Optimization Techniques for Multi-Relay Multi-Cell Massive MIMO NOMA Systems

Index Terms

NOMA, WMMSE, SSUM.

I. Introduction

It has been shown that non-orthogonal multiple access, also known as NOMA, has a greater SE than its orthogonal counterpart, OMA. NOMA does this by using overlap coding to multiplex the signals of the many users in the power domain. [1]–[3]. At the receiver, users of NOMA employ a technique known as successive interference cancellation (SIC), which helps to buffer the IUI impact.Integrating NOMA in mMIMO systems has lately been looked into in [2]–[8] due to the fact that doing so may greatly increase the system SE. Within the single cell mMIMO NOMA system, channel estimation and straightforward linear BS precoding techniques are used to efficiently minimise IUI [2]. When there are several cells involved, and there is a finite coherence interval, the pilot sequences are replayed in the cells that are close to each other. As a consequence, the channel estimates of users who share pilots get tainted, which brings to a decline in the quality of the SIC and, ultimately, the SE. [6], [8]. The paper [4] which is the base for this paper, works on closed form MRC SE of the Multi-Cell Multi-Relay NOMA Systems.

In the recent decade Massive MIMO(mMIMO) communications are widely popular due to their capability to provide substantial spectral efficiency (SE) and energy efficiency (EE) gains, higher reliability and lower end-to-end latency. Massive antenna arrays are installed at base stations (BSs) in cellular mMIMO systems, which maximises spatial diversity to multiplex tens of users sharing a single spectrum resource while increasing system SE and reliability. In order to limit the amount of inter-user interference in a multi-user mMIMO system, orthogonal multiple access (OMA) techniques are employed which reduce the amount of inter-user interference that occurs. These approaches distribute orthogonal time-frequency resources to each of the system's users (IUI). Even while OMA is very efficient at suppressing IUI, it lowers system SE when users observe channels with low quality. [1], [2].

The system coverage area, as well as its SE in [9], may be improved by the use of cooperative relaying, which involves the BS serving several customers by way of relays. a summary of the uses of NOMA in the relay-aided, single hop mMIMO systems and cooperative relaying was presented in Vaezi *et al.* in [9]. Recent research conducted in has focused on analysing multi-relay-aided uni-cell mMIMO NOMA systems [10]–[15].

The WMMSE algorithm was first proposed in [16] to optimize the weighted sum rate to design a linear transmit filter. The fundamental concept behind the approach is to transform the objective problem into a WMMSE maximisation, in which the weights are modified in an iterative fashion. The algorithm covers a wide range of issues, including sum rate use.

The sum SE and GEE metrics in the aforementioned optimization works [17]–[21], are stochastic functions of random fading channels and receiver noise. These works adopted a statistically approximated average (SAA) optimization approach, wherein an approximate closed-form expression of SE/GEE is first obtained by statistically averaging out the randomness due to channel fading, and the resultant deterministic expression is then optimized using an iterative algorithm. As a result SAA approach requires more memory and has higher computational and time complexity [22]. Stochastic optimization, in contrast to SAA, constructs an approximate sample objective function that captures the randomness in real-time and then solves it using iterative algorithm [22]. The stochastic optimization approach, therefore, has much lesser time complexity than SAA method [22], [23].

The base paper for this part of the thesis [4] uses the low-complexity inexpensive algorithm to solve the *deterministic* GEE and sum SE non-convex problems. The algorithm of [4] tries to convert the hard non-convex *deterministic* problems into a convex approximate using Quadratic Transform (QT) and Lagrangian Dual transform (LDT) which is solved through approach based on modified low-complexity(AMM).

In this work, we propose a generic framework to optimize the sum SE of a Multi-Cell, Multi-Relay Downlink mMIMO NOMA system, which solves both **stochastic** and **deterministic** sum SE non-convex problem with iterative closed form solutions and a very-low time complexity based on the WMMSE approach. This framework is then compared with the framework of the base paper [4]. The paper also extends its work to obtain **optimal deterministic precoder allocation** based on sum SE maximization which reduces the time and memory complexity of precoder formation and sum SE optimization.

