



AI based Pilot Contamination Analysis for 5G MIMO based on Multi Antenna Routing Networks and Multi-User Pilot Scheduling

Dr. Nitin Sherje

Associate Professor and Head of Department,
Smt. Kashibai Navale College Of Engineering, Pune, India
npsherje@sinhgad.edu

Abstract:

Pilot contamination, a type of inter-cell interference, limits performance of large multi-input multi-output (MIMO) antenna systems. A drawback that results in ineffective bandwidth use is the burden of pilots who must estimate the channel regularly due to the acquisition of channel state data for channel estimation. Thus, there is a trade-off between spectral efficiency (SE) as well as quantity of pilots needed to evaluate channel. This research proposes novel technique in pilot contamination analysis (PCA) for 5G network based on MIMO by multi antenna routing system. The main aim is to detect pilot contamination and enhance spectral efficiency of the network. Here pilot contamination is detected using multi-user pilot scheduling with convolutional adversarial training model. As a result, a security breach occurs when crucial information slips to Eve during downlink transmission. Ability of the legitimate user to maintain secrecy can be greatly enhanced by knowing of an active eavesdropper. We also analyse the likelihood of detection, the likelihood of a false alarm, and the likelihood of a detection error. Simulation results show that suggested strategy to find PCA is effective. The proposed technique attained SINR of 72%, spectral efficiency of 85%, normalized MSE of 73%, PCA detection accuracy of 95%.

Keywords: MIMO, pilot contamination analysis, 5G network, multi antenna routing system, pilot scheduling.

1. Introduction:

One of the crucial technologies for the next 5G networks is MIMO. A cellular base station (BS) in a 5G network is expected to have a very wide antenna array, such as hundreds or more antennas, which will greatly increase the transmission rate compared to traditional MIMO systems. Huge MIMO has drawn a lot of attention from researchers recently despite the fact that MIMO is a well-researched wireless communications paradigm. This is because massive MIMO needs novel solutions to address fresh design issues (see, for example, and the references therein). Pilot contamination (PC) is one of the key problems in huge MIMO systems [1]. CSI between BS and different users must be frequently calculated using uplink pilot broadcasts due to the huge number of antennas at BS and relatively short channel coherence time. With assumption of channel reciprocity, the BS uses these CSI calculations for downlink data transmissions. However, users in adjacent cells might exchange pilots because there are so few orthogonal pilot sequences (for instance, 42 [2]). These pilots' interference results in inaccurate channel estimates at the BS, which negatively affects system

performance. Massive MIMO, also known as large-scale MIMO, is superior to regular MIMO technology in terms of both architecture and total performance [3]. One is that a significantly larger antenna array is used in Massive MIMO instead of the several antennas used in regular MIMO. Second, and this is a huge innovation as well as advancement for conventional MIMO method, Massive MIMO method uses a smaller hardware device with reduced power consumption to minimize complexity of hardware. Additionally, the Massive MIMO system substitutes the classic MIMO system's joint use of TDD as well as Frequency Division Duplex (FDD) systems with Time Division Duplexing (TDD) [4]. Three things are the benefits of Massive MIMO. First, numerous users can be served simultaneously by the cellular network; second, the system's data transmission rate can be considerably increased; and third, spatial diversity and multiplexing are effectively used. In conclusion, large-scale MIMO method maximises all of the performance of conventional MIMO method in addition to inheriting it. However, the base communication station is unable to properly discriminate between distinct signals when using the same orthogonal





pilot sequence in several locations [5]. Physical layer (PHY) security has become the subject of research in recent years because it offers an alternative to traditional cryptographic algorithms at higher network layers. Two major methods for achieving security for wireless networks are secret key generation from channel randomness as well as secrecy capacity. Early PHY security efforts concentrated on an data theoretic technique to block access to private messages to Eve. Due to its enormous capacity, MaMIMO systems operate in TDD mode as well as strong candidates for 5G wireless communications technology. Using the predicted channel from the reverse training, the BS creates the precoder for the downlink broadcast in the MaMIMO system. Due to precoder architecture for downlink transmission, which reduces data leaking towards Eve, passive eavesdropping is not effective. BS broadcasts artificial noise (AN) in the null-space of the channel matrix of LUs to further increase secrecy capacity. The NOMA situation is not covered by the PCA countermeasures now in place. The majority of PCA detection techniques now in use consider that pilot signals from various users are divided into distinct orthogonal resource blocks, where they are identified by their orthogonality or the difference in timing between two pilot signals [6].

Contribution of this research is as follows:

1. To propose novel method in pilot contamination analysis (PCA) for 5G network based on MIMO by multi antenna routing system
2. To detect pilot contamination using multi-user pilot scheduling with convolutional adversarial training model
3. The experimental analysis has been carried out based on number of cells.

