BLAST Architectures

Eduardo Zacarías B.
Signal Processing Laboratory
ezacaria@wooster.hut.fi

Abstract— **Multiple Input Multiple Output (MIMO) systems have been extensively studied in the context of wireless communications, promising both increased capacity and link level reliability. Following the proposal by Foschini at Bell Labs [1], a family of architectures emerged for systems employing multiple antenna arrays at transmit and receive end, collectively known as Bell Labs Layered Space-Time (BLAST) architectures. This paper reviews the essential aspects of three of the best known members of the family, namely D-BLAST, V-BLAST and Turbo-BLAST.**

I. Introduction

Wireless communications systems with multi-antenna arrays have been a field of intensive research on the last years. Space time layered architectures offer a big increase in capacity, promising a linear growth with the size of the antenna array under some circumstances [2]. In 1996, G.J. Foschini proposed a diagonal layered architecture [1] (note that the analysis of Telatar in [2] is dated in 1995 in the Bell Labs internal files), now widely known as D-BLAST, and from which stemmed several derivations, namely V-BLAST [3] and Turbo-BLAST [4].

This paper will review the most important facts of the aforementioned architectures, giving some performance examples found in literature.

II. System model

The system is a discrete time MIMO setup with a single user, N_T antennas at the transmit side and N_R antennas at the receive side. It is also assumed to use FDD and to be frequency flat. Therefore:

$$\mathbf{y} = \mathbf{HWx} + \text{noise} \tag{1}$$

where the equation relates the output of transmit side antenna array $\mathbf{x}$ ($N_b \times 1$, which represents a system with N_b beams) undergoing flat frequency fading for one symbol period, to the output of the receive side antenna array $\mathbf{y}$.

The channel matrix $\mathbf{H}$ is an $N_R \times N_T$ complex matrix, and its elements can be correlated or uncorrelated. The transmit power is restricted and normalized to one. The noise is a WSS complex Gaussian process with i.i.d. vector components and the average noise power on each Rx antenna is N_0. The beamforming matrix is an $N_T \times N_b$ complex matrix. For a single beam system with no beamforming applied, $\mathbf{W}$ is $N_T \times 1$ with elements equal to $1/\sqrt{N_T}$. In the case of the BLAST family without beamforming, $\mathbf{W} = \mathbf{I}_{N_T}/\sqrt{N_T}$ where $\mathbf{I}_n$ is the identity matrix of dimension n, since $N_b = N_T = N_R = n$ and the beams are the canonic basis scaled to satisfy the power constraint.

System model is shown in Figure 1.

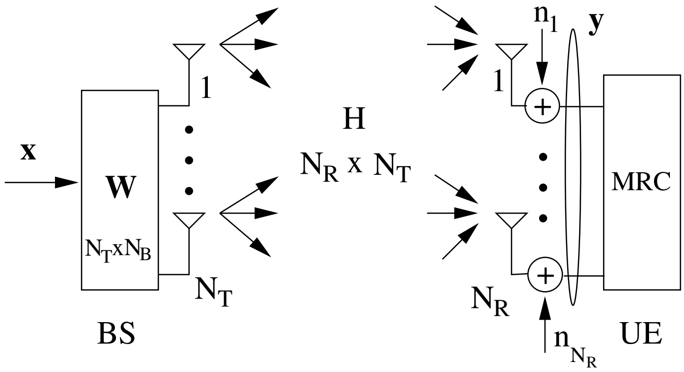

Fig. 1. System model

III. Architectures' description

A. D-BLAST

Originally proposed by Foschini in 1996 [1], this architecture is now considered the reference in performance for MIMO systems, since it can reach capacities near the Shannon limit. However, the complexity is still too high to be practical.

D-BLAST is also described in standard textbooks of communications [5],[6].

1) Encoder: The encoder uses a space time arrangement that corresponds to a diagonal layering. The information bit stream coming from the source is demultiplexed into several substreams (serial to parallel), and each substream is coded separately and mapped to complex symbols. Then the symbols of each substream are dispersed "diagonally" across antennas and time. Figure 2 shows the antenna and instant where symbols associated to each layer are transmitted, for a system with four transmit antennas (figure adapted from [6], fig. 6.10). Note that the layer might have more symbols than the number of transmit antennas, and the frame can be very long.

