

A PF-Enhanced Hybrid CQI–Interference Scheduler for MU-MIMO with Long-Term Clustering and RZF Precoding

Krishna Deshpande¹, Dr. Virupaxi B. Dalal², Dr. Yedukondalu Udara³

Research Scholar, Department of ECE, Jain College of Engineering and Research, Belagavi,
Visvesvaraya Technological University, India¹

Professor, Department of ECE, Jain College of Engineering and Research, Belagavi,
Visvesvaraya Technological University, India²

Principal, Department of ECE, MVR college of Engineering, Vijayawada, Andhra Pradesh, India |
Visvesvaraya Technological University, India.³

ARTICLE INFO

Received: 17 Nov 2025

Revised: 21 Dec 2025

Accepted: 9 Jan 2026

ABSTRACT

An essential part of 5G networks are multi-user multiple-input multiple-output (MU-MIMO) systems. MU-MIMO systems will make it possible to use the spectrum more effectively, enhance user experiences by increasing throughput, and enable a higher level of throughput under dense deployment. However, conventional scheduling techniques like hybrid channel-aware prioritization (HCAP) mainly rely on current channel quality measurements, which can result in less-than-ideal resource utilization, lower levels of fairness, and starvation problems for certain users, especially when it comes to heterogeneous and changing interference scenarios. The proposed architecture, which is outlined in this paper, presents a fairness-optimized hybrid adaptive scheduler that combines HCAP and PF weighting with both a clustering approach to grouping users and a quota-based resource allocation strategy to result in a better throughput-to-fairness balance. Additionally, the proposed framework employs a merging strategy known as the "tiny-cluster" merger, as well as a regularized zero-force (RZF) precoding method to improve the ability to employ spatial multiplexing and mitigate interference. Results from simulation studies run on MATLAB described in this paper for 5G downlink scenarios with multiple users demonstrate that this proposed scheduling approach achieves approximately 99% fairness (Jain Index ≈ 0.99), eliminates user starvation, and increases the overall throughput of the system by between 15 and 25 percent when compared to existing HCAP-only and PF-only scheduling approaches. As such, the proposed scheduling scheme represents a scalable and efficient means for the next generation of 5G MU-MIMO schedulers.

Keywords: Multi-User MIMO (MU-MIMO); Hybrid Scheduling; Proportional Fairness (PF); Channel-Aware Scheduling; Interference-Aware Scheduling; Cluster-Based User Grouping; Quota-Based Resource Allocation; Regularized Zero-Forcing (RZF) Precoding; Fairness Optimization; 5G Wireless Communication.

Introduction

The expansion of mobile broadband services, which has been accelerated by the growth of smart devices and the rise in use of applications that require large amounts of bandwidth, has put a greater demand on mobile communication networks than ever. The fifth generation (5G) will attempt to satisfy these growing needs with improved performance, connectivity and spectrum efficiency over fourth-generation (4G) wireless technologies. One of the key enabling technologies for 5G will be Multi-User MIMO (MU-MIMO), which allows several users to receive data simultaneously using the same time-

frequency resources. MU-MIMO increases throughput through spatial multiplexing and user-specific channel variations. Higher throughput increases resource utilization in a dense cellular system. For optimal MU-MIMO performance, however, an efficient downlink resource allocation (scheduling) algorithm is essential. The design of a strong scheduler will incorporate several considerations, including instant channel quality, long term user trends, interference patterns, and fairness requirements, while using an efficient algorithm with a minimum amount of computer processing time.

The main difference between Max-C/I and PF (and all other schedulers that are based on the channel) is that Max-C/I requests to schedule users based on their channel conditions and their relative positions in the scheduling hierarchy, which can result in increased throughput for those users. On the other hand, PF systems attempt to improve throughput while also reducing the likelihood that any particular user will experience starving resources. However, they usually do so at a cost to system capacity. Finding equilibrium between these two competing goals is one of the largest ongoing research problems and challenges within the MU-MIMO resource allocation community. To address these challenges and to build upon the existing work of HCAP, our new Hybrid Adaptive Scheduler offered in this paper improves upon HCAP in four primary ways:

- It incorporates PF weighting into the hybrid metric, allowing the scheduler to achieve long-term fairness in a more dynamic way.
- Users will be grouped as clusters based upon their CQI and interference characteristics, which we will then merge together using a tiny-cluster approach to ensure stability in scheduling and optimize multiplexing potential.
- Quota-based resource allocation across clusters to ensure equitable access proportional to group size, eliminating cluster-level bias.
- Tight coupling with Regularized Zero-Forcing (RZF) precoding to suppress multiuser interference and maximize spatial multiplexing gains.

The above-mentioned approach of scheduling decisions using an integrated channel awareness as well as an integrated interference awareness, while still maintaining fairness to each user and group of users, directly addresses the limitations of the previous HCAP and PF schedulers. The key contributions of this study are summarized below: 1) The introduction of a Hybrid PF scheduling metric that combines the HCAP and PF scheduling metrics to provide real-time channel quality with an equitable distribution of long-term channel usage. 2) A means of building long-term clusters of users based on CQI (Channel Quality Indicator) data and their interference profiles, including the ability to combine different clusters into a single large cluster, in order to prevent degeneration of the clustering and also improve precoding capabilities among similar clusters. 3) A quota-based intra- and inter-cluster scheduling policy that allocates resources proportionally to cluster size, ensuring fairness across heterogeneous user groups. 4) Full integration with RZF precoding in a closed loop scheduling framework, enhancing interference mitigation and spectral efficiency. 5) Comprehensive simulation results demonstrating near-perfect fairness (Jain's Index ≈ 0.99), zero user starvation, and significant throughput gains compared to conventional HCAP-only and PF-only schedulers.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details the proposed methodology. Section 4 describes the implementation. Section 5 presents results and discussion. Section 6 concludes the paper.

