IEEE P802.11  
Wireless LANs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Proposed IEEE 802.11 AIML TIG Technical Report Text for the CSI Compression Use Case | | | | |
| Date: 2023-05-30 | | | | |
| Author(s): | | | | |
| Name | Affiliation | Address | Phone | Email |
| Zinan Lin | InterDigital | 111 West 33rd Street  New York, NY 10120  USA |  | [Zinan.lin@gmail.com](mailto:Zinan.lin@gmail.com) |
| Rui Yang | InterDigital |  |  |
| Xiaofei Wang | InterDigital Inc. | +1-607-592-2727 | [Xiaofei.wang@interdigital.com](mailto:Xiaofei.wang@interdigital.com) |
| Liangxiao Xin | Zeku |  |  |  |
| Ziming He | Samsung Cambridge Solution Centre |  |  | [Ziming.he@samsung.com](mailto:Ziming.he@samsung.com) |
| Eunsung Jeon | Samsung | 1-1, Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do 18448 Korea | +82-10-2317-5808 | [eunsung.jeon@samsung.com](C:\\DOCUMENTS\\EFSS\\TEMPFILES\\CHECKOUT\\DATA\\D_829b54882_56_\\eunsung.jeon@samsung.com) |

Abstract

This document contains the proposed technical report text of the IEEE 802.11 AIML TIG, especially for the CSI compression use case.

Revision history:

r0: Copied from 802.11-23/0906r1 and add the dual CSI feedback use case in Section 2.1.2.

r1: Copied from 802.11-23/0906r2 and add the dual CSI feedback use case in Section 2.1.1.

# Table of Contents

1. **Introduction**
   1. Terminologies
   2. Background information
2. **AIML Use cases for IEEE 802.11**

Note: use cases potentially can be organized into different categories

Note: use cases potentially can identify KPIs

## Use case 1: CSI feedback compression

### Use case description: system throughput improvement

In 802.11ax [1] and the draft of 802.11be [2], the AP initiates the sounding sequence by transmitting the NDPA frame followed by a NDP which is used for the generation of V matrix at the beamformee. Upon the receipt of the NDP from the beamformer, the beamformee applies a compression scheme (i.e., Givens rotations) on the V matrix and feeds back the angels in the beamforming report frame.

It is indicated in [3] that higher number of spatial streams has been an inevitable trend in WiFi for more than a decade. The preliminary results [3] [4] [5] show that MIMO with a large number transmitter antennas and a large number of spatial streams (e.g., 16 spatial streams) offer remarkable system performance gains on both SU-MIMO and MU-MIMO cases. Multi AP (MAP) may be one potential feature in the next 802.11 generation, e.g. UHR [6] - [9] . Large number of spatial streams combined with MAP feature may further increase the sounding feedback airtime overhead if coordination between APs (e.g., joint transmission/reception, coordinated beamforming) is applied. Large amount of overhead or prolonged sounding procedures may negatively impact the latency and limit the system performance. Therefore, there is a need to reduce the CSI overhead especially when the number of transmitter antennas goes higher or multiple APs perform joint or coordinated transmission.