Unfortunately, given the structure of the decoder, the space time wastage is necessary. This ultimately makes D-BLAST unable to reach the capacity limit, since the wastage is repeated every time a new set of layers are to be transmitted.

Note that since the symbols are spread across antennas, this scheme captures transmit diversity.

2) Decoder: The decoder proceeds to decode one layer after another. They first symbol of the layer is guaranteed to be detected without errors, since it is transmitted alone (but the system pays the space-time wastage). After that, the next symbol on the layer is demodulated and detected, facing one interferer. The next will face two interferers, and so on. Once all the symbols of the first layer are demodulated, the substream associated to the layer can be decoded. This

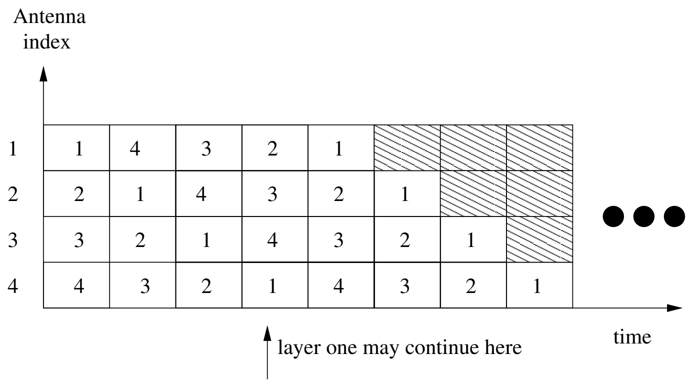

Fig. 2. D-BLAST: diagonal layering. Numbers in blocks represent the layer that can transmit its symbols at that antenna and symbol period. Filled blocks represent space time wastage.

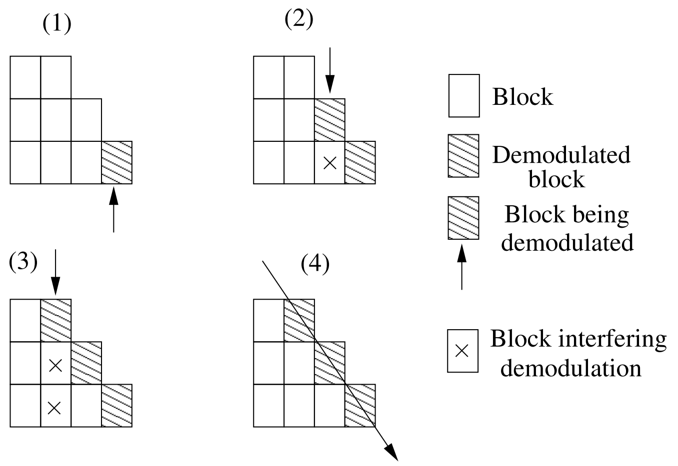

Fig. 3. D-BLAST decoder (adapted from [6])

decoding should be error free, otherwise the whole process would suffer from error propagation. In order to ensure the absence of errors, the channel code associated to the stream must be powerful and the stream must be long. Once the layer is decoded, it can be subtracted, thus "peeling it" and "exposing" the next one, for which the aforementioned process is repeated. This process is illustrated on figure 3, where the layer consists of only three symbols.

B. V-BLAST

The Vertical BLAST or V-BLAST architecture [3] is a simplified version of D-BLAST, that tries to reduce its computational complexity. But in doing so, the transmit diversity is lost.

1) Encoder: As in D-BLAST, the information bit stream is separated in substreams, and each can undergo its own channel coder. However, the layering is horizontal, meaning that all the symbols of a certain stream are transmitted through the same antenna (one stream per antenna).

This eliminates the space time wastage, but loses the transmit diversity, since each stream is "tied" to its antenna.

This scheme is also known as vector modulation [7], since it involves just a serial to parallel operation. At a certain symbol instant, the output of the transmission antenna array is a vector $[s_{1k}, \ldots, s_{N_T k}]^T$ where s_{ik} represents the k-th symbol of the $i-th$ stream.

2) Decoder: The decoder needs to demodulate the symbols on the received vector. If channel coding is used, then the demodulated symbols need to be buffered until the whole block can be decoded. Otherwise, the demodulation can be done immediately. Several decoders are possible for this architecture.