2. Review of literature:

A acceptable pre-coding system, improved channel estimation, and the use of a sensible pilot allocation strategy are the three major ways to lessen pilot contamination [7]. For MaMIMO systems, active eavesdropping is more efficient than passive attack. The analysis in [8] examines how PCA affects the potential for secrecy. To guide a portion of the beam towards Eve during PCA, Eve transmits a pilot sequence of LU to BS in the pilot stage. To prevent private communications from Eve, PCA must be detected. The phase shift keying (PSK) symbols in the pilot stage should be modified for forward as well as reverse channel estimates, according to [9]. The energy ratio detector in [10] makes use of uplink and downlink channels' channel norms,

which necessitates the estimation of these channels during a pilot period. Self-contamination as well as usage of a minimum description length (MDL) were introduced by PCA detection approach in [11] as well as references therein. In low SNR regime, MDL-based detector experiences performance decrease. In [12], a random symbols-based PCA detector shows improved performance at low SNR. It should be noted that the more advanced procedures in [12] use second order statistics. PCA was initially suggested in [13]. Later, [14] provided a summary of PCA for massive MIMO. Random pilots as well as specialised beamforming are the two basic techniques that can be employed to detect PCA, according to [15]. A better version that takes the case of three or more observations into account was offered in [16], and random PSK pilot was initially proposed in [17]. In essence, this plan consists of two steps. First, a series of randomly selected PSK pilot symbols—chosen from an N-PSK constellation independently—are transmitted by the legitimate user (LU). Secondly, after getting these PSK symbols, BS determines whether there are active attackers over this time period by estimating phase change between two symbols. Based on random frequency shifts, other random pilot approach [18] is used. The ground-breaking research in [19] examined PCA for BS with multiple antennas. The random pilot sequence was employed in work in [20] for PCA detection. A variety of Neyman-Pearson-based PCA detection methods for MIMO systems that assume knowledge of channel as well as noise covariance matrices were provided in [21]. The energy ratio-based PCA detector described in [22] took use of received signal's asymmetric power levels at legal emitter and receiver.

3. Proposed Pilot contamination Model for MIMO system:

This section discuss novel technique in pilot contamination analysis (PCA) for 5G network based on MIMO by multi antenna routing system. The main aim is to detect pilot contamination and enhance the spectral efficiency of the network. Here the pilot contamination is detected using multi-user pilot scheduling with convolutional adversarial training model. Multi-cell MIMO cell structure is represented in figure-1.



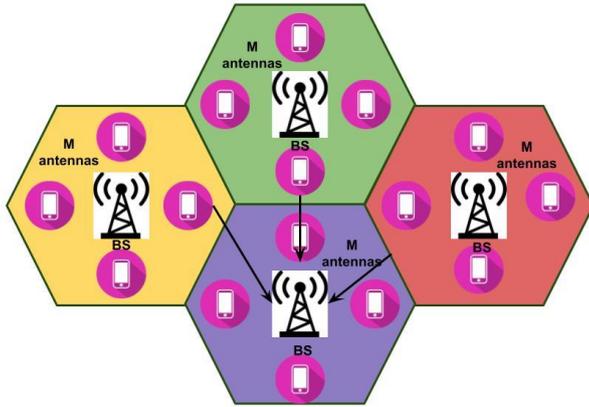


Figure-1 Multi-cell_multi antenna MIMO cell structure

PCA DETECTION AND ANALYSIS:

LU broadcasts a pilot sequence to the BS during the training phase so that it can estimate the channel. The BS creates a precoder to create a beam for the LU. When PCA occurs, Eve transmits the identical pilot sequence in time with LU. PCA allows BS to estimate sum $h_u + h_E$. The precoder design threatens information privacy when PCA is present since it emits two beams toward LU and Eve. The goal of this research is to identify PCA by processing training-phase observations. Received signal Y 's matrix model at BS during training is by eq. (1)

$$Y = \sqrt{P_u} \mathbf{h}_u \mathbf{x}^T + I_e \sqrt{P_E} \mathbf{h}_E \mathbf{x}^T + N \quad (1)$$

where N is a noise matrix, \mathbf{x} is a training sequence, and $I_e \in \{0, 1\}$ is an indicator function with a range of 0 to 1. The noise matrix N 's elements are i.i.d with a Gaussian distribution that has a zero mean and variance σ^2 . LU and Eve both use power $\sqrt{P_u}$ and $\sqrt{P_E}$ to send the identical pilot sequence. We assume that $\sqrt{P_u} = 1$ and $\sqrt{P_E} = 1$. Altered observation signal \mathbf{z} vector is eq. (2)

$$\begin{aligned} \mathbf{z} &= Y \frac{\mathbf{x}^*}{\|\mathbf{x}\|^2} = \mathbf{h}_u \mathbf{x}^T \frac{\mathbf{x}^*}{\|\mathbf{x}\|^2} + I_e \mathbf{h}_E \frac{\mathbf{x}^T \mathbf{x}^*}{\|\mathbf{x}\|^2} + \frac{N \mathbf{x}}{\|\mathbf{x}\|^2} \\ &= \mathbf{h}_u + I_e \mathbf{h}_E + \frac{N}{\tau} \end{aligned} \quad (2)$$

Each pilot symbol is thought to have one unit of energy. Consequently, $\|\mathbf{x}\|^2 = \tau$ training duration during the training phase. It turns out that the observation vector \mathbf{z} is an approximated channel. In PCA, Eve's vector \mathbf{z} represents an estimate of the sum $h_u + h_E$. The pseudo-inverse of the pilot sequence is multiplied by the observation matrix to get the low complexity channel estimator. The updated findings for hypotheses H_0 and H_1 are eq. (3)

$$\begin{aligned} \mathcal{H}_0: \mathbf{z} &= \mathbf{h}_u + \frac{N \mathbf{x}}{\tau} \\ \mathcal{H}_1: \mathbf{z} &= \mathbf{h}_u + \mathbf{h}_E + \frac{N \mathbf{x}}{\tau} \end{aligned} \quad (3)$$