II. System Model

The system model is acquired from [4]. and this section follows the order of the [4]. This is a briefed section and for any explanation revert back to the [4]. The paper investigates the downlink communication of an multi-cell (L) mMIMO NOMA system where multi-antenna(N) BS supply data to the single-antenna cellular users which forms clusters by deploying NOMA through half-duplex single-antenna AF relays which acts as the intermediate between BS and Users. These relays are installed such that the cluster of users forms around them and are positioned so that both the BS-relay and the user-relay channels include LoS and NLoS components. The cluster $\mathcal{U}_{lk}$ is the user from lth cell and R_{lk} relay .

The communication runs in the TDD mode, and its coherence interval of τ_c symbols is separated into the channel estimation (CE) phase, which consists of the transmission of pilot symbols, and the data transmission (DT) phase, which consists of τ and $(\tau_c$ - $\tau)$ symbols respectively. During the CE phase, the pilots are sent from the relays to the base stations (BS) as well as the Users for the purposes of transmit beamforming and receive forming, respectively. During the DT phase, the NOMA architecture is put into place, and users are serviced through relays.

The communication operates in TDD mode, with τ_c symbols coherence interval is divided into channel estimation (CE) by transmitting pilots and data transmission (DT) phases of τ and $(\tau_c$ - $\tau)$ symbols respectively. In the CE phase, the pilots are transmitted from the relays to both BS and the Users for transmit beamforming and receive forming respectively. In DT phase, NOMA is deployed and users are served via relays.

A. Channel Modeling

As there are two hops, one from BS-relay and the other from relay-BS which channel modeling. As it based on the paper [4], We just provide the mathematical expression for the channel.

For the modelling of BS-relay channel, as mentioned earlier that channel contain LoS and NLos components and which forms Rician channel. $\mathbf{h}_{lk}^{j}$ is the label given for the channel that runs from the l'th BS to R_{lk}.In light of this, the representation of it is as follows:

$$\mathbf{h}_{lk}^{l'} = \overline{\mathbf{h}}_{lk}^{l'} + (\mathbf{R}_{lk}^{l'})^{\frac{1}{2}} \mathbf{h}_{lk}^{l',\text{NLoS}} \tag{1}$$

In order to define the spatial correlation of the NLoS component, the matrix $\overline{\mathbf{R}}_{lk}^{k'}$ is used. The $\mathbf{h}_{lk}^{l',\text{NLoS}}$ is random variable which follows complex normal pdf. And the vector $\overline{\mathbf{h}}_{lk}^{l'}$ is the LoS component of the channel. The second-order statistics and its generation are explained in [4].

The relay-user channel is a scalar as both are single antennas. The channel pdf is a rician pdf as it contains Los and NLoS components. $g_{lk,n}^{l'k'}$ is the lable given for the channel scalar to the nth user $\mathcal{U}_{lk}$ from $R_{l'k'}$.In light of this, the representation of it is as follows:

$$g_{lk,n}^{l'k'} = \overline{g}_{lk,n}^{l'k'} + (\upsilon_{lk,n}^{l'k'})^{\frac{1}{2}} g_{lk,n}^{l'k',\text{NLoS}}. \tag{2}$$

The scalar $g_{lk,n}^{jk',\text{NLoS}}$ is random variable with complex normal distribution. In order to define the spatial correlation of the NLoS component, $\upsilon_{lk,n}^{l'k'}$ is used.

The CE phase: In this system model which follows [4], there is no direct link between users and BS as we considering they are far apart and causes huge shadowing and path loss and which leads to not having to estimate the BSs and user channel end-to-end. Therefore the relay helps to estimate the channel between user-relay and BS-relay, the local CSI. To estimate the local CSI, the K relays in the cluster R_{lk} transmits $K = \tau_p$ pilots (ψ_k) which are mutually orthogonal and with magnitude K to both lth BS and cluster of users $\mathcal{U}_{lk}$. Each cell's relays share the pilots, causing pilot contamination.