The Maximum Likelihood (ML) decoder solves:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \left\| y - \sqrt{\frac{E_s}{N_T}} \mathbf{H}\mathbf{s} \right\|_F^2 \tag{2}$$

Nearly optimal decoders with reduced complexity based on the sphere decoder principle have been proposed [8], so that implementing Maximum Likelihood decoding becomes practical (otherwise exhaustive search is too complex for higher order constellations).

Linear decoders are possible as well. Well known decoders are the Zero Forcing (ZF) and Minimum Mean Square Error (MMSE) decoder. Both apply slicing (over the appropriate symbol constellation) to the post-processed vector $\mathbf{Gy}$. These are given by:

$$\mathbf{G}_{zf} = \sqrt{\frac{N_T}{E_s}} \mathbf{H}^{\dagger} \tag{3}$$

where † denotes Moore-Penrose pseudo inverse and H denotes hermitian transpose.

$$\mathbf{G}_{mmse} = \sqrt{\frac{N_T}{E_s}} \left(\mathbf{H}^H \mathbf{H} + \frac{N_T}{\rho} \mathbf{I}_{N_T} \right)^{-1} \mathbf{H}^H \tag{4}$$

The ZF decoder attempts to invert the channel, but amplifies the noise in the process. The MMSE decoder attempts to both invert the channel, but keep the noise amplification controlled, in a MMSE sense. More details can be found in [6], ch. 7.

One iterative strategy is represented by the Ordered Serial Interference Cancellation (OSIC) decoder, which estimates the substream with strongest signal to noise ratio, demodulates it, and then subtracts it from the output of the array. After that, the next substream with highest SNR is decoded and subtracted, and so on. This is explained with more detail in [3] and also briefly in [6], ch. 7.

C. Turbo-BLAST

Sellathurai and Haykin proposed another BLAST architecture [4], based on th Turbo principle, which was afterwards generalized by the Threaded Space-Time Architecture (TST) by ElGamal and Hammons [9].

1) Encoder: A Random Layered Space-Time (RLST) coding scheme is employed before transmission. The information bit stream is also demultiplexed and substreams obtained thus are independently encoded with the same block FEC, as in D-BLAST. Then the substreams are *bit interleaved* in space using a diagonal interleaver ("random space time interleaver" described in [4]). Finally, the "mixed" streams are mapped to symbols and transmitted. Each symbol can have bits coming from more than one stream, and therefore a symbol error

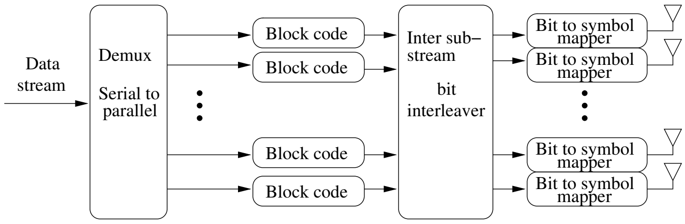

Fig. 4. Turbo-BLAST encoder.

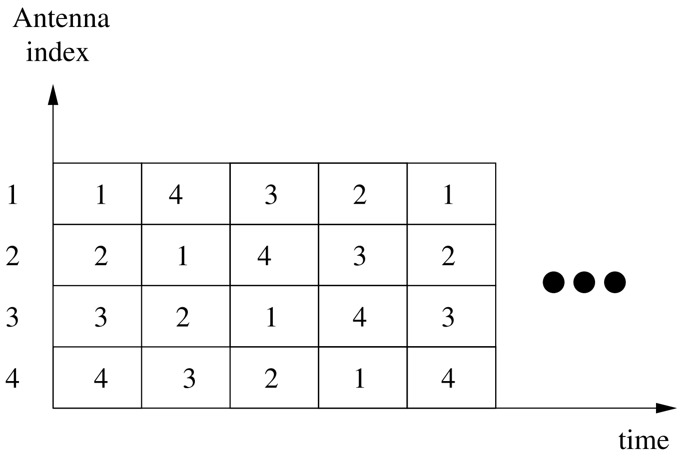

Fig. 5. Turbo-BLAST diagonal bit interleaver.

spreads the bit errors across streams, thus making the error correction easier for the block decoders. The encoder is shown in figure 4 (figure adapted from [4]).

The inter-stream bit interleaver is similar to the diagonal scheme in D-BLAST, but it has no space time wastage, as shown in figure 5.