Some studies (e.g., [10] [11] [12] [13] [18]) have shown that AI/ML can efficiently reduce the CSI feedback and improve the system throughput. For example, motivated by the nature that the CSI may fall into different clusters due to the channel similarity of nearby STAs, iFOR algorithm [10] applies the unsupervised learning, K-means, to the CSI compression to classify the angle vectors which are derived from V matrix. Simulation results show that for an 8x2 SU-MIMO, iFOR uses around 8% of the number of bits required by the existing feedback mechanism (802.11ax) and boosts the system throughput by up to 52%. In [18], the dual CSI compression combines the codebook and Givens rotation. It uses the fact that any untiary matrix can be decomposed by multiplication of two unitary matrices, i.e., **V** = **V**1 **V**2. By using this, the dual CSI compression decomposes a large size CSI into a subband CSI (**V**1) and a subcarrier CSI (**V**2), giving lower feedback rate for the slow-varying subband CSI (**V**1) using the codebook while allocating higher feeback rate for the frequency-selective subcarrier CSI (**V**2) using the Givens rotaion. In order to improve the reliability for subband CSI (**V**1), the K-means algorithm is exploited for the codebook generation. The simulation results show that for 8x2 SU-MIMO, the AI/ML aided dual CSI compression can reduce the feedback overhead more than 50% compared with the conventional scheme. In addition, the throughput improvement from the reduced overhead is about 20%. In [11] , another unsupervised learning, Deep Neural Network Autoencoder (DNN-AE) is applied to CSI angle vectors and further compresses the derived angles (LB-SciFi) by leveraging the compression capability of DNNs. Experimental results show that LB-SciFi reduces the feedback overhead by 73% and increases the network throughput by 69% on average.

This use case proposes to apply AI/ML technique to CSI feedback schemes to reduce the CSI overhead with minimum loss of PER performance, and eventually increase the system throughput.

### Use case description: compression complexity reduction

On the other hand, a large number transmitter antennas and a large number of spatial streams will significantly increase the computational complexity of the CSI feedback compression. For example, the complexity comes from the Givens rotations in the V matrix compression. A highly complex compression scheme may negatively impact the processing time (introduces latency) and the power consumption at non-AP STAs [20].

A study in [20] has shown that AI/ML can reduce the computational complexity of the CSI feedback compression by nearly 60% without degrading the system throughput. The idea is to compress V matrix directly by avoiding the computation in the Givens rotations, using a low complexity Autoencoder.

This use case proposes to apply AI/ML technique to CSI feedback schemes to reduce the computational complexity of the feedback compression without loss of the system throughput.

### KPIs

KPIs considered in this use case are proposed as follows:

1. Number of feedback bits per subcarrier group
2. Achieved PER
   1. Both SU-MIMO and MU-MIMO cases need to be considered
3. Additional AIML overhead compared with compression saving
   1. One example is the ratio between the number of additional bits required by AIML process (including data used for model training/inference [14] , the model parameters, the additional signaling) and the number of bits saved by the CSI feedback scheme. In this example, if the data used for model training that is performed by the AP fully relies on the legacy CSI report, then the additional AIML overhead used for model training/inference may be negligible.
4. Computation complexity/Latency:
   1. Additional delay or computation is introduced by AIML processing.

Evaluation methodology needs to be established.

### Requirements

1. Backward compatibility with legacy 802.11
   1. Support backward compatibility and coexistence with legacy 802.11 CSI report schemes
2. Performance should follow the guidance below:
   1. **CSI airtime reduction**: achieve airtime reduction of CSI feedback over 802.11be for a given Nr x Nc MIMO, where Nr is the number of rows in the compressed beamforming feedback matrix, Nc is the number of columns in the compressed beamforming feedback matrix.
   2. **Additional AIML overhead**: minimize the additional overhead used for AIML process. Additional AIML overhead may include the data used for AIML model training/inference [14], the model parameters and additional signaling. The data used for AIML model training/inference [14] can reuse the legacy CSI report data.
   3. **Packet Error rate (PER)**: guarantee minimum SNR loss compared with 802.11be to achieve the target PER (e.g., 1% and/or 10%) at a given MCS in all types of channels [15] .
   4. **Computation complexity/Latency**: minimize the additional computation complexity or latency required by AIML process
   5. **Additional storage for AIML:** minimize the additional storage used for AIML process. Additional AIML storage may include the storage required for AIML model training data and the model parameters. For example, the STAs performing training may need storage for the training data or/and the model parameters, the STAs performing inference needs storage for the model parameters [17]. The model parameters could be the codebook used in the K-means based schemes (e.g., [10], [17], [18]), or the neural network coefficients in the autoencoder based schemes (e.g., [11], [20]), or both the codebook and neural network coefficients in the autoencoder based scheme in [19].