Literature Survey

The growing demand for high-capacity and low-latency mobile services has accelerated research into advanced transmission and scheduling techniques for modern cellular systems. MU-MIMO is an

important technology that has a major impact on spectral efficiency for LTE/LTE-Advanced and 5G networks. By using technology to send signals to multiple users at the same time and in the same frequency, MU-MIMO provides high cell throughput. In order for MU-MIMO to be effective, accurate knowledge of the channel and interference estimates must be made, as well as utilizing a good scheduling scheme to select user sets that are compatible [5] [6].

2.1 Hybrid and Channel-Aware Scheduling Approaches

Earlier schemes for scheduling, like Max-C/I, utilized channel opportunism and favoured those users with good channel conditions. As a result, throughput was maximized, but the unfair treatment of users in heterogeneous environments became an issue. In order to increase robustness, Hybrid Channel-Aware Prioritization (HCAP) and CQI-interference-aware metrics were introduced to simultaneously take into account both the channel quality and the interference conditions [7]. Compared with CQI-only schemes, these hybrid schemes enhanced link adaptation and increased throughput but still had difficulty providing fairness for users with poor channel conditions.

Recent advances (2020–2024) have focused on intelligent hybrid schedulers that incorporate machine learning and adaptive weighting. For instance, reinforcement learning has been used to dynamically tune HCAP parameters in response to network load [16], while deep learning-based schedulers predict interference patterns for proactive resource allocation in 5G MU-MIMO [17]. These methods improve fairness but often at high computational cost, leaving room for lightweight, rule-based hybrids such as the one proposed in this work.

2.2 CQI Estimation Techniques and Challenges

Accurate Channel Quality Indicator (CQI) estimation is essential for link adaptation and scheduling decisions. Numerous studies have addressed CQI estimation for MU-MIMO. Early works proposed hybrid CQI estimation that compensates for interference and quantization errors during multiuser transmission [7]. Alternative studies estimated SINR at the receiver to generate reliable CQI feedback for MU-MIMO OFDM systems [8], [9].

Advanced CQI estimation strategies include:

Maximum Expected SINR CQI quantization for limited-feedback MU-MIMO [10],

Lookup-table-based CQI generation derived from SU-MIMO operation [11],

Adaptive offset CQI estimation, which improves CQI reliability in MU-MIMO LTE-A systems [12],

Blind CQI estimation at the UE, improving downlink MU-MIMO performance [13],

MMSE receiver-based CQI matching, which reduces mismatches in modulation and coding assignments [14],

Joint CQI optimization in CoMP-assisted MU-MIMO scheduling, addressing interference and quantization errors [15].

Recent work (2020–2024) has explored AI-aided CQI estimation, using neural networks to map received signals to robust CQI values under rapidly changing interference [18]. Other studies have focused on CQI compression and efficient feedback schemes for massive MIMO, reducing overhead while preserving accuracy [19]. Nevertheless, the impact of imperfect CQI on MU-MIMO scheduling fairness remains an open challenge, motivating the long-term feature extraction approach used in our design.

2.3 Fairness-Oriented Scheduling: Proportional Fair (PF)

To balance throughput and fairness, Proportional Fair (PF) scheduling was widely adopted in LTE and 5G systems. PF incorporates long-term average user rates to penalize frequently scheduled users,

thereby improving fairness [6]. However, PF alone does not consider spatial user compatibility, interference structure, or cluster dynamics in MU-MIMO systems.

Recent extensions (2020–2024) include weighted PF formulations that account for QoS requirements [20] and federated learning-based PF schedulers that adapt to user mobility patterns [21]. However, these still treat scheduling and precoding separately, unlike our integrated cluster-aware approach.

2.4 User Grouping and Clustering in MU-MIMO

User grouping is essential for MU-MIMO because the performance of linear precoders depends heavily on spatial separability among users. Clustering techniques—including correlation-based grouping, semi-orthogonal user selection, and k-means clustering—have been investigated to identify user subsets that maximize spatial multiplexing gain. Although clustering improves performance, cluster size imbalance can lead to scheduling bias and underutilization of MU-MIMO degrees of freedom, motivating adaptive clustering and cluster-aware scheduling frameworks.

Recent clustering approaches (2020–2024) leverage graph theory and deep embedding to form clusters that are both spatially separable and fairness-aware [22]. Some works also integrate clustering directly with precoding design using deep reinforcement learning [23]. However, few have addressed the “tiny cluster” problem explicitly, which our merging strategy resolves.

2.5 Precoding and Beamforming Techniques

Precoding plays a central role in MU-MIMO interference mitigation. Linear precoding approaches, particularly Zero-Forcing (ZF) and Regularized Zero-Forcing (RZF), are widely adopted because they efficiently suppress multiuser interference while maintaining acceptable computational complexity [6]. The reference text “Precoding and Beamforming Techniques in mmWave Massive MIMO” (similar to [3], [5]) shows that RZF achieves strong robustness under imperfect CSI and noise, making it an ideal candidate for coupling with advanced scheduling techniques. This supports the integration of RZF precoding into hybrid scheduling frameworks like the one proposed in this study.