2) Decoder: Unfortunately, decoding the interweaved streams is very expensive computationally, being exponential in the number of substreams (i.e. the number of Tx antennas), constellation and block sizes. An iterative suboptimal algorithm is proposed in [4], based on decoding of serially concatenated turbo codes. The idea is based on the interpretation of the Turbo-BLAST encoder as a group of block codes ("outer coder") connected with an "inner coder" through parallel interleavers. Thus, the inner decoder is meant to cope with Inter Symbol Interference (ISI) coming from multipath fading in the channel, and the outer decoder aims to correct symbol errors occurred during the transmission over the first channel path. Both decoders output soft decisions, which are ultimately sent to hard limiters after the required iterations. A block diagram scheme is shown in figure 6 for a system with 4 receive antennas (figure adapted from fig. 6.38 in [5]).

IV. Capacity aspects

For D-BLAST, symbols in a layer are extracted from successive frames through an MMSE receiver, and then assembled into a single stream for optimal detection (the layer spans several frames). This procedure is repeated for all the layers. With this scheme, *if space time wastage is neglected and the frames are suitably long*, the D-BLAST scheme is able to reach the channel capacity of the fading channel (see [6], sec.

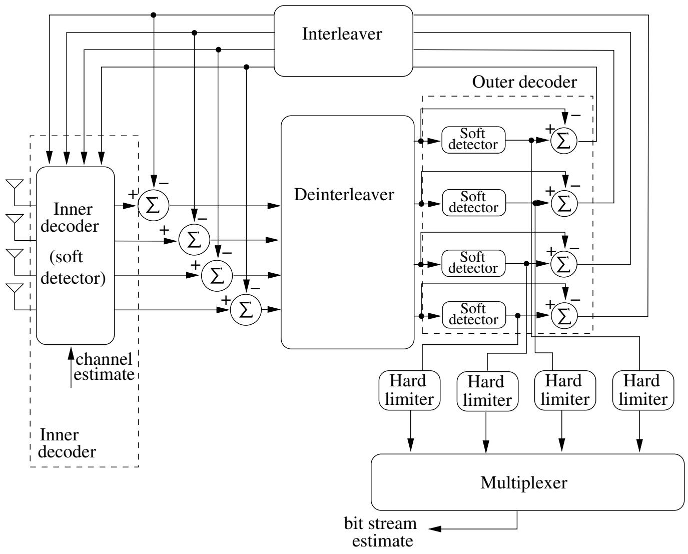

Fig. 6. Turbo-BLAST iterative detection and decoding scheme.

12.4.1). However, the frames can't be arbitrarily long, due to delay, memory and complexity restrictions, and therefore a nearly optimal capacity in practice would require a very complex and expensive system.

Information rates for V-BLAST are reduced since the scheme does not exploit transmit diversity. Therefore it is a suboptimal scheme.

Closed form expression for the capacity of the Turbo-BLAST architecture does not seem to exist yet, but the authors claim to obtain big performance improvements in Bit Error Rate (BER), when compared to V-BLAST. They also claim that Turbo-BLAST can cope with the asymetric $N_T > N_R$ case. This is shown in figures 10 and 10.

V. Performance

A. Complexity issues

The most expensive (computationally) member of the family is D-BLAST. After it, but probably depending on the choice of decoders, comes Turbo-BLAST and finally the simplest is V-BLAST. A more thorough complexity comparison was not found in the reviewed literature.

B. Fading environment requirements

The performance of the BLAST architectures depends on three key factors. Quoting [5]:

1) The system operates in a rich Rayleigh scattering environment.
2) Appropriate coding structures are used.
3) Error-free decisions are available in the interference-cancellation schemes. This condition assumes the combined use of arbitrarily long (and therefore powerful) FEC codes and perfect decoding.

Therefore, fading correlation and limited code lengths will diminish the ability of BLAST architectures to provide the promised high data rates. See for example [10].

BLAST architectures described here do not assume any Channel Side Information (CSI) at the transmitter side. In Time Division Duplexing systems this information comes from uplink measurements and it can be exploited to boost the performance to nearly optimal levels (see for example [6], ch. 8). In Frequency Division Duplexing (FDD) systems, the CSI must come from the closed loop feedback channel. Techniques exploiting the feedback information are the subject of a future talk in this course.