### Technical Feasibility Analysis

### Standard Impart

The standard impact may include:

* Additional signaling (e.g., between AP and non-AP STAs) required by AIML process.

### Technical feasibility

The following metrics will be studied:

1. **Backward compatibility**: The STAs that supports AIML enabled CSI feedback compression shall support the legacy 802.11 CSI report scheme. This compatibility is expected to be supported since AIML capable STAs are expected to support legacy CSI report scheme.
2. **Data availability and accessibility:** There are some STAs that are able to use the data to perform AIML model training and/or inference[14] . The data used for model training and/or inference shall be accessible for these STAs.
   * + - AP/edge computing based AIML: Data may be collected from non-AP STAs. The legacy 802.11 CSI reports may be used as training data.
       - Device computing based AIML: Data should be available at all STAs that support AIML process.
3. **Hardware/software capability**: The STAs that use AIML to generate the AIML enabled CSI feedback compression shall have the hardware and software capability to support AIML algorithm(s).
   * + - AP/edge computing based AIML [16] : Extra data and model (e.g., model parameters) exchange may be required to support AP/edge computing based AIML. However, computation is not expected to be located at AP or edge computing resources for which higher computation capabilities is expected.
       - Device computing based AIML [16] : STAs that support AIML may be required to have extra computation capability. Extra data and model (e.g., model parameters) exchange between STAs may also be required to support device computing based AIML.
   1. Use case 2
   2. Use case N
4. **Summary**
5. **References**
6. IEEE 802.11-REVme D2.0, October 2022
7. IEEE P802.11be D2.2, October 2022
8. 802.11-18/0818r3, 16 Spatial Stream Support in Next Generation WLAN
9. 802.11-20/1877r1, 16 Spatial Stream Support
10. 802.11-20/1535r66, Compendium of straw polls and potential changes to the Specification Framework Document Part 2
11. 802.11-22/1515, A candidate feature: Multi-AP
12. 802.11-22/1394, Virtual BSS And Multi AP Transmissions
13. 802.11-22/1516 Considerations on Multi-AP Coordination
14. 802.11-22/1512 Multi-AP Coordination for UHR
15. M. Deshmukh, Z. Lin, H. Lou, M. Kamel, R. Yang, I. Güvenç, “Intelligent Feedback Overhead Reduction (iFOR) in Wi-Fi 7 and Beyond,” in Proceedings of 2022 VTC-Spring
16. P. K. Sangdeh, H. Pirayesh, A. Mobiny, H. Zeng, “LB-SciFi: Online Learning-Based Channel Feedback for MU-MIMO in Wireless LANs, ” in Proceedings of 2020 IEEE 28th ICNP
17. P. K. Sangdeh and H. Zeng, “DeepMux: Deep-Learning-Based Channel Sounding and Resource Allocation for IEEE 802.11ax,” IEEE Journal on Selected Areas in Communications, Vol. 39, No. 9, Aug. 2021
18. S. Szott, K. Kosek-Szott, P. Gawłoicz, J. T. Gómez, B. Bellalta, A. Zubow, F. Dressler, “WiFi Meets ML: A Survey on Improving IEEE 802.11 Performance with Machine Learning,” IEEE Communication Surveys & Tutorials, Vol.24, Issue 3, Juen 2022
19. 802.11-22/0723r1, Further discussion on next generation WLAN
20. 802.11-19/0719r1, IEEE 802.11be Channel Model Document
21. 802.11-22/1443r0, Wi-Fi Meets ML: Re-thinking Next-generation Wi-Fi Networks
22. 802.11-23/0275r2, Improved AIML Enabled Index Based Beamforming CSI Feedback Schemes
23. 802.11-23/0280r0, ML aided Dual CSI Feedback for Next Generation WLANs
24. 802.11-23/0290r2, Study on AI CSI Compression
25. 802.11-23/0755r0, AIML Assisted Complexity Reduction For Beamforming CSI Feedback Using Autoencoder