BS-Relay Estimate: The channel estimate of $\mathbf{h}_{lk}^{l}$ can be considered as a simple MMSE channel estimation between BS and relay. From [4], without using math, the MMSE estimate of channel $\mathbf{h}_{lk}^{l}$ is:

$$\widehat{\mathbf{h}}_{lk}^{l} = \overline{\mathbf{h}}_{lk}^{l} + \sqrt{p_p}\mathbf{R}_{lk}^{l}\mathbf{\Psi}_{lk}\left[\tilde{\mathbf{y}}_{k}^{l} - \overline{\mathbf{y}}_{k}^{l}\right], \tag{3}$$

where $\mathbf{\Psi}_{lk} = \left(\mathbf{I}_N + \tau p_p \sum\limits_{l'=1}^{L} \mathbf{R}_{l'k}^{l}\right)^{-1}$, $\overline{\mathbf{y}}_{k}^{l} = \sum\limits_{l'=1}^{L} \sqrt{p_p}\tau\overline{\mathbf{h}}_{l'k}^{l}$ and $\tilde{\mathbf{y}}_{k}^{l} = \sum_{l'=1}^{L} \sqrt{p_p}\tau\mathbf{h}_{l'k}^{l} + \mathbf{N}_{k}^{l\,*}$ and p_p is the pilot power and $\boldsymbol{N}_{k}^{l*}$ is AWGN of the lth BS from kth relay.

Relay-User Estimate: The channel estimate of $g_{lk,n}^{lk}$, is can be formed similar to the MMSE channel estimate of D2D section mentioned in the previous chapter, The MMSE estimate of $g_{lk,n}^{lk}$ is obtained from [4] is expressed as:

$$\widehat{g}_{lk,n}^{lk} = \overline{g}_{lk,n}^{lk} + \frac{\sqrt{p_p}\gamma_{lk,n}^{l'k}}{1 + \sum\limits_{l'=1}^{L} \tau p_p \gamma_{lk,n}^{l'k}}\left[\tilde{y}_{lk,n}^{p} - \overline{y}_{lk,n}^{p}\right], \quad \text{where } \overline{y}_{lk,n}^{p} = \sum_{l'=1}^{L} \sqrt{p_p}\tau\overline{g}_{lk,n}^{l'k}. \tag{4}$$

where $\tilde{y}_{lk,n}^{p} = \sum\limits_{l'=1}^{L} \sqrt{p_p}\tau g_{lk,n}^{l'k} + \mathbf{n}_{lk,n}^{p*}$

The DT phase: Two time slots are necessary for the transfer of data from the base station to the users through relay. The operation of these time slots is outlined as follows:

1) The 1st time slot: The First slot is used for BS-to-Relay transmission, and NOMA superposes user transmit signals in its cell, then precodes and sends them to relays. The equation for calculating NOMA precoded signal that is transmitted by the lth BS, with the $s_{lk,n}$ denoting nth users signal in U_{lk} obtained from [4] and is expressed as:

$$\mathbf{x}^l = \sum_{k=1}^{K} \mathbf{w}_{lk} \sum_{n=1}^{\mathcal{U}_{lk}} \sqrt{p_{lk,n}} s_{lk,n} \triangleq \sum_{k=1}^{K} \mathbf{w}_{lk} x_{lk}. \tag{5}$$

Here x_{lk} is the signal transmitted for the cluster $\mathcal{U}_{jk}$ to relay R_{lk} and the precoder $\mathbf{w}_{lk} \in \mathbb{C}^{N\times 1}$ is based on the channel estimate of $\mathbf{H}$ with unit norm due to power constraints. The received signal from the BS l to the kth relay in lth cell is given as

$$y_{R_{lk}} = \underbrace{\sum_{(l',k')} (\mathbf{h}_{lk}^{l'})^T \mathbf{w}_{l'k'} x_{l'k'}}_{\tilde{y}_{R_{lk}}} + z_{R_{lk}}. \tag{6}$$

The scalar $z_{R_{lk}}$ is the AWGN at the relay R_{lk}.