Both LU and Eve have M flat-fading routes that lead to BS. All path gains have an i.i.d Gaussian distribution with a variance of $\sigma_h^2 = \frac{1}{M}$ and a zero mean. Thus, $\mathbf{z} | H_0 \sim \mathcal{CN} \left(0, \left(\frac{1}{M} + \frac{\sigma^2}{M\tau} \right) \mathbf{I}_M \right)$ and $\mathbf{z} | H_1 \sim \mathcal{CN} \left(\mathbf{0}, \left(\frac{2}{M} + \frac{\sigma^2}{M\tau} \right) \mathbf{I}_M \right)$ are distributions of \mathbf{z} under hypothesis H_0 and H_1 . Under hypothesis H_0 and H_1 , we have eq. (4)

$$\begin{aligned} f_{\mathbf{z}|\mathcal{H}_0}(\mathbf{z}) &= \frac{1}{\sqrt{(2\pi\sigma_0^2)^M}} e^{-\frac{E}{2\sigma_0^2}} \\ f_{\mathbf{z}|\mathcal{H}_1}(\mathbf{z}) &= \frac{1}{\sqrt{(2\pi\sigma_1^2)^{MH}}} e^{-\frac{1}{2\sigma_1^2}} \end{aligned} \quad (4)$$

Where $\|\mathbf{z}\|^2 = E$, $\sigma_0^2 = \frac{1}{M} + \frac{\sigma^2}{M\tau}$ and $\sigma_1^2 = \frac{2}{M} + \frac{\sigma^2}{M\tau}$ detector of hypothesis H_0 and H_1 is by eq. (5)

$$\frac{1}{\sqrt{(2\pi\sigma_1^2)^M}} \exp \left(-\frac{E}{2\sigma_1^2} \right) \sum_{\mathcal{H}_0} \frac{1}{\sqrt{(2\pi\sigma_0^2)^M}} \exp \left(-\frac{E}{2\sigma_0^2} \right) \quad (5)$$

By substituting σ_0^2 and σ_1^2 in (6) we have

$$E \sum_{\mathcal{H}_0} \eta = \left(1 + \frac{\sigma^2}{\tau} \right) \left(2 + \frac{\sigma^2}{\tau} \right) \ln \left(\frac{2 + \frac{\sigma^2}{\tau}}{1 + \frac{\sigma^2}{\tau}} \right) \quad (6)$$

The suggested detector in (6) is simple. The observation matrix Y is multiplied with pseudo-inverse of pilot sequence vector \mathbf{x} to estimate the of the reverse channels for LU as well as Eve. The current PCA detectors rely on second order statistics of considerable complexity. It should be noted that the i.i.d. random vectors vector has complicated elements. According to hypotheses H_0 and H_1 , the variance for each dimension is $\frac{\sigma_0^2}{2}$ and $\frac{\sigma_1^2}{2}$, respectively. The $\mathbf{z}|H_0$ and $\mathbf{z}|H_1$ average energies are eq. (7)

$$\begin{aligned} E_{a0} &= E[\mathbf{z}^H \mathbf{z} | \mathcal{H}_0] = \text{Tr}\{E[\mathbf{z}\mathbf{z}^H | \mathcal{H}_0]\} = 1 + \frac{\sigma^2}{\tau} \\ E_{a1} &= E[\mathbf{z}^H \mathbf{z} | \mathcal{H}_1] = \text{Tr}\{E[\mathbf{z}\mathbf{z}^H | \mathcal{H}_1]\} = 2 + \frac{\sigma^2}{\tau} \end{aligned} \quad (7)$$

We are now assessing the analytical PF, PM, PD, and Pe. Uninterruptible power channel estimate's $x = \|\mathbf{z}\|^2$ is a chi-square random variable with a $2M$ degree. Assume that the instantaneous energies under hypotheses H_0 and H_1 , respectively, are $E_0 = E|H_0$ and $E_1 = E|H_1$. Likelihood of missing PM and experiencing a false alarm PF are by eq. (8)

$$\begin{aligned} P_M &= P \left(E_1 = E | \mathcal{H}_1 \right) \\ &< \left(1 + \frac{\sigma^2}{\tau} \right) \left(2 + \frac{\sigma^2}{\tau} \right) \ln \left(\frac{2 + \frac{\sigma^2}{\tau}}{1 + \frac{\sigma^2}{\tau}} \right) \end{aligned}$$

$$P_F = P \left(E_0 = E \mid \mathcal{H}_0 > \left(1 + \frac{\sigma^2}{\tau} \right) \left(2 + \frac{\sigma^2}{\tau} \right) \ln \left(\frac{2 + \frac{\sigma^2}{\tau}}{1 + \frac{\sigma^2}{\tau}} \right) \right) \quad (8)$$

respectively. Thus by eq. (9),

$$\begin{aligned} P_M &= \int_0^\eta f_{\chi^2|\mathcal{H}_1}(x) dx \\ &= \int_0^\eta \frac{1}{\left(\frac{\sigma_1}{2}\right)^n 2^{\frac{n}{2}} \Gamma\left(\frac{1}{2}n\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{\sigma_1^2}} dx \\ &= 1 - \exp\left(-\frac{\eta}{\sigma_1^2}\right)^{m-1} \frac{1}{c!} \left(\frac{\eta}{\sigma_1^2}\right)^c \end{aligned} \quad (9)$$

where $m = M$. Similarly by eq. (10),

$$\begin{aligned} P_F &= 1 - \left(1 - \exp\left(-\frac{\eta}{\sigma_0^2}\right)^{m-1} \frac{1}{c!} \left(\frac{\eta}{\sigma_0^2}\right)^c \right) \\ &= \exp\left(-\frac{\eta}{\sigma_0^2}\right)^{m-1} \frac{1}{c!} \left(\frac{\eta}{\sigma_0^2}\right)^c \end{aligned} \quad (10)$$

