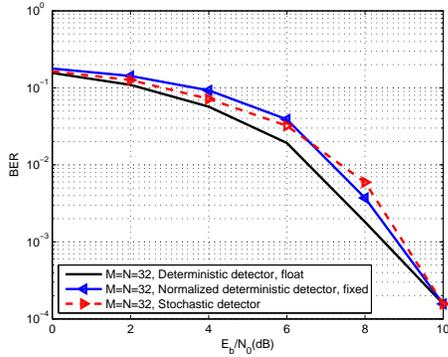


Fig. 6. Performances of deterministic and stochastic detectors (4-QAM).

For 32×32 MIMO system, Fig. 6 shows the BER performances of the float-point deterministic BP detector, fixed-point normalized deterministic BP detector with 1 sign-bit and $k = 12$ data-bits, and stochastic BP detector with $L = 2^{12}$. The iteration numbers are 7, 12, and 12, respectively. According to Fig. 6, to achieve similar performance, fixed-point methods need more iterations than the float-point one. Also, the performance gap between the two fixed-point detectors are

negligible. Therefore, the stochastic BP detector with $L = 2^{12}$ is feasible for hardware implementation.

B. Hardware Architecture

For 4-QAM large-scale MIMO systems, the overall hardware architecture of BP stochastic detector is presented in Fig. 7. It is composed of 3 main modules: message updating module for observation nodes, message updating module for symbol nodes, and decision output module. Details of PE1 to PE6 are given in Fig. 8. PE7 to PE9 shown in Fig. 3 to 5 carry out stochastic operations of observation nodes, symbol nodes, and output decision, respectively. Units within the pink solid bordered rectangle are the same. The scarlet letters at the bottom of PEs denote the concurrent numbers.

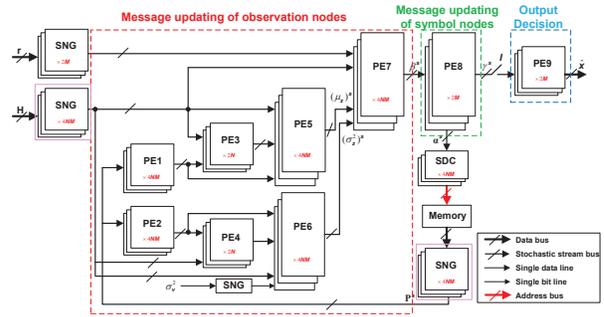


Fig. 7. Overall architecture of stochastic MIMO detector.

In real model of 4-QAM MIMO systems, assume the normalized transmitted signal x equals to $\frac{1}{\sqrt{2}}$ with probability p . The mean and variance of x are given by

$$\mu_x = \frac{1}{\sqrt{2}}(p + p + (-1)), \quad \sigma_x^2 = (p - p^2) + (p - p^2), \quad (3)$$

where $\mu, \sigma_x^2 \in [-1, 1]$. Therefore, Eq. (3) can be implemented by PE1 and PE2 in Fig. 8, where $1/\sqrt{2}$ is replaced by its fixed-point approximation. Square operation is implemented by the structure in the green dotted bordered rectangle, where the register D generates an uncorrelated bit stream with probability $p_{j,i}^s$. $4NM$ PE1's and $4NM$ PE2's are employed to generate the stochastic streams of μ_x and σ_x^2 , respectively.

Since μ_{z_j} and $\sigma_{z_j}^2$ are pre-computed, complexity of updating matrices μ_z and σ_z^2 is reduced from $\mathcal{O}(NM^2)$ to $\mathcal{O}(NM)$. Because $\mu_{z_j}, \sigma_{z_j}^2 \in [-1, 1]$, stochastic implementations of μ_{z_j} and $\sigma_{z_j}^2$ are shown as PE3 and PE4 in Fig. 8. Structures in the blue dotted bordered rectangles perform addition operations. $2N$ PE3's and $2N$ PE4's generate the stochastic streams of vectors μ_z and σ_z^2 , respectively. Stochastic stream of matrix μ_z is obtained by the parallel architecture of $4NM$ PE5's. The similar implementation applies matrix σ_z^2 .