2) The 2nd time slot: The second slot is used for Relay-BS transmission after the signal received to Relay k. The received signal is transmitted again after the amplification of the signal by AF relay to the users of the cluster $\mathcal{U}_{lk}$. The transmit relay power is limited to q_{lk}, by $\overline{\mu}_{lk}$ which is the amplification factor. The $\overline{\mu}_{lk}$ cited from [4] is given as:

$$\overline{\mu}_{lk}^2 \mathbf{E}(|y_{R_{lk}}|^2) = q_{lk} \implies \overline{\mu}_{lk} = \sqrt{\frac{q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k} + 1 \right)}}. \tag{7}$$

where $p_{l'k'}$ is sum of the user transmit power from the cluster $\mathcal{U}_{l'k'}$. And the variables $\kappa_{l'k',lk}$, $\rho_{l'k,lk}$ are provided in the appendix A of [4]. The nth user's received NOMA signal is given by equation (11) in [4] is:

$$\begin{aligned} y_{lk,n} &= f_{lk,n} g_{lk,n}^{lk} \mu_{lk} \mathbf{h}_{lk}^{lT} \mathbf{w}_{lk} \sqrt{p_{lk,n}} s_{lk,n} + f_{lk,n} g_{lk,n}^{lk} \mu_{lk} \mathbf{h}_{lk}^{lT} \mathbf{w}_{lk} \sum_{n' \neq n}^{\mathcal{U}_{lk}} \sqrt{p_{lk,n'}} s_{lk,n'} \\ &+ \sum_{l' \neq l}^{L} f_{lk,n} g_{lk,n}^{lk} \mu_{lk} \mathbf{h}_{lk}^{l'T} \mathbf{w}_{l'k} x_{l',k'} + \sum_{l'=1}^{L} \sum_{k' \neq k}^{K} f_{lk,n} g_{lk,n}^{lk} \mu_{lk} \mathbf{h}_{lk}^{l'T} \mathbf{w}_{l'k'} x_{l'k'} + \sum_{l' \neq l}^{L} f_{k,n} g_{lk,n}^{l'k} \mu_{l'k} \tilde{y}_{R_{l'k}} \\ &+ \sum_{l'=1}^{L} \sum_{k' \neq k}^{K} f_{lk,n} g_{lk,n}^{l'k'} \mu_{l'k'} \tilde{y}_{R_{l'k'}} + \sum_{l'=1}^{L} \sum_{k'=1}^{K} f_{lk,n} g_{lk,n}^{l'k'} \mu_{l'k'} z_{R_{l'k'}} + f_{k,n} z_{k,n}. \end{aligned} \tag{8}$$

The explanation of the equation is provided in the system model section of the paper [4]. Relay-associated users execute SIC to minimise inter-relay interference. We suppose that these users are sorted by decreasing path loss. After cancelling IRI from users connected with the nth relay, using the SIC method [24], the first n-1 users' signals are treated as intrinsic intra-cluster interference [24]. $\mathbb{E}\left[\mathbf{h}_k^T \mathbf{w}_k\right]$ and $\hat{g}_{k,k,n}$ are used by the user to execute SIC which is expresse in the below equation:

$$f_{lk,n}g^{lk}_{lk,n}\mu_{lk}\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}\sum_{n'=1}^{n-1}\sqrt{p_{lk,n'}}s_{lk,n'}+\sum_{n'=n+1}^{\mathcal{U}_{lk}}\mu_{lk}\left[f_{lk,n}g^{lk}_{lk,n}\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}-f_{lk,n}\widehat{g}^{lk}_{lk,n}\mathbb{E}\left[\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}\right]\right]\sqrt{p_{lk,n'}}s_{lk,n'}.$$