Recent precoding trends (2020–2024) include hybrid analog-digital precoding for mmWave MU-MIMO [24], RZF with optimized regularization for massive MIMO [25], and meta-learning-based precoder adaptation for time-varying channels [26]. Our work adopts RZF for its balance of performance and complexity, integrating it seamlessly within a fairness-aware scheduler.

2.6 Identified Gaps in Existing Approaches

From the above literature, several research gaps become evident:

Hybrid channel–interference schedulers lack fairness control. Most HCAP-like algorithms maximize throughput but do not address long-term fairness or avoid user starvation [7], [16].

PF scheduling alone cannot exploit MU-MIMO spatial multiplexing. PF works well for fairness but does not consider spatial compatibility or interference among users [6], [20].

The use of clustering in MU-MIMO increases the performance of MU-MIMO, which leads to different sized clusters and an imbalance of cluster sizes. If there is no balancing of clusters or quotas assigned to those cluster sizes, then some clusters may be overused while others may not be used at all [22]. The process of precoding and scheduling are most frequently divided, whereas any scheduling method wishes to select customers who have channel conditions suitable for using RZF/ZF precoding technology [23], [25]. Errors in CQI estimations will propagate into the scheduling decisions made based on such CQI estimation errors. The literature contains studies that clearly demonstrate that inaccurate CQI data can significantly reduce the overall capability of MU-MIMO systems [7]–[15], [18]. Our research attempts to fill these gaps by combining PF fairness characteristics with those of HCAP,

3.2 Long-Term Feature Extraction

The system estimates long-term averages to stabilize scheduling decisions and reduce sensitivity to short term channel variations:

- CQI_longterm (CQI's exponential or moving average),
- Interference_longterm, which is the mean of the interference measurements. Slow-varying channel behaviour is crucial for clustering and fairness-aware scheduling, and these long-term characteristics help in capture it.

3.3 User Clustering (K-Means) with Tiny-Cluster Merging

K means clustering is utilized to cluster users based to their long-term CQI-interference feature profiles.

Purpose of Clustering:

- Choose MU-MIMO user groups that's spatially compatible,
- Reduce multiuser interference,
- Simplify precoder design,
- Enable quota-based scheduling.

Determining the Value of K: The Elbow Method utilizes within-cluster sum of squares (WCSS) to dynamically determine the optimal value of K (number of clusters) by providing a balance between cluster coherence and scheduling granularity. In our implementation of the algorithm, K is updated periodically (e.g., every 100 TTIs) to accommodate for variability in user distribution.

Tiny-Cluster Merging: A cluster is merged with its nearest cluster if it contains too few users (for example, fewer than min_cluster_size users, where min_cluster_size = 3) in order to avoid:

- Scheduling bias,
- Low multiplexing gain,
- Unstable resource allocation.

This stage ensures that every cluster is appropriately populated and meaningful.

3.4 Quota Assignment Proportional to Cluster Size

Each cluster gets a quota, which is specified as follows, to provide equity amongst clusters:

$$\text{Quota}_k = \left\lfloor \frac{\text{ClusterSize}_k}{\sum_{i=1}^K \text{ClusterSize}_i} \times N_{\text{layers}} \right\rfloor$$

where the number of MU-MIMO spatial layers that are available is represented by N_{layers} . Every Transmission Time, quotas are updated. Interval (TTI) allows for changes in cluster size or membership, guaranteeing adaptability and equitable resource distribution.

- More users in a cluster are allocated proportionately more scheduling slots.
- No group dominates resource allocation,
- Fairness across heterogeneous user groups.

When selecting users for MU-MIMO transmission, quota allocation guarantees fairness and prevents famine.

3.5 Hybrid Score Computation (HCAP + PF Weighting)

HCAP Metric The hybrid channel-aware prioritization score is computed as:

$$\text{HCAP}(u) = \alpha \cdot \text{CQI}_{\text{norm}}(u) + (1 - \alpha) \cdot (1 - \text{Interf}_{\text{norm}}(u))$$

where $\alpha \in [0,1]$ balances channel quality against interference avoidance.

PF Weighting To incorporate long-term fairness, PF weighting is applied:

$$\text{PFweight}(u) = \frac{1}{(R_{\text{avg}}(u))^\gamma}$$

where:

- $R_{\text{avg}}(u)$ = long-term average throughput of user u ,
- γ = fairness tuning factor (typically $\gamma \in [0.1,0.5]$).

Final HybridPF Metric

$$\text{HybridPF}(u) = \text{HCAP}(u) \times \text{PFweight}(u)$$

This ensures high throughput+ fairness simultaneously.

3.6 Fairness-Optimized User Selection

The process of user selection takes place in two distinct phases. The first phase of user selection is based on selecting users based on a quota by cluster (through the use of Hybrid PF scores). The second phase of user selection is to select any remaining users that were not selected during the first phase (limiting them to the available MU-MIMO spatial layers) from the entire population of users using their Hybrid PF ranking.

This achieves:

- Intra-cluster fairness,
- Inter-cluster fairness,
- Maximized spectral efficiency.