PD is evaluated as $PD = 1 - PM$. Probability of error $P_c = \frac{1}{2}(P_M + P_F)$ is by eq. (11)

$$\begin{aligned} P_e &= \frac{1}{2} - \frac{1}{2} \exp\left(-\frac{\eta}{\sigma_1^2}\right)^{m-1} \frac{1}{c!} \left(\frac{\eta}{\sigma_1^2}\right)^c + \\ &\frac{1}{2} \exp\left(-\frac{\eta}{\sigma_0^2}\right)^{m-1} \frac{1}{c!} \left(\frac{\eta}{\sigma_0^2}\right)^c \end{aligned} \quad (11)$$

Multi-user pilot scheduling with convolutional adversarial training model:

In Fig. 2, we examine a $L \times 1$ cell, multi-cell, multi-antenna MIMO method with K single antenna users ($M \gg K$) and a BS with M antennas in each cell. A user who is being attacked in reference (0-th) cell has their pilot contaminated by a single active antenna, Eve. TDD mode as well as channel reciprocity are presumptions. For data transmission in downlink direction, BS modules precoders utilizing an evaluate of reverse CSI of every user in cell. In order to taint LU's channel estimation, an active Eve in reference cell initiates PCA by broadcasting a synchronised pilot sequence of LU that is being attacked to the BS.

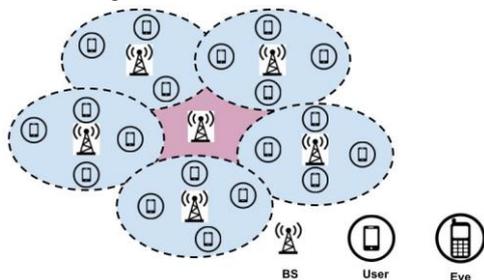


Figure- 2 Multiuser_multi antenna architecture with eavesdropper

The suggested technique for pilot assignment maximises the feasible total rate for the target cell based on the

performance degradation of users who have not been allocated pilot sequences. Taking i th cell as target cell, we consider pilot assignment in that cell, while pilot assignments in other cells are separately controlled by their respective BSs. We first determine user $U_k \in U$ optimal pilot sequence, or pilot that can produce highest utility function eq. (12),

$$\phi^{\text{opt}}(U_k) = \arg \max_{\phi \in \Phi} \left\{ \Theta(U_k, \phi_1), \dots, \Theta(U_k, \phi_{N_\phi}) \right\} \quad (12)$$

When every available pilot sequence except optimal pilot sequence is allocated to user, suboptimal pilot sequence of user U_k is described as pilot that can acquire highest utility function by eq. (13).

$$\phi^{\text{sub}}(U_k) = \arg \max_{\phi \in \Phi} \left\{ \Theta(U_k, \phi_1) \setminus \Theta(U_k, \phi^{\text{opt}}(U_k)) \right\} \quad (13)$$

So, the following formula can be used to calculate the user U_k 's performance degradation by eq. (14):

$$d(U_k) = \Theta(U_k, \phi^{\text{opt}}(U_k)) - \Theta(U_k, \phi^{\text{sub}}(U_k)) \quad (14)$$

The deterioration of each user who has never received a pilot sequence is then calculated. We identify the user among these who experiences the highest decline based on eq. (15):

$$U_{\text{sens}} = \arg \max_{U_k \in U} \left\{ d(U_1), \dots, d(U_{N_U}) \right\} \quad (15)$$

Therefore, in the existing assignment system, the user U_{sens} is the most sensitive user. Thus, the most delicate user U_{sens} is given the best pilot sequence, which is $\phi^{\text{opt}}(U_{\text{sens}})$. The user U_{sens} must then be eliminated from set U , and pilot sequence $\phi^{\text{opt}}(U_{\text{sens}})$ must be eliminated from pool of potential pilot sequences by eq. (16).

$$\begin{aligned} U &\leftarrow U \setminus U_{\text{sens}} \\ \Phi &\leftarrow \Phi \setminus \phi^{\text{opt}}(U_{\text{sens}}) \end{aligned} \quad (16)$$

Additionally, the set U and element counts need to be adjusted as eq. (17):

$$\begin{aligned} N_U &\leftarrow N_U - 1 \\ N_\Phi &\leftarrow N_\Phi - 1 \end{aligned} \quad (17)$$

Until set U is empty, that is, all pilot sequences have been allotted, these operations will be repeated. Only customers that repeatedly employ the pilot sequences see widespread fading, which is related to the uplink asymptotic SINR. The square of the associated large-scale fading is how we define the users' channel quality by eq. (18).

$$S_{ik}^{\text{sig}} = \beta_{iik}^2 \quad (18)$$

As a result, users are divided into groups based on the S_{ik}^{sig} channel quality. Here, we can assume that each cell has an equal number of edge users, where K_e stands for edge users and K_c for centre users, i.e., $K_c = K_{jc}$, $K_e = K_{je}$ and $j = 1, \dots, L$. However, the suggested user grouping-based pilot

scheduling technique raises the pilot overhead, which is determined by eq. (19)

$$\mu_0 \times \frac{L_{K_e+K_c}}{K} \leftarrow \mu_0 \quad (19)$$

where K is number of orthogonal pilot sequences set when FRPS method is selected, and represents number of users in every cell. When suggested technique is used, number of orthogonal pilot sequences is indicated by $L_{K_e} + K_c$. According to enhanced method based on large-scale fading coefficients, grouping criterion is given as when grouping parameter λ is fixed by eq. (20).