C. Switching between V-BLAST and OSTBC

The information coming from the closed loop feedback channel can be exploited to adapt the transmission to the channel conditions. For example, WCDMA's High Speed Downlink Packet Access (HSDPA) [11] makes extensive use of Adaptive Modulation and Coding (AMC).

Another example is to switch to an Orthogonal Space Time Block Code (OSTBC) when the channel conditions are not suitable for V-BLAST. This has been studied in [12], where the authors derive the probability of having an instantaneous channel that is better suited for uncoded V-BLAST, rather than for a Space Time Code exploiting transmit diversity. They also derive a metric that must be fed back in order to make the decision, which is based on the Euclidean distance of the "received constellations" (this is, the symbol constellations for V-BLAST and the STC after the channel). Note that the OSTBC and the V-BLAST scheme must have the same data rate, in order to make a fair comparison.

If the hypothetical constellations for V-BLAST and STC have minimum Euclidean distances d_{BLAST} and d_{STC} respectively (e.g. QPSK and 16QAM for a 2x2 array), then V-BLAST is preferred whenever the channel satisfies:

$$\kappa \leq \frac{d_{BLAST}}{d_{STC}} \tag{5}$$

where $\kappa := \frac{||\mathbf{H}||_F}{\lambda_{\min}}$ and $\lambda_{\min}$ is the smallest singular value of the channel matrix $\mathbf{H}$. This condition is valid only for Rayleigh fading channels.

A performance example is taken from [12] and shown in figure 7. As expected, the "composite" BER curve is better than both the "individual" curves.

The probability of selecting uncoded V-BLAST instead of OSTBC can be measured experimentally. It is found that for uncorrelated channels, it agrees with the theoretical expression, while for correlated channels the situation is completely different. This is shown in table I, where the constellation is QPSK for V-BLAST and appropriate QAM for the uncoded V-BLAST.

D. V-BLAST, beamforming and waterfilling

When there is some CSI available, the optimal strategy is to use waterfilling over the parallel sub-channels. This is absolutely necessary, as the smallest eigenmode of the channel is very weak most of the time. As the number of antennas increases, the number of dominant (usable) beams increases as well. For example, in 2x2 system in general the second beam can not be used often. In a 4x2 system, it is more "reasonable" to use the first two eigenbeams.

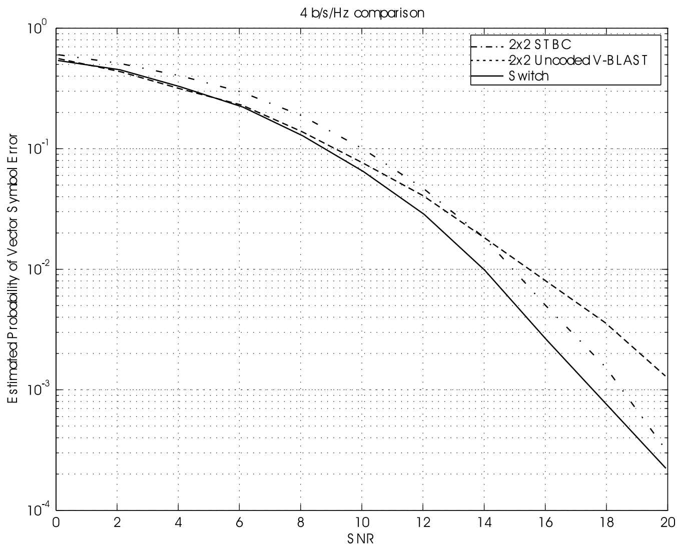

Fig. 7. Switching between V-BLAST and STC based on Euclidean distance.

Rate	N	P_{HP}	Uncorrelated	3GPP C2	3GPP C4
4	2	0.216	0.215	0.002	0.056
8	2	0.687	0.688	0.038	0.354
8	4	0.485	0.484	0.000	0.028

TABLE I
MEASURED PROBABILITY FOR V-BLAST BEING PREFERRED OVER OSTBC FOR SQUARE SYSTEMS, COMPARED TO EXPRESSION OBTAINED BY HEATH AND PAULRAJ (P_{HP}) [12]. RATE IS GIVEN IN BITS PER CHANNEL USE.

Therefore, when exploiting the CSI, the "practical rank" of the channel matrix indicates how many parallel substreams the channel can support at a given time. For more details, see [7], sec. 12.5.