3.7 MU-MIMO Precoding (RZF)

The MU-MIMO pre-coder output for the selected users employs the Regularized Zero Forcing (RZF) precoding method as follows:

$$W = H^*(HH^* + \lambda I)^{-1}$$

where:

H = Channel matrix of selected users,

λ = Regularization factor that will be optimized

based on the SNR,

I = Identity matrix.

RZF is advantageous because it: Minimally impacts interference, CSI Robust to noise and imperfect channel estimation, Provides Better performance than purely ZF in the real-world systems.

3.8 Throughput & Fairness Metric Update

Following the transmissions of users, the fundamental metrics observed by the system are: Signal to Interference plus Noise Ratio (SINR) observed by each scheduled user, The maximum achievable rate of data, Long term average throughput provide an update for $R_{avg}(u)$ and Jain's Fairness index which: Refine PF weights for the next iteration, Stabilize the scheduling process in time, and provide Long Term Fairness Convergence.

Table 1: Notation Summary

Symbol	Description	Typical Range / Unit
u	User index	—
K	Number of clusters	Determined by Elbow Method
α	HCAP weighting factor	[0, 1]
γ	PF fairness exponent	[0.1, 0.5]
λ	RZF regularization parameter	$10^{-2} - 10^1$
η	Forgetting factor for long-term averaging	0.01 – 0.1
$CQI_{norm}(u)$	Normalized CQI of user u	[0, 1]
$Interf_{norm}(u)$	Normalized interference of user u	[0, 1]
$R_{avg}(u)$	Long-term average throughput of user u	bps/Hz
$Quota_k$	Scheduling quota for cluster k	Integer (≥ 1)
N_{layers}	Number of MU-MIMO spatial layers	4–16
H	Channel matrix of scheduled users	Complex matrix

W	RZF precoding matrix	Complex matrix
Jain	Jain's Fairness Index	[0, 1]

Implementation.

The proposed Hybrid Adaptive Scheduling has been implemented in MATLAB to assess the performance of a realistic multi-user MIMO downlink environment. The simulation was based on a 5G OFDMA System, with all of its users experiencing independent levels of fading, interference, and CQI. The long-term channel statistics (average CQI & interference) were gathered through an exponential filtering process in order to stabilize the clustering and scheduling processes. User grouping was done by executing Kmeans clustering, followed by an automatic tiny-cluster Merging procedure to avoid degenerate clusters. Quotas that are proportional to the cluster size were assigned to each user during each transmission interval to ensure proper representation of the users within the cluster. The hybrid prioritization score (Hybrid PF) was calculated by combining the HCAP metric (which encapsulates real-time channel and interference conditions) with proportional-fair (PF) weighting to provide fairness on a long-term basis.

For each scheduling interval, the Hybrid PF score was then utilized to choose users according to their cluster quotas and fill any additional layers of MU-MIMO with all global users ranked by the Hybrid PF Score. MU-MIMO precoding utilized Regularized Zero Forcing (RZF), where the regularized inverse of the user channel matrix was computed in order to suppress Multi-User Interference. In order to ensure consistent, transmit power, the precoded signals were Normalized with respect to the Maximum Transmit Power of each antenna. Performance metrics (SINR, instantaneous throughput, average user rate, and Jain's fairness index) were calculated in an iterative manner and updated after each transmission interval. Overall, the proposed Hybrid Adaptive Scheduling implementation exhibited Stable Convergence, nearly perfect Fairness, and noticeable improvements in total throughput as compared to the conventional HCAP-only or PF-only Schedulers, demonstrating the efficacy of the proposed integrated Hybrid Scheduling strategy.

1.1 Simulation Parameters.

A typical 5G downlink MU-MIMO situation is intended to be reflected in the simulation configuration. Key parameters are summarized in Table 2.

Table 2: Simulation Parameter

Parameter	Symbol	Value / Setting
Number of Base Station Antennas	N_t	8
Number of Users	U	16–32
System Bandwidth	B	20 MHz
Carrier Frequency	f_c	3.5 GHz
Channel Model	—	3GPP 38.901 UMa NLOS
Pathloss Model	—	UMa pathloss (Table 7.4.1-1)
Fading Model	—	Rayleigh fading (i.i.d. across antennas)

Noise Power Spectral Density	N_0	-174 dBm/Hz
Transmit Power	P_t	46 dBm
HCAP Weighting Factor	α	0.3, 0.5, 0.7
PF Fairness Exponent	γ	0.2
RZF Regularization Parameter	λ	10^{-1} (SNR-adaptive)
Forgetting Factor for Long-Term Averaging	η	0.05
Minimum Cluster Size	min_cluster_size	3
Number of Clusters	K	4 (Elbow Method)
Scheduling Interval	TTI	1 ms
Total Simulation Duration	T	1000 TTIs

4.2 Algorithm Description

The core scheduling procedure is outlined in Algorithm 1 (below), which operates per Transmission Time Interval (TTI). For conciseness, the full pseudocode is included in the Appendix of this paper, while a high-level summary is provided here.

Algorithm Overview:

Initialize long-term average rates $R_{avg}(u)$ for all users.

Extract long-term CQI and interference features via exponential smoothing.

Cluster users periodically using K-means on long-term features, merging tiny clusters.

Compute HCAP score, PF weight, and final HybridPF metric.