$$S_{ik}^{sig} \geq \lambda \rho_i \rightarrow \begin{cases} \text{Yes} & \rightarrow \text{center users} \\ \text{No} & \rightarrow \text{edge users} \end{cases} \quad (20)$$

where grouping threshold value is indicated by $\lambda \rho_i$. Both the user with the best channel condition as well as user with the poorest channel condition must be taken into account. Accordingly, ρ_i can be written as eq. (21)

$$\rho_i = \frac{\max\{s_{i1}^{sig}, s_{i2}^{sig}, \dots, s_{iK}^{sig}\} + \min\{s_{i1}^{sig}, s_{i2}^{sig}, \dots, s_{iK}^{sig}\}}{2} \quad (21)$$

As before, the enhanced user grouping-based pilot scheduling approach results in an additional pilot overhead, which is given by eq. (22)

$$\frac{\mu_0}{K} \left\{ \sum_{j=1}^L K_{j_e} + \max\{K_{1c}, K_{2c}, \dots, K_{Lc}\} \right\} \leftarrow \mu_0 \quad (22)$$

Additionally, a user grouping-based pilot scheduling strategy is suggested to improve method based on deterioration when an additional set of orthogonal pilot sequences is provided. Average uplink reachable rates of centre as well as edge users using the modified pilot allocation technique based on user grouping are given by eq. (23)

$$C_{ik}^{uc} = \left(1 - \frac{\mu_0}{K} \left(\sum_{j=1}^L K_{j_e} + \max\{K_{1c}, K_{2c}, \dots, K_{Lc}\} \right) \right) E\{\log_2(1 + SINR_{ik}^{uc})\}$$

$$C_{ik}^{ue} = \left(1 - \frac{\mu_0}{K} \left(\sum_{j=1}^L K_{j_e} + \max\{K_{1c}, K_{2c}, \dots, K_{Lc}\} \right) \right) E\{\log_2(1 + SINR_{ik}^{ue})\} \quad (23)$$

where $E\{\cdot\}$ denotes expectation operation.

Based on the SINR, each user's transmission speed is chosen. Using pilot levels, large-scale fading, and range of nearby cells, the CATM algorithm must identify the optimal cell-edge users in terms of SINR. The length of output vector and ability to locate a characteristic are both indicated. The probability or length of vector remain same as detected property travels or changes its position. Convolution's output vector is written as eq. (24)

$$U_i = \frac{\|r_i\|^2 r_j}{1 + \|r_j\|^2 \|r_j\|} \quad (24)$$

When training the convolutional adversarial network, the margin loss is applied by eq. (25).

$$K_l = t_l \text{MAX}(0, N^+ - U_l)^2 + \lambda(1 - t_l) \text{MAX}(0, u_l - n^-)^2 \quad (25)$$

Consideration is made easier by the channel estimation process for closest P (=2) sub-lines, l_0 and $l_0 + 1$, notwithstanding the slight case extension employing P > 2. Let's suppose without losing generality that eq. (26)

$$Z_l = Z, E_l = E \text{ and } Y_l = \sqrt{Q}J \text{ for } l \in \{l_0, l_0 + 1\} \quad (26)$$

then Q is used to represent the transmitted power. The matrix of the X_l pilot signal changes to Eq (27),

$$X_l = \sqrt{Q}Z^G G_l H + M_l \quad (27)$$

The H_k and H_s matrices will be utilised to process the X_l during the preliminary estimation. Here is the G_l , which is the result of the estimating process by eq. (28).

$$S_l = H_K X_l H_S = \sqrt{Q} H_K Z^G G_l G H_S + G_K M_l G_S$$

$$H_K = \begin{cases} Z, & N_s < M_s, \\ (ZZ^G)^{-1}Z, & N_s \geq M_s, \end{cases}$$

$$H_S = \begin{cases} E^G & N_t < M_t, \\ E^G (E E^G)^{-1} & N_t \geq M_t, \end{cases} \quad (28)$$

The CATM, which reflects evaluated channel dimensions through $G \hat{l}_0$ and $G \hat{l}_0+1$ mapping connections, simultaneously includes the tentatively estimated dimensions S_{l_0} and S_{l_0+1} by eq. (29).

$$\{\hat{G}_{l_0}, \hat{G}_{l_0+1}\} = e_\phi(S_{l_0}, S_{l_0+1}; \phi) \quad (29)$$

where the CATM parameter configuration is shown by ϕ . Reducing MSE losses in Eq is the CATM's main goal (30),

$$\text{MSE}_{L,SS} = \frac{1}{M_{TS} D^2} \sum_{j=0}^{M_{TS}} \sum_{p=1}^2 \|G_{l_0+p-1}^j - \hat{G}_{l_0+p-1}^j\|_E^2 \quad (30)$$

Eq. (31) provides information on the computational complexity based on CATM's channel estimate,

$$C_{CCNN} \sim O(P M_t M_s (M_t + M_s) + M_t M_s \sum_{k=1}^{K_d} E_T^2 M_{T-1} M_k) \quad (31)$$

A deep learning method is adversarial training networks. It consists of the Discriminator (D) and Generator (G), two separate neural networks. G creates a synthetic sample $G(z)$ using distribution P_g after receiving random noise (z) from a distribution P_z . To minimise the ability of D to distinguish between actual data and synthetic (fake) data, G learns distribution of real data (P_x) during training. This allows G to produce synthetic data that are confirmed by D as real data. The following is how objective function of G is given by eq. (32):

$$\max_D V(D, G) = E_{x \sim p_x} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (32)$$

where z represents random noise from a normal distribution, $E_{z \sim p_z}$ represents expected value on synthetic samples, $G(z)$ represents output of G, indicating likelihood that data are

artificial or real. Similar to D, which aims to increase the likelihood of correctly distinguishing between valid (actual) and invalid (false) data by eq. (33):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] \quad (33)$$

D(x) is result of D for x, which comes from valid data, and x is sample in question. The two networks (D and G) engage in a minimax game, which is modelled as eq. (34):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] \quad (34)$$

Our GAN design's minimax loss function is modified as eq. (35):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_x} [\log D(x)] + \mathbb{E}_{y \sim p_y} [\log (1 - D(G(Y, \phi)))] \quad (35)$$

where G(Y, φ) is channel produced by G (i.e., H) and H is the legal channel matrix.