Mentioned in Section II, $p_{i,j}(+1)$ is obtained from an LUT. In initialization, the signed fixed-point number $A \in [-1, 1]$ with 13 bits is employed as address to build up a data memory, where the data $D \in (0, 1)$. The relationship of A and D is:

$$D = (e^{\frac{A}{s/\sqrt{2}}}) / (1 + e^{\frac{A}{s/\sqrt{2}}}), \quad (4)$$

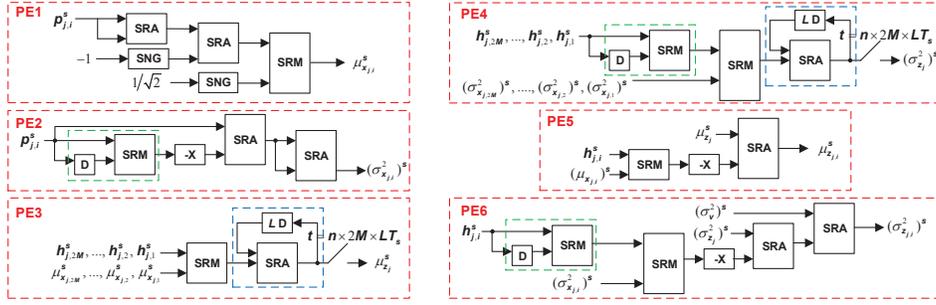


Fig. 8. Hardware architectures of PE1 to PE6.

where $s/\sqrt{2}$ is the scaling factor in Eq. (2). In the iterations, stochastic expression α^s is first converted to deterministic matrix α . Then, we access the data memory of D in parallel by employing the 13 bits of α as address bus.

IV. PERFORMANCE AND COMPLEXITY

A. Numerical Simulation Results

Simulation results of proposed fixed-point normalized deterministic ($I = 7$) and stochastic ($I = 12$) MIMO detectors with different antenna configurations are given in Fig. 9. Stochastic stream length $L = 2^{12}$. 3 configurations are: $M = N = 8, 16,$ and 32 . Seen from Fig. 9, when $M = N$ is small, the performance of stochastic detector is poorer than the deterministic one. However, its performance improves as $M = N$ increases, approaching that of the fixed-point deterministic detector. This matches the theoretical behavior of large-scale MIMO systems.

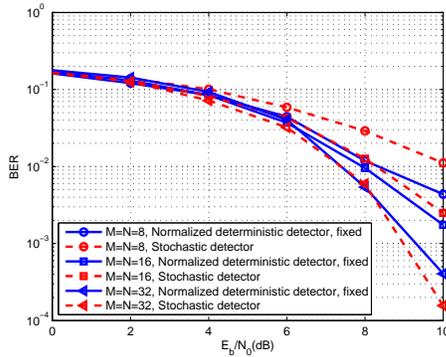


Fig. 9. Performances with different antenna configurations (4-QAM).

B. Complexity and Latency Analysis

Assume the memory access latency is nT_s . According to the operating principle, SDC causes a delay of $(L + \log_2 L)T_s$ in each operation. Seen in Fig. 2, 2 parallel registers in SRA introduce a delay of T_s for each addition. The JK flip-flop leads to a delay of T_s every time. For $M \times N$ 4-QAM MIMO systems, the hardware and latency of proposed stochastic MIMO detector are listed in Table I. Seen from Table I, if M (or N) is constant, the hardware and latency increase linearly

with N (or M). This means the stochastic BP detector is suitable for large-scale MIMO systems.

TABLE I
HARDWARE AND LATENCY OF THE PROPOSED STOCHASTIC DETECTOR.