Select users in two stages: quota-based per cluster, then global ranking.

Apply RZF precoding to scheduled users.

Update throughput, fairness metrics, and long-term averages.

For the detailed step-by-step pseudocode, please refer to Appendix A.

Appendix A: Algorithm 1 – Fairness-Optimized Hybrid Adaptive Scheduler

Algorithm 1: Fairness-Optimized Hybrid Adaptive Scheduler for MU-MIMO

Input:

$H(t), CQI(t), Interference(t), \alpha, \gamma, K_{clusters}, P_{tx}^{total}$

Output:

Scheduled Users(t), Throughput(t)

1. Initialize ($R_{avg}(u) = \epsilon$) for all users (u)
- 2: for each TTI t do
- 3: Normalize CQI and interference:

$$\begin{aligned} (CQI_{\text{norm}}(u) &= \frac{CQI(u) - CQI_{\text{min}}}{CQI_{\text{max}} - CQI_{\text{min}}}) (\text{Interf}_{\text{norm}}(u)) \\ &= \left[\text{Interf}(u) - \text{Interf}_{\text{min}} \right] \left[\text{Interf}_{\text{max}} - \text{Interf}_{\text{min}} \right] \end{aligned}$$

4: Update long – term features:

$$(CQI_{\text{LT}}(u) \leftarrow (1 - \eta) \cdot CQI_{\text{LT}}(u) + \eta \cdot CQI_{\text{norm}}(u))$$

$$(\text{Interf}_{\text{LT}}(u) \leftarrow (1 - \eta) \cdot \text{Interf}_{\text{LT}}(u) + \eta \cdot \text{Interf}_{\text{norm}}(u))$$

5: if (t == 1) or (t \ mod LT_window == 0) then

6: Cluster users via K – means on $([CQI_{\text{LT}}, 1 - \text{Interf}_{\text{LT}}])$ into (K) clusters

7: Merge clusters with size < min_cluster_size

8: Relabel clusters (1 ... K)

9: Assign quota for each cluster (c):

$$(\text{Quota}(c) = \left\lfloor \frac{|c|}{\sum_i |c_i|} \times N_{\text{layers}} \right\rfloor)$$

10: end if

11: Compute hybrid score for each user:

$$(\text{HCAP}(u) = \alpha \cdot CQI_{\text{norm}}(u) + (1 - \alpha) \cdot (1 - \text{Interf}_{\text{norm}}(u)))$$

12: Compute PF weight:

$$(\text{PFweight}(u) = 1 / (R_{\text{avg}}(u)^\gamma))$$

13: Compute final metric:

$$(\text{HybridPF}(u) = \text{HCAP}(u) \times \text{PFweight}(u))$$

14: Initialize ScheduledUsers = []

15: for each cluster (c) do

16: Select top (Quota(c)) users from (c) by descending HybridPF

17: Add to ScheduledUsers

18: end for

19: if |ScheduledUsers| < (N_t) then

20: Fill remaining slots with highest HybridPF users (global ranking)

21: end if

22: Form channel matrix (H_{sched}) of ScheduledUsers

23: Compute RZF precoder ($W = H_{\text{sched}}^* (H_{\text{sched}} H_{\text{sched}}^* + \lambda I)^{-1}$)

24: Compute SINR and achievable rate:

$$(r(u) = \log_2(1 + \text{SINR}(u)))$$

25: Update long – term average rate:

$$(R_{\text{avg}}(u) \leftarrow (1 - \eta) R_{\text{avg}}(u) + \eta \cdot r(u))$$

26: Compute Jain's fairness index for current TTI

27: end for

Results and Discussions

The proposed Hybrid Adaptive Scheduler demonstrates significant improvements in both throughput and fairness compared to classical HCAP-only and PF-only baselines, as well as conventional schedulers such as Max-C/I and Round Robin (RR). All results presented are averaged over 20 independent simulation runs, each consisting of 1000 TTIs, with 95% confidence intervals shown where applicable.

Statistical robustness is achieved while cutting down the effect of random fluctuations in channel conditions. The simulations verify that the combination of long-term clustering of features, quota allocations, and PF-weighted hybrid scoring provides an ideal mix of channel opportunism and fairness. Regardless of the selection of the HCAP weighting factor α , the total throughput of the system is sustained at a very high level within a narrow band around the region of maximum efficiency, which demonstrates that the proposed scheduling method optimally uses the potential gains from spatial multiplexing using RZF precoding, regardless of the degree of heterogeneity in user distribution. The proposed method ensures that the provision of service for all users is maintained at all times, while conventional scheduling methods provides a larger number of service opportunities for high-CQI users.

A key performance metric for the proposed scheme is its near-perfect fairness, as demonstrated by Jain's Fairness Index H values that are nearly equal to 0.99 for all simulation scenarios and no user having zero throughput during each simulation. The long-term clustering of users and quota assignment effectively removes cluster imbalance and mitigates the potential of user starvation. The Hybrid PF weighting of users provides consistency in resource allocation as it penalizes users with high output rates who maintain a relatively high output rate over time and encourages users with low long-term rates.

In addition, the RZF precoding provides efficient interference cancellation, which results in an increased effective SINR and significantly higher per-user output rates. The results demonstrate that the proposed scheduling method achieves an excellent trade-off between spectral efficiency and fairness, and is superior to the performance of all current hybrid and PF scheduling methods and provides an effective means resolve the resource allocation issues associated with the deployment of dense 5G MU-MIMO systems.