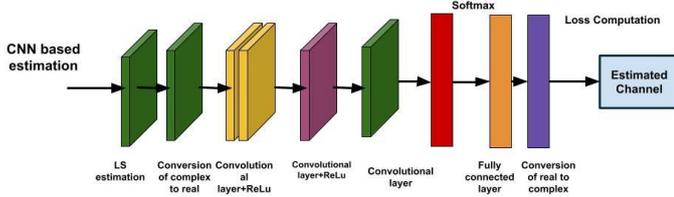


Figure- 3 Proposed neural network framework in PCA.

The structure for our suggested methods is represented in Figure 3, where NN is divided into three layers and coupled to each layer by activation functions. The model in the upper portion uses all three fully connected layers, and model in the lower part uses convolutional layers to simulate a CNN. We refer to approach using the first model as NN-based estimate and approach using second as CNN-based estimation. The neural network receives as input data the LS evaluated channel vector $\hat{h}_{lk}^S \in \mathbb{C}^M$ for incoming pilot signals, and it outputs the desired channel vector $\hat{h}_{lk}^{prop} \in \mathbb{C}^M$.

The neural network's internal weight parameters compute the input-output correspondence, and underlying loss function is specified as eq. (36).

$$\mathcal{L} = \mathbb{E} [\|\hat{h}_{lk} - f(\hat{h}_{lk}^S)\|_2^2] \quad (36)$$

There are 4M nodes in first layer, 64M in second layer, and 2M in final layer in NN-based estimating model. The activation function is the rectified linear unit (ReLU), which has the formula $r(x) = \max(x, 0)$. Technique of NN-based estimation fNN is described as follows if weight specifications at i-th layer are denoted by W_i and b_i by eq. (37).

$$f_{NN}(x) = W_3 \cdot r(W_2 \cdot r(W_1 x + b_1) + b_2) + b_3 \quad (37)$$

Here, $x \in \mathbb{R}^{2M}$ is input for fNN, and is being converted from complex into real,

$x = [\text{Re}(\hat{h}_{lk}^{LS}), \text{Im}(\hat{h}_{lk}^{LS})]^T$. Thus, desired channel is obtained from $\hat{h}_{lk}^{prop} = f(\hat{h}_{lk}^{LS}) = f_{NN}(x)$. For the sake of convenience, the input batch size is set to one here, but in reality, numerous batches of datasets can be handled concurrently.

4. Experimental analysis:

using simulation to assess our job. We simulate a multi-cell network as well as generate training data using MatlabTM. The training and detection procedures are then tested using Python in Google Colaboratory (Colab) setting. Considered is a standard hexagonal cellular network with L cells, each of which contains K users with a single antenna and BS with M antennas. Target cell is core cell surrounded by surrounding cells. Table 1 represents a summary of system parameters.

Table 1: Basic parameters of system simulation

Parameters	Values
Number of cells L	7
Cell radius R	500m
Number of BS antennas M	16-1024
Number of users in every cell k	[10,20]
Path loss	3,8
Average SNR	20dB
Log normal shadowing fading	0-8dB
Spectral efficiency	0.05
Pilot transmission power	30dBm
Data transmission power	30dBm

Table-2 Comparative analysis of SINR

Number of cells	RP_P SK	PCA_MA_M IMO	AI_PCA_5G_MIMO_MAR N_MUPS
20	31	45	51
40	33	48	53
60	36	49	55
80	39	52	59
100	41	56	61
120	43	57	63
140	45	61	66
160	49	63	69
180	51	65	71
200	53	66	72

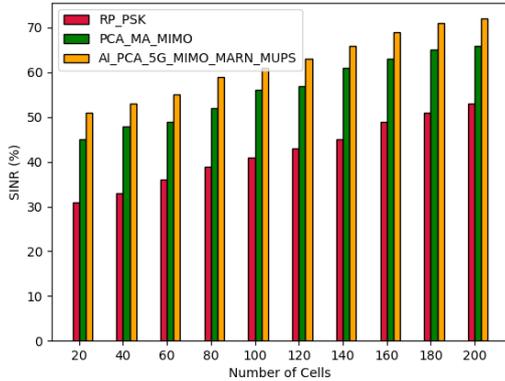


Figure-4 Comparison of SINR

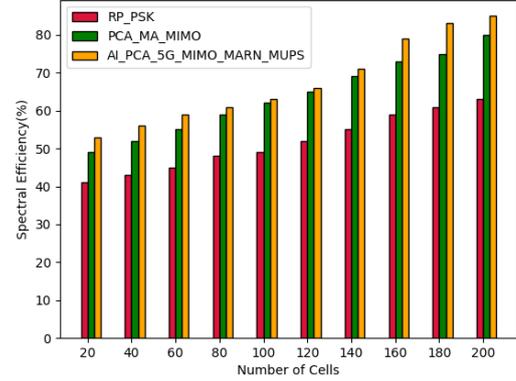


Figure-5 Comparative analysis of Spectral efficiency

The above table-2 and figure-4 shows comparative analysis of SINR between proposed and existing technique based on number of cells. Here the existing technique compared are RP_PSK and PCA_MA_MIMO with proposed AI_PCA_5G_MIMO_MARN_MUPS. In wireless communication systems, such as networks, SINR is a measurement utilized to provide theoretical upper constraints on channel capacity. SINR is described as power of an interest signal divided by total of interference power and some background noise, similar to SNR frequently utilized in wired communications systems. SINR becomes SIR if power of the noise term is zero. In contrast, zero interference transforms SINR to SNR, which is less frequently utilised in the construction of mathematical models of wireless networks like cellular networks. Proposed technique attained SINR of 72% for 200 cells, existing RP_PSK attained 53% and PCA_MA_MIMO attained 66%.