As an example, the received symbols for uncoded V-BLAST in a 2x2 system with MMSE decoder are shown in figure 8. The instantaneous eigenbeams (left singular vectors) of the channel matrix have been used (i.e., full CSI). Statistics of the singular values of a strongly correlated channel are shown in figure 9.

E. Diversity order and multiplexing gain

The *diversity order* is defined as the asymptotic rate at which the Frame Error Rate (FER) curve falls as a function of the SNR in a log-log plot [5]:

$$d_0 = -\lim_{\rho \to \infty} \left\{ \frac{\log(FER(\rho))}{\log(\rho)} \right\} \tag{6}$$

The maximal diversity order for the MIMO system is $N_T N_R$.

The *multiplexing gain* is defined [5] as the asymptotic increase of the ergodic capacity as function of SNR, also in log-log plot:

$$r = \lim_{\rho \to \infty} \frac{C(\rho)}{\log(\rho)} \tag{7}$$

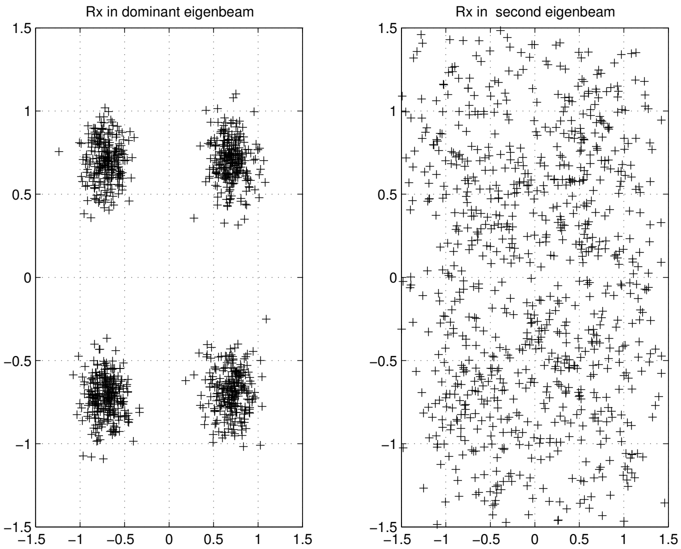

Fig. 8. Received symbols (MMSE) in 2x2 V-BLAST system with full CSI and fixed modulation scheme (no waterfilling).

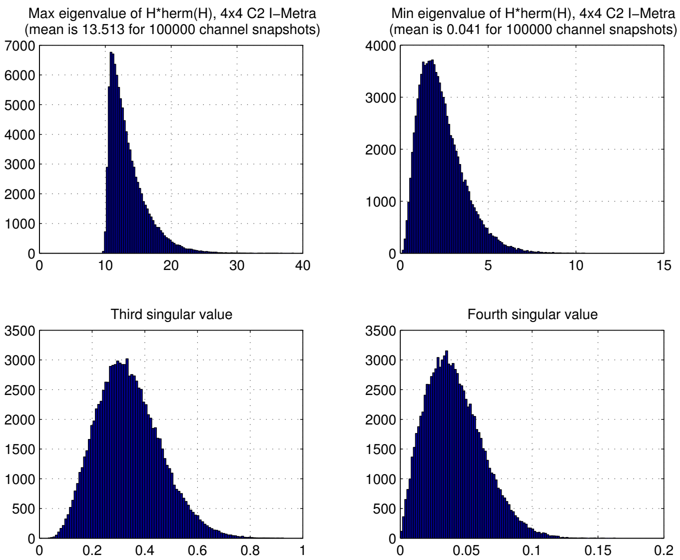

Fig. 9. Statistical characterization of singular values of 4x4 strongly correlated channel.

and its maximum is $\min(N_T, N_R)$.

Haykin argues that there is a fundamental trade off between diversity order and multiplexing gain for BLAST architectures, since they are designed to maximize the ergodic capacity, as opposed to the Space-Time Codes, which intend to maximize the diversity order. This is equivalent to the trade off between data rate and link reliability.

In general, the STBC are capacity suboptimal (see [6], sec. 12.4). D-BLAST is able to reach maximal diversity order on ideal conditions and some OSTBC are able to reach capacity (e.g. the Alamouti code). V-BLAST with OSIC has diversity order $N_R - N_T + 1$ [5].