Modules	CMP		MUX	
	$24NM + 2M + 1$		$72NM + 8N + 4M$	
	CNT	JK	SWT	
	$8NM + 4M$	$4NM$	$4(N + M)$	
D flip-flops				
$4(L + 2 \log_2 L + 20)NM + 2(2L + 5)N + 2(\log_2 L + 2)M + L$				
Gates	XOR		AND	
	$64NM + 10N + 2M$		$208NM + 26N + 10M$	
	NOT		XNOR & OR	
	$20NM + 2M$		$72NM + 8N + 4M$	
Memory	$8K$			
Latency (T_s)	$((2N + 2M + 1)L + 2 \log_2 L + 12)I + L + n$			

V. CONCLUSION

To sum up, the proposed stochastic BP detection works well for large-scale MIMO systems, no matter what antenna configuration is employed. This method is matrix inversion free. Its performance improves with the increase of antenna number, at the cost of linear increase of complexity and latency. Therefore, it is suitable for large-scale MIMO systems. Further work will be directed towards stochastic BP detector with a low latency and high order modulations.

ACKNOWLEDGEMENT

This work is partially supported by NSFC under grant 61501116, International Science & Technology Cooperation Program of China under grant 2014DFA11640, Student Research Training Program of Southeast University, Huawei HIRP Flagship under grant YB201504, Jiangsu Provincial NSF under grant BK20140636, Intel Collaborative Research Institute for MNC, the Fundamental Research Funds for the Central Universities under grants 3204004202, 3204004102, and 3204005101, the Research Fund of National Mobile Communications Research Laboratory, Southeast University under grant 2014B02, and the Project Sponsored by the SRF for the Returned Overseas Chinese Scholars of State Education Ministry.

REFERENCES

- [1] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, no. 3, pp. 1–10, Mar. 2009.
- [2] I. Pefkianakis, L. Suk-Bok, and L. Songwu, "Towards MIMO-aware 802.11n rate adaptation," *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 692–705, Jun. 2013.
- [3] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta *et al.*, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [4] S. S. Tehrani, A. Naderi, and *et. al.*, "Majority-based tracking forecast memories for stochastic LDPC decoding," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4883–4896, Sep. 2010.
- [5] G. Sarkis and W. J. Gross, "Efficient stochastic decoding of non-binary LDPC codes with degree-two variable nodes," *IEEE Commun. Lett.*, vol. 16, no. 3, pp. 389–391, Mar. 2011.
- [6] B. Yuan and K. K. Parhi, "Successive Cancellation Decoding of Polar Codes using Stochastic Computing," in *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, Lisbon, Portugal, May 2015, pp. 3040–3043.
- [7] J. Chen, J. Hu, and S. Li, "Low power digital signal processing scheme via stochastic logic protection," in *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, Seoul, Korea, May 2012, pp. 3077–3080.
- [8] J. Chen, J. Hu, and G. E. Sobelman, "Stochastic MIMO Detector Based on the Markov Chain Monte Carlo Algorithm," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1454–1463, Mar. 2014.
- [9] J. Chen, J. Hu, and G. E. Sobelman, "Stochastic Iterative MIMO Detection System: Algorithm and Hardware Design," *IEEE Trans. Circuits Syst. I*, vol. 62, no. 4, pp. 1205–1214, Apr. 2015.
- [10] S. Haykin, M. Sellathurai, Y. de Jong, and T. Willink, "Turbo-MIMO for wireless communications," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 48–53, Oct. 2004.
- [11] J. Yang, C. Zhang, X. Liang, S. Xu, and X. You, "Improved symbol-based belief propagation detection for large-scale MIMO," in *Proc. IEEE Workshop on Signal Processing Systems (SiPS)*, 2015, accepted.
- [12] S. L. Toral, J. M. Quero, and L. G. Franquelo, "Stochastic pulse coded arithmetic," in *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, Geneva, Switzerland, May 2000, pp. 599–602.