The above table-3 and figure-5 shows comparative analysis of Spectral efficiency between proposed and existing technique. A cellular network's spectral efficiency, also known as bandwidth efficiency, is comparable to highest number of bits of data that are sent to a particular number of users per second while maintaining a respectable level of service. When we talk about spectral efficiency, we often refer to the total spectral efficiency of all transmissions within a cellular network cell. It is expressed as bit/s/Hz. You may calculate the cell throughput in bit/s by multiplying it by the bandwidth. Spectral efficiency attained by proposed technique is 85%, existing RP_PSK attained 63% and PCA_MA_MIMO attained 80% for 200 cells.

Table- 3 comparison of Spectral efficiency

Number of cells	RP_PSK	PCA_MA_MIMO	AI_PCA_5G_MIMO_MARN_MUPS
20	41	49	53
40	43	52	56
60	45	55	59
80	48	59	61
100	49	62	63
120	52	65	66
140	55	69	71
160	59	73	79
180	61	75	83
200	63	80	85

Table-4 Comparative analysis of Normalized MSE

Number of cells	RP_PSK	PCA_MA_MIMO	AI_PCA_5G_MIMO_MARN_MUPS
20	22	31	45
40	25	35	49
60	29	39	53
80	31	42	56
100	35	45	59
120	41	51	63
140	43	53	65
160	45	56	66
180	51	61	71
200	55	63	73

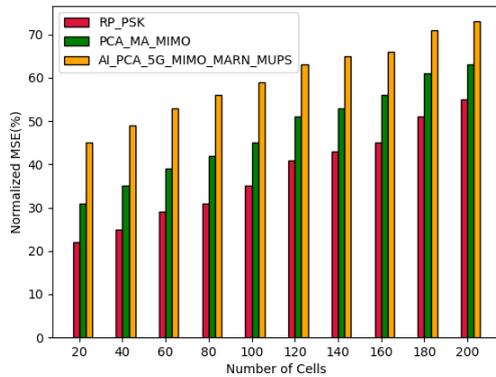


Figure-6 Comparative analysis of Normalized MSE

From above table-4 and figure-6 shows comparative analysis of normalized MSE between proposed and existing technique. When comparing the outcomes of proficiency tests, normalised error, a statistical analysis that takes measurement uncertainty into account, is utilised. In proficiency testing, it is typically the first evaluation used to evaluate conformity or nonconformance (i.e., Pass/Fail). The proposed technique attained normalized MSE of 73%, existingRP_PSK attained 55% and PCA_MA_MIMO attained 63% for 200 cells.

Table-5 Comparative analysis of PCA detection accuracy

Number of cells	RP_PSK	PCA_MA_MIMO	AI_PCA_5G_MIMO_MARN_MUPS
20	65	71	75
40	66	73	77
60	69	77	79
80	73	79	81
100	75	81	83
120	77	83	85
140	79	85	89
160	81	89	92
180	83	91	93
200	85	93	95

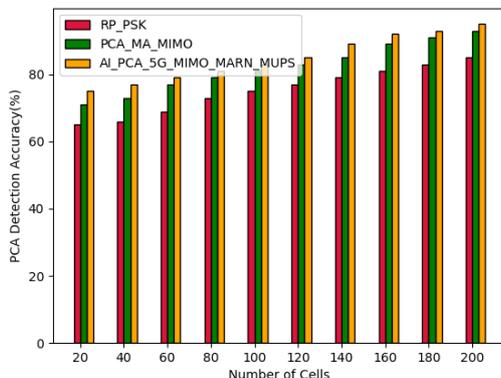


Figure-7 Comparison of PCA detection accuracy

Table-5 and figure-7 shows comparative analysis of PCA detection accuracy between proposed and existing technique. When two terminals use same pilot sequence, pilot contamination happens. Different pilots are used in adjacent cells to suppress it, for instance by having a large number of pilot sequences as well as rotating between them at random. Here the PCA detection accuracy of proposed technique is 95% and existingRP_PSK attained 85% and PCA_MA_MIMO attained 93% for 200 cells.

5. Conclusion:

This research designed novel framework for detection of pilot contamination based on machine learning techniques. proposed design developed PCA using multi-user pilot scheduling with convolutional adversarial training model wherespectral efficiency of the network has been enhanced with PCA detection accuracy. We also came up with a rough analytical MSE expression for the suggested estimator, which is helpful for system design and performance analysis. Additionally, we've demonstrated how this MSE expression asymptotically approaches that of the MMSE estimator. Experimental results has been analysed in terms of SINR, spectral efficiency, normalized MSE and PCA detection accuracy. the proposed technique attained SINR of 72% , spectral efficiency of 85%, normalized MSE of 73%, PCA detection accuracy of 95%.