Equation (12) from [4] is used to calculate the received NOMA signal at the n user of $\mathcal{U}_{lk}$ after the SIC and is stated as:

$$\begin{aligned}\bar{y}_{lk,n} &= f_{lk,n}g^{lk}_{lk,n}\mu_{lk}\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}\sqrt{p_{lk,n}}s_{lk,n}+f_{lk,n}g^{lk}_{lk,n}\mu_{lk}\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}\sum_{n'=1}^{n-1}\sqrt{p_{lk,n'}}s_{lk,n'} \\ &+\sum_{n'=n+1}^{\mathcal{U}_{lk}}\mu_{lk}\left[f_{lk,n}g^{lk}_{lk,n}\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}-f_{lk,n}\widehat{g}^{lk}_{lk,n}\mathbf{E}\left[\mathbf{h}^{l^T}_{lk}\mathbf{w}_{lk}\right]\right]\sqrt{p_{lk,n'}}s_{lk,n'} \\ &+\sum_{l'\neq l}^{L}f_{lk,n}g^{lk}_{lk,n}\mu_{lk}\mathbf{h}^{l'^T}_{lk}\mathbf{w}_{l'k}x_{l',k'}+\sum_{l'=1}^{L}\sum_{k'\neq k}^{K}f_{lk,n}g^{lk}_{lk,n}\mu_{lk}\mathbf{h}^{l'^T}_{lk}\mathbf{w}_{l'k'}x_{l'k'}+\sum_{l'\neq l}^{L}f_{k,n}g^{l'k}_{lk,n}\mu_{l'k}\tilde{y}_{R_{l'k}} \\ &+\sum_{l'=1}^{L}\sum_{k'\neq k}^{K}f_{lk,n}g^{l'k'}_{lk,n}\mu_{l'k'}\tilde{y}_{R_{l'k'}}+\sum_{l'=1}^{L}\sum_{k'=1}^{K}f_{lk,n}g^{l'k'}_{lk,n}\mu_{l'k'}z_{R_{l'k'}}+f_{k,n}z_{k,n}.\end{aligned} \tag{9}$$

BS combiners: In this work, we analyze the system performance for three combining schemes namely MRC, ZF and MMSE. The MMSE and ZF combiners are designed to cancel the inter- and intra-cell interference experienced by the user. These techniques, however, cannot cancel the Relay to User interference, of the system. The ZF and MMSE schemes, by utilizing the statistics and the channel realizations of the relay transmitters, mitigates the intra-cell, inter-cell of the relay interference. The ZF and MMSE combiners are given as

$$\overline{\mathbf{W}_l}=\begin{cases}\mathbf{H}_l, & \text{for MRC}\\ \mathbf{H}_l\left[\mathbf{H}_l^H\mathbf{H}_j\right]^{-1}, & \text{for ZF}\\ \left[\sum\limits_{l'=1}^{L}\mathbf{H}_{l'}\overline{\mathbf{P}}^{cd}_{l'}\mathbf{H}^H_{l'}+\mathbf{I}_M\right]^{-1}\mathbf{H}_j\overline{\mathbf{P}}_j, & \text{for MMSE.}\end{cases} \tag{10}$$

Here $\overline{\mathbf{W}^l}=[\mathbf{w}_{l1},\cdots,\mathbf{w}_{lK}]\in\mathbf{C}^{N\times K}$ denotes the set of combiners used by lth BS for the relays in the lth cell. The matrices $\mathbf{H}_l=[\mathbf{h}^l_{l1},\cdots,\mathbf{h}^l_{lK}]\in\mathbf{C}^{N\times K}$ denote the set of channels from K relays in lth cell to the lth BS respectively. Further $\mathbf{P}_{l'}=\text{diag}(p_{l'1},\cdots,p_{l'K})\in\mathbb{R}^{K\times K}_+$ where $p_{l'k}=\sum\limits_{n'=1}^{\mathcal{U}_{l'k}}p_{l'k,n'}$contains the sum of transmit powers of the users of cluster $\mathcal{U}_{l'k}$.