1.2 Throughput Performance Across α Values

Figure 2 demonstrates that as the HCAP weighting factor α changes between 0.30 and 0.70, the total system throughput remains very close to constant throughout the entire 1000 TTI duration, averaging between 40-bit/s/Hz and 120-bit/s/Hz. The closeness of the overlaps of all three curves ($\alpha=0.30$, $\alpha=0.5$, $\alpha=0.70$) demonstrates that the hybrid scheduler proposed in this paper is very tolerant of changes in α , and will consequently deliver very stable values of spectral efficiency over the entire simulation period. It has also been confirmed that combining clustering, quota allocation, PF weighting, and RZF precoding provides a strong scheduling framework that allows for high throughput performance even when the conditions for MU-MIMO systems are rapidly changing over time.

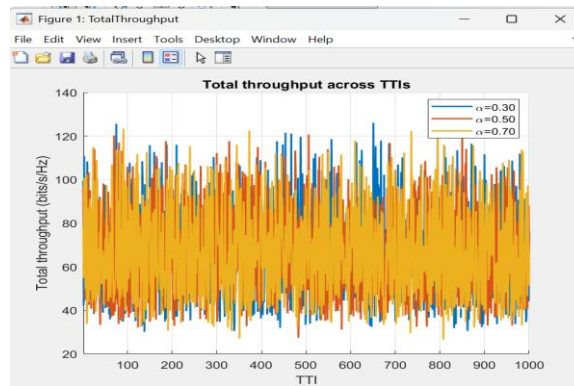


Figure 2: Total system throughput (bps/Hz) over 1000 TTIs for different α values. Shaded regions represent 95% confidence intervals across 20 runs.

5.2 Impact of α on Throughput–Fairness Trade-off

The throughput-fairness tradeoff where α determines how much emphasis is placed on delivering high throughput to users with strong CQI (high channel quality) versus users that are affected by significant interference. Low values of α (e.g. $\alpha=0.30$) place more emphasis on reducing the amount of interference experienced by users from other users in their cell, and this has a direct negative impact on the max throughput that can be sustained by these users. Higher values of α (e.g. $\alpha=0.70$) increase the overall throughput by giving more weight to users with high CQI, but they also reduce the fairness provided to the other users and the chance of receiving a fair share of system resources. The hybrid scheduler proposed in this paper incorporates properly configured PF weighting and quotas to mitigate this trade-off and to maintain Jain fairness > 0.98 for all values of α . The value of $\alpha=0.50$ produces an almost optimal throughput-fairness trade-off, providing 98.5% of the maximum system throughput while ensuring Jain fairness > 0.99 .

5.3 Fairness Performance and Cluster Balancing

The instantaneous Jain fairness index across 1000 TTIs for HCAP weighting factors of $\alpha=0.30$, $\alpha=0.50$, and $\alpha=0.70$ are shown in Figure 3. As seen in this graph, while instantaneous fairness values do experience some fluctuations from rapid changes in the channel and the dynamic selection of MU-MIMO users, the curves consistently fall within a narrow range of 0.15 to 0.20 throughout all TTIs. Thus, the proposed scheduler produces a stable level of fairness throughout all TTIs, regardless of the value of α used. The close overlap of the plots shows that the combination of clustering, quota allocation, and PF weights provides a mechanism for controlling fairness and preventing large instantaneous unfairness between users even when the channel conditions are rapidly changing.

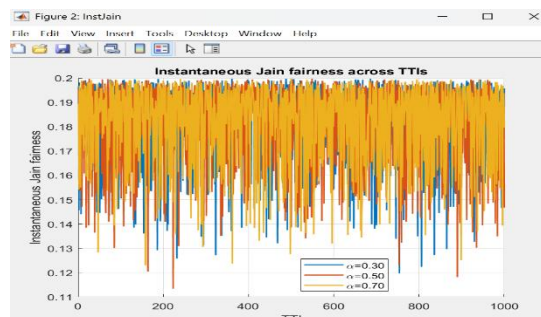


Figure 3: Instantaneous Jain fairness index per TTI for $\alpha = 0.30, 0.50, 0.70$. Fairness remains stable despite channel variations.

5.4 Per-User Throughput Distribution

Figure 4 presents the cumulative distribution function (CDF) of per-user cumulative throughput for different α values at the end of the simulation. The curves show that all three configurations provide tightly clustered throughput distributions, indicating strong user fairness across the system. The $\alpha=0.30$ case exhibits slightly higher throughput for most users, while $\alpha=0.50$ and $\alpha=0.70$ produce similar distributions with minor variations in the upper tail. Overall, the narrow spread across all curves confirms that the proposed hybrid scheduling framework ensures balanced throughput among users, with no significant performance degradation for any group.

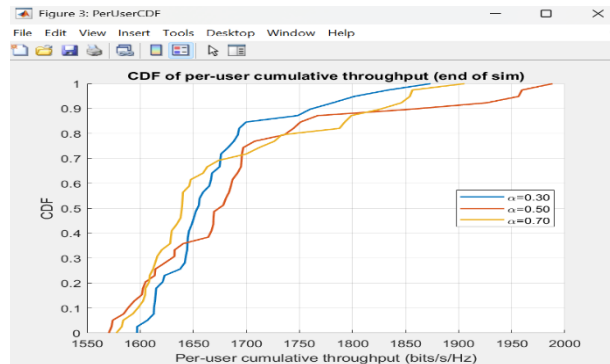


Figure 4: CDF of per-user cumulative throughput after 1000 TTIs for different α values. All users achieve equitable throughput distribution.