Reference:

- [1]. Banaeizadeh, F., Barbeau, M., Garcia-Alfaro, J., Kranakis, E., & Wan, T. (2021, September). Pilot Contamination Attack Detection in 5G Massive MIMO Systems Using Generative Adversarial Networks. In *2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)* (pp. 479-484). IEEE.
- [2]. Akgun, B., Krunz, M., & Koyluoglu, O. O. (2017, October). Pilot contamination attacks in massive MIMO systems. In *2017 IEEE Conference on Communications and Network Security (CNS)* (pp. 1-9). IEEE.
- [3]. Wu, Y., Liu, T., Cao, M., Li, L., & Xu, W. (2018). Pilot contamination reduction in massive MIMO systems based on pilot scheduling. *EURASIP Journal on Wireless Communications and Networking*, 2018(1), 1-9.
- [4]. Hassan, M., Ahmed, A., & Zia, M. (2018, October). Detection of Pilot Contamination Attack in Massive MIMO System. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (pp. 1674-1678). IEEE.
- [5]. Kapetanović, D., Zheng, G., Wong, K. K., & Ottersten, B. (2013, September). Detection of pilot contamination attack using random training and massive MIMO. In *2013 IEEE 24th Annual International Symposium on*



- Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 13-18). IEEE.
- [6]. Hassan, M., Zia, M., Ahmed, A., & Bhatti, N. (2020). Pilot contamination attack detection for multi-cell MU-massive MIMO system. *AEU-International Journal of Electronics and Communications*, 113, 152945.
- [7]. Dash, L., &Thampy, A. S. (2022). Optimal Pilot Contamination Mitigation-Based Channel Estimation for Massive MIMO System Using Hybrid Machine Learning Technique. In *Advances in Intelligent Computing and Communication* (pp. 309-321). Springer, Singapore.
- [8]. Hirose, H., Ohtsuki, T., &Gui, G. (2020). Deep learning-based channel estimation for massive MIMO systems with pilot contamination. *IEEE Open Journal of Vehicular Technology*, 2, 67-77.
- [9]. Gholami, R., Cottatellucci, L., & Slock, D. (2021, July). Tackling pilot contamination in cell-free massive mimo by joint channel estimation and linear multi-user detection. In *2021 IEEE International Symposium on Information Theory (ISIT)* (pp. 2828-2833). IEEE.
- [10]. Boulouird, M., Belhabib, A., &Riadi, A. (2021). Encryption based strategy to overcome the problem of pilot contamination within multi-cellular massive MIMO systems. *Wireless Personal Communications*, 119(3), 2639-2655.
- [11]. Li, J., Xue, P., & Wang, W. (2022). Pilot contamination suppression method for massive MIMO system based on ant colony optimization. *Wireless Networks*, 1-10.
- [12]. Belhabib, A., Amadid, J., Boulourd, M., Hassani, M. M. R., &Zeroual, A. (2022). Pilot Contamination Suppression Based Coordination in Multi-cell Massive MIMO Systems. *Wireless Personal Communications*, 1-12.
- [13]. Belhabib, A., Amadid, J., Khabba, A., &Zeroual, A. (2022). Ant Colony-based Optimization algorithm to overcome the pilot contamination issue within multi-cell Massive MIMO systems. In *E3S Web of Conferences* (Vol. 351, p. 01078). EDP Sciences.
- [14]. BELHABIB, A., Amadid, J., &Zeroual, A. (2022). Large-Scale Fading Coefficients Classification-Aided Power Control Strategy for Mitigating the Pilot Contamination in Massive MIMO Systems. *Statistics, Optimization & Information Computing*, 10(1), 107-118.
- [15]. Yuan, Q., Li, L., Xia, W., Zhu, G., Zhu, X., & Wu, J. (2021, July). Research on anti-pilot contamination in massive MIMO. In *Journal of Physics: Conference Series* (Vol. 1983, No. 1, p. 012046). IOP Publishing.
- [16]. Banoori, F., Shi, J., Khan, K., Han, R., & Irfan, M. (2021). Pilot contamination mitigation under smart pilot allocation strategies within massive MIMO-5G system. *Physical Communication*, 47, 101344.
- [17]. Huang, Y., Wu, Q., Lu, R., Peng, X., & Zhang, R. (2021). Massive MIMO for cellular-connected UAV: Challenges and promising solutions. *IEEE Communications Magazine*, 59(2), 84-90.
- [18]. Deshpande, S., Sharma, P., Aggarwal, M., & Ahuja, S. (2021). Mitigating pilot contamination in Rician faded massive MIMO 5G systems using enhanced zero forcing precoding and ring partitioning. *Physical Communication*, 49, 101467.
- [19]. Chen, Y., Ding, M., López-Pérez, D., Yao, X., Lin, Z., & Mao, G. (2021). On the Theoretical Analysis of Network-Wide Massive MIMO Performance and Pilot Contamination. *IEEE Transactions on Wireless Communications*, 21(2), 1077-1091.
- [20]. Dey, A., &Pattanayak, P. (2022). Inter-intra cellular pilot contamination mitigation for heterogeneous massive MIMO cellular systems. *Telecommunication Systems*, 80(1), 91-103.
- [21]. Garg, N., &Ratnarajah, T. (2022). Generalized Superimposed Training Scheme In Cell-free Massive MIMO Systems. *IEEE Transactions on Wireless Communications*.
- [22]. KÖSE, E. C., Yayilkan, A., Kulac, S., &Hakkı, İ. L. K. (2022). Contemporary Approaches to Mitigate Pilot Contamination in Massive Mimo Systems. *AvrupaBilimveTeknolojiDergisi*, (36), 197-206.