III. SE ANALYSIS

1) Ergodic SE Analysis: From the equation (12) of the base paper [4], The Ergodic sum SE of the system using Genie-bound with finite BS antennas of the system model, is given as

$$R^{\text{e}}_{\text{sum}}=\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}}\mathbb{E}\left[\frac{1}{2}\left(1-\frac{\tau}{\tau_c}\right)\log_2\left(1+\frac{\overline{\Delta}^i_{lk,n}}{\overline{\Omega}^i_{lk,n}}\right)\right],\ \text{where}\ \ \overline{\Omega}^i_{lk,n}=\sum_{m=1}^{5}\hat{I}^{(m)}_{lk,n}+1, \tag{11}$$

$$\hat{\Delta}_{lk,n}=\hat{A}_{lk,n}\overline{\mu}_{lk}^2 p_{lk,n},\ \hat{I}_{lk,n}^{(1)}=\sum_{n'=1}^{n-1}\hat{C}_{lk,n}^{(1)}\overline{\mu}_{lk}^2 p_{lk,n'},\ \hat{I}_{lk,n}^{(2)}=\sum_{n'=n+1}^{\mathcal{U}_{lk}}\hat{C}_{lk,n}^{(2)}p_{lk,n'}\overline{\mu}_{lk}^2,$$

$$\hat{I}_{lk,n}^{(3)}=\sum_{(l',k')\neq(l,k)}\sum_{n'=1}^{\mathcal{U}_{l'k'}}\hat{C}_{l'k',lk,n}^{(3)}p_{l'k',n'}\overline{\mu}_{l'k'}^2\ \hat{I}_{lk,n}^{(4)}=\sum_{(l',k')}\sum_{(l'',k'')\neq(l',k')}\sum_{n'=1}^{\mathcal{U}_{l''k''}}C_{\substack{l''k'',l'k',\\ lk,n}}^{(4)}p_{l''k'',n'}\overline{\mu}_{l'k'}^2$$

$$\hat{I}_{k,n}^{(5)}=\sum_{l'=1}^{L}\sum_{k'=1}^{K}\hat{C}_{l'k',lk,n}^{(5)}\overline{\mu}_{l'k'}^2,\ \text{and}\ \overline{\mu}_{lk}=\sqrt{\frac{q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'}+\sum\limits_{l'=1}^{L}\iota_{l'k,lk}p_{l'k}+1\right)}}.$$

The terms $\hat{A}_{lk,n}$, $\hat{C}_{lk,n}^{(1)}$, $\hat{C}_{lk,n}^{(2)}$, $\hat{C}_{l'k',lk,n}^{(3)}$, $\hat{C}_{\substack{l''k'',l'k',\\ lk,n}}^{(4)}$ and $\hat{C}_{l'k',lk,n}^{(5)}$ are functions of instantaneous channel realizations which are given as

$$\hat{A}_{lk,n}=|f_{lk,n}g_{lk,n}^{lk}\mathbf{h}_{lk}^{l^H}\mathbf{w}_{lk}|^2,\ \hat{C}_{lk,n}^{(1)}=|f_{lk,n}g_{lk,n}^{lk}\mathbf{h}_{lk}^{l^H}\mathbf{w}_{lk}|^2,\ \hat{C}_{lk,n}^{(2)}=|f_{lk,n}g_{lk,n}^{lk}\mathbf{h}_{lk}^{l^H}\mathbf{w}_{lk}-f_{lk,n}\hat{g}_{lk,n}^{lk}\mathbb{E}[\mathbf{h}_{lk}^{l^H}\mathbf{w}_{lk}]|^2,$$
$$\hat{C}_{l'k',lk,n}^{(3)}=|f_{lk,n}g_{lk,n}^{l'k'}\mathbf{h}_{l'k'}^{l'^H}\mathbf{w}_{l'k'}|^2,\ \hat{C}_{\substack{l''k'',l'k',\\ lk,n}}^{(4)}=|f_{lk,n}g_{lk,n}^{l'k'}\mathbf{h}_{l'k'}^{l''^H}\mathbf{w}_{l''k''}|^2,\ \hat{C}_{l'k',lk,n}^{(5)}=|f_{lk,n}g_{lk,n}^{l'k'}|^2 \tag{12}$$