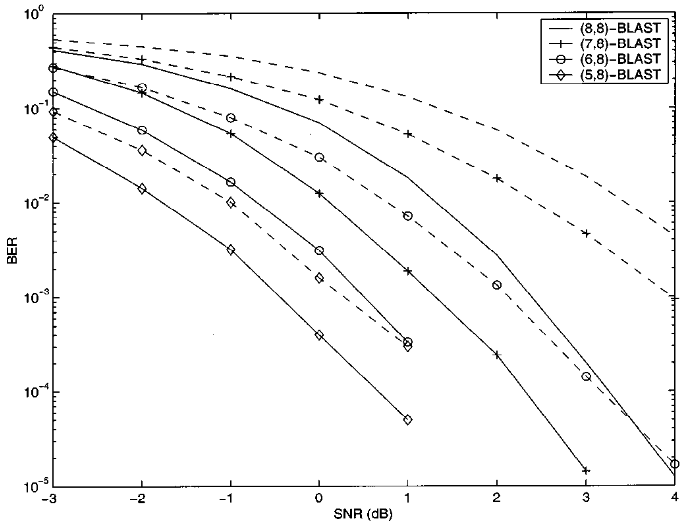

Fig. 10. Experimental Turbo-BLAST versus V-BLAST results.

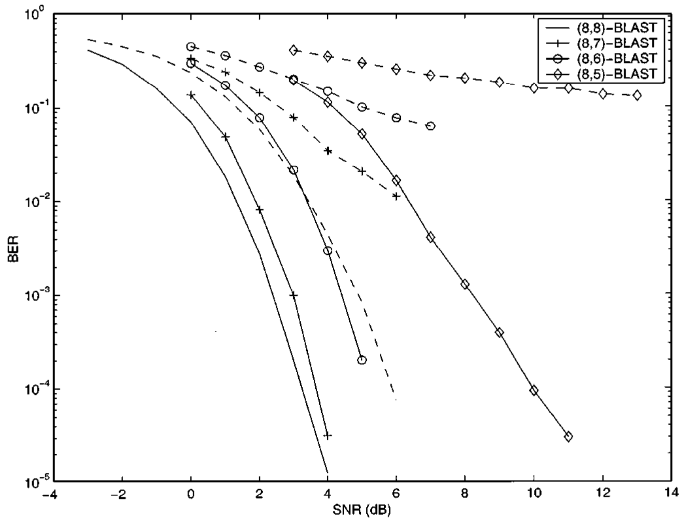

Fig. 11. Experimental Turbo-BLAST versus V-BLAST results when $N_T > N_R$.

F. Comparisons between architectures

Following comparisons are found in [4] from real Laboratory experiments. In figure 10 Turbo-BLAST is compared to V-BLAST for several configurations with $N_R = 8$, while in figure 11, Turbo-BLAST is compared to V-BLAST when $N_T > N_R$.

VI. Conclusions

A brief overview of the main characteristics of three BLAST architectures has been presented: D-BLAST, V-BLAST and Turbo-BLAST. The architectures' encoding and decoding strategies are described, and considerations about performance are given. Some results are presented as to illustrate particular details and to compare the performance between architectures.

References

[1] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," Bell Labs, Technichal Journal 2, 1996, appeared in Volume 1, number 2, pp 41-59.

[2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November/december 1999.

[3] P. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *URSI International Symposium on Signals, Systems, and Electronics, 1998. ISSSE 98.*, 1998.

[4] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Transactions on Signal Processing*, 2002.

[5] S. Haykin and M. Moher, *Modern Wireless Communications (International Edition)*. Pearson Education, March 2004.

[6] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, May 2003.

[7] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-antenna transceiver techniques for 3G and beyond*. John Wiley and Sons, January 2003.

[8] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639 – 1642, July 1999.

[9] H. El Gamal and A. Hammons Jr., "New approach to layered space-time transmitter receiver design," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2321 – 2334, September 2001.

[10] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502 – 513, March 2000.

[11] T. 3GPP, "TR 25.858: Physical layer aspects of utra high speed downlink packet access," 3GPP, Tech. Rep., 2002.

[12] A. Paulraj and R. J. Heath, "Characterization of MIMO channels for spatial multiplexing systems," *IEEE International Conference on Communications*, vol. 2, no. 11-14, pp. 591 – 595, June 2001.