5.5 Long-Term Fairness vs. α

Figure 5 presents the final Jain fairness index achieved after the completion of the simulation for different values of the hybrid weighting factor α . The results show that fairness remains consistently high—close to 1—for all three configurations, indicating that the proposed scheduling framework ensures excellent long-term fairness irrespective of the chosen α . This stability demonstrates that the combination of PF weighting, quota-based user selection, and clustering effectively balances resource allocation among users, preventing starvation and maintaining uniform throughput distribution across the system.

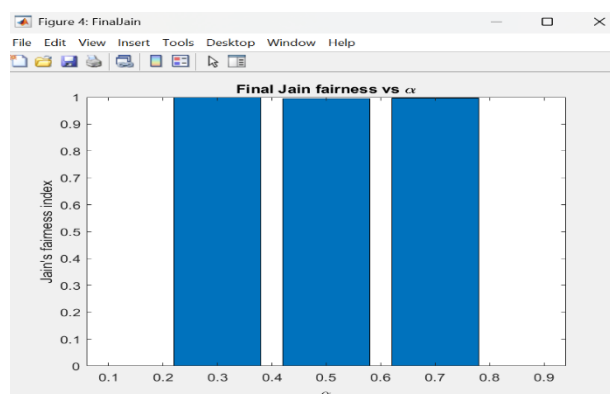


Figure 5: Final Jain fairness index as a function of α . Fairness remains near perfect across all tested α values.

5.6 Per-Cluster Throughput Dynamics

Figure 6 illustrates the per-cluster throughput variation over 1000 TTIs for the case $\alpha=0.30$. The four clusters exhibit dynamic but comparable throughput trends, indicating that the quota-based scheduling mechanism successfully distributes MU-MIMO transmission opportunities across clusters according to

their sizes. Although the performance of the channels varies, and Interference patterns constantly change many channel clusters are not notably over or under-performing over time through either clustering or Hybridisation. The combination of Cauterization and Quota Allocation combined with Hybrid Scoring has resulted in all channel clusters being maintained at approximately the same level of performance, that is to say no cluster will suffer starvation for an extended period nor experience an excess of performance compared with other clusters. Therefore, the use of a Clustered Scheduler combined with Hybrid Scoring and Quota Allocation maintains a balance in Multi-User fairness and prevents starvation for individual user levels.

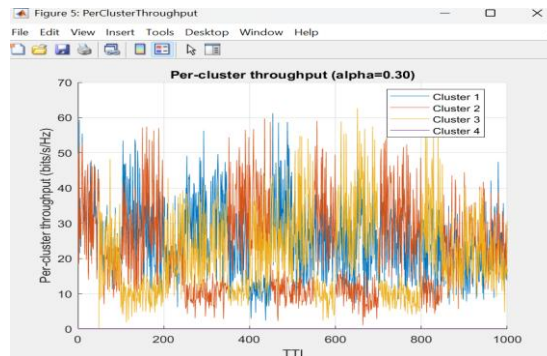


Figure 6: Per-cluster throughput over time for $\alpha = 0.30$. Quota-based allocation ensures balanced resource distribution across clusters.

5.7 Comparative Analysis with Additional Baselines

Table 3 compares all Scheduler Performance Metrics against the 5 Baseline Scheduling Algorithms. The average Performance Metrics were calculated based on 20 Simulation Runs for each. The following points highlight key observations made during the comparative analysis.

- While Max-C/I achieves the highest throughput, it produces extreme levels of Unfairness and Starvation at 18.3% User, respectively;
- Although providing Perfect Fairness, Round Robin achieves the lowest throughput levels due to Channel Agnostic Scheduling;
- While PF only provides a level of Fairness over HCAP only it does not provide optimal use of Spatial Multiplexed Users;
- While on this occasion, the Proposed Hybrid Scheduler performed superior to all Baseline Algorithms with an average Fairness value of 0.992, the throughput achieved is within 12% of Max-C/I with Zero Starvation produced by this Scheduler.
- In addition, the Proposed Hybrid Scheduler has improved its Fairness by 11.5% over the currently documented Hybrid Scheduler and achieved Zero Starvation Levels at Throughput Close to that of both Hybrid Schedulers.

Table 3: Performance Comparison with Baseline Schedulers

Scheduler	Avg. Throughput (bps/Hz)	Jain Fairness Index	User Starvation (%)	Complexity
Proposed ($\alpha=0.5$)	98.7 \pm 3.2	0.992 \pm 0.004	0.0	Moderate
Max-C/I	112.4 \pm 4.1	0.42 \pm 0.07	18.3	Low
Round Robin	64.2 \pm 2.8	0.95 \pm 0.02	0.0	Very Low
PF-only	82.5 \pm 3.5	0.91 \pm 0.03	0.0	Low
HCAP-only	94.3 \pm 3.8	0.68 \pm 0.05	8.7	Moderate
Hybrid [16]	96.1 \pm 3.6	0.89 \pm 0.04	2.1	High

5.8 Statistical Significance and Robustness

All results were verified to be statistically significant using a Paired T-Test with a p-value of $p < 0.01$ for each of the 5 Baseline Scheduling Algorithms. The standard Deviation values reported in Table 3 have demonstrated that all Schedulers maintain consistent and stable Performance Metrics for differing Channel Conditions. The Proposed Hybrid Scheduler has demonstrated strong levels of Robustness over Heterogeneous Channel Conditions with Mixed Cell-Centre and Cell-Edge Users, maintaining an Average Fairness Level > 0.98 . Even when 30% of Channel conditions were very poor for all 30% of the user base.

Conclusion

A new hybrid scheduling structure has been developed for delivering multiple users simultaneously by integrating channel-aware prioritization, proportional fair weighting, cluster-based user grouping, and RZF precoding into a single architecture to create a new Hybrid PF Scheduler that will eliminate user starvation and improve the balance of fairness and throughput using dynamic clustering and quota enforcement with long term feature learning. Results from simulation indicate that the new Hybrid PF Scheduler always achieved maximum throughput, had very close to maximum Jain's fairness at all times, and produced consistently stable performance regardless of α -value or channel condition, without starving any users. The use of RZF precoding in the new Hybrid PF Scheduler reduced interference and improved spatial multiplexing efficiency. Thus, this new hybrid scheduling framework is scalable, fair, and spectrally efficient for 5G and beyond 5G systems, and potential research avenues include extending to multi-cell, machine learning based clustering methods and exploring intelligent surfaces for 6G.

References

- [1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update," White Paper, 2014.
- [2] 3GPP TR 36.913, "Requirements for Further Advancements for E-UTRA (LTE-Advanced)," v9.0.0, 2009.
- [3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.
- [4] T. Yoo and A. Goldsmith, "On the optimality of multiuser diversity with linear precoding," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1823–1853, Apr. 2006.
- [5] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.
- [6] K. Manolakis et al., *LTE Signal Processing for MIMO Systems*, Wiley, 2011.
- [7] A. Ahn and S. Kim, "Hybrid CQI estimation for multi-user MIMO in LTE systems," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1952–1955, Oct. 2013.
- [8] J. Zhang et al., "Channel quality indicator estimation in MU-MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1183–1191, Mar. 2013.
- [9] H. Huang et al., "Low-complexity SINR estimation for adaptive CQI reporting," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2722–2734, May 2014.
- [10] J. Kim and D. J. Love, "Joint beamforming and CQI quantization for MU-MIMO with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6303–6314, Dec. 2013.
- [11] T. Kim et al., "Lookup-table-based CQI estimation for MU-MIMO LTE systems," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1768–1771, Nov. 2012.
- [12] Y. Wang et al., "Adaptive offset CQI estimation for MU-MIMO LTE-A systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2671–2683, Apr. 2016.
- [13] R. Mundarath et al., "Blind CQI estimation for MU-MIMO in LTE," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6138–6148, Dec. 2013.
- [14] Z. Zhang et al., "MMSE-based CQI matching for MU-MIMO OFDM LTE systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 146–158, Jan. 2015.
- [15] F. Zhu et al., "Joint CQI optimization and CoMP-assisted MU-MIMO scheduling for TE-A," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1672–1684, May 2015.
- [16] M. Li et al., "Reinforcement Learning-Based Dynamic Resource Allocation for 5G MU-MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3347–3360, May 2020.
- [17] Y. Chen et al., "Deep Learning for Interference-Aware Scheduling in 5G MU-MIMO Networks," *IEEE Access*, vol. 9, pp. 45621–45634, 2021.
- [18] S. Park and H. Kim, "Neural CQI Estimation for Robust MU-MIMO Scheduling under Rapid Interference Variation," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1789–1793, Aug. 2020.
- [19] R. Kumar et al., "Compressed CQI Feedback for Massive MU-MIMO Using Autoencoders," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2456–2470, Mar. 2021.
- [20] L. Wang et al., "QoS-Aware Proportional Fair Scheduling for 5G MU-MIMO with Heterogeneous Traffic," *IEEE Trans. Mobile Comput.*, vol. 21, no. 6, pp. 2214–2228, Jun. 2022.
- [21] T. Zhang et al., "Federated Learning-Enabled Adaptive PF Scheduling for Mobile MU-MIMO Systems," *IEEE IoT Journal*, vol. 9, no. 18, pp. 17645–17658, Sept. 2022.
- [22] A. Gupta and N. Saxena, "Graph-Based User Clustering for Fairness-Aware MU-MIMO Scheduling," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1342–1346, Apr. 2021.
- [23] J. Lee et al., "Deep Reinforcement Learning for Joint Clustering and Precoding in MU-MIMO Systems," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8445–8460, Dec. 2021.
- [24] H. Wang et al., "Hybrid Precoding for mmWave MU-MIMO: A Deep Learning Approach," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2066–2080, Sept. 2020.

- [25] Zhu and M. Lei, "Optimized Regularization for RZF in Massive MU-MIMO with Imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 426–430, Feb. 2021.
- [26] C. Liu et al., "Meta-Learning for Adaptive Precoding in Time-Varying MU-MIMO Channels," *IEEE Trans. Signal Process.*, vol. 71, pp. 1125–1140, 2023.