2) UATF closed-form SE: From the derivation of the base paper [4], The UATF sum SE of the system with MRC precoder using Hard-bound technique with finite BS antennas of the system model, is given as:

$$R_{\text{sum}}^c=\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}}\frac{1}{2}\left(1-\frac{\tau}{\tau_c}\right)\log_2\left(1+\frac{\overline{\Delta}_{lk,n}^c}{\overline{\Omega}_{lk,n}^c}\right),\ \text{where}\ \overline{\Omega}_{lk,n}=\sum_{m=0}^{7}\overline{I}_{lk,n}^{(m)}+1, \tag{13}$$

$$\overline{\Delta}_{lk,n}=A_{lk,n}\overline{\mu}_{lk}^2 p_{lk,n},\ \overline{I}_{lk,n}^{(0)}=C_{lk,n}^{(0)}p_{lk,n}\overline{\mu}_{lk}^2,\ \overline{I}_{lk,n}^{(1)}=C_{lk,n}^{(1)}\sum_{n'=1}^{n-1}\overline{\mu}_{lk}^2 p_{lk,n'},\ \overline{I}_{lk,n}^{(2)}=C_{lk,n}^{(2)}\sum_{n'=n+1}^{\mathcal{U}_{lk}}p_{lk,n'}\overline{\mu}_{lk}^2,$$

$$\sum_{m=3}^{6}\overline{I}_{lk,n}^{(m)}=\sum_{(l',k')\neq(l,k)}\sum_{n'=1}^{\mathcal{U}_{l'k'}}C_{l'k',lk,n}^{(3)}p_{l'k',n'}\mu_{l'k'}^2+\sum_{(l',k')}\sum_{(l'',k'')\neq(l',k')}\sum_{n'=1}^{\mathcal{U}_{l''k''}}C_{\substack{l''k'',l'k',\\ lk,n}}^{(4)}p_{l''k'',n'}\overline{\mu}_{l'k'}^2,$$

$$\overline{I}_{k,n}^{(7)}=\sum_{l'=1}^{L}\sum_{k'=1}^{K}C_{l'k',lk,n}^{(5)}\overline{\mu}_{l'k'}^2,\ \text{and}\ \overline{\mu}_{lk}=\sqrt{\frac{q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'}+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}+1\right)}}. \tag{14}$$

Here $p_{lk}=\sum\limits_{n=1}^{\mathcal{U}_{lk}}p_{lk,n}$, $A_{lk,n}=\frac{\pi\upsilon_{lk,n}^{lk}\delta_{lk}}{4}\left[L_{1/2}\left(-\left|\overline{g}_{lk,n}^{lk}\right|^2/\upsilon_{lk,n}^{lk}\right)\right]^2$, with $L_{\frac{1}{2}}(\cdot)$ being Laguerre polynomial [25]. The terms $C_{lk,n}^{(0)}$, $C_{lk,n}^{(1)}$, $C_{lk,n}^{(2)}$, $C_{l'k',lk,n}^{(3)}$, $C_{\substack{l''k'',l'k',\\ lk,n}}^{(4)}$ and $C_{l'k',lk,n}^{(5)}$ are functions are provided in the base paper's [4] Appendix A.

IV. Preface to WMMSE Algorithm

The WMMSE algorithm was first proposed in [16] to optimize the weighted sum rate to design a linear transmit filter. The fundamental concept behind the approach is to transform the objective problem into a WMMSE maximisation, in which the weights are modified in an iterative fashion. The algorithm covers a wide range of issues, including sum rate use. The following fundamental equation serves as the foundation for the WMMSE method.

$$\text{SINR} = \max_{u} \gamma = \max_{u} \frac{1}{e} - 1 \tag{15}$$

where u can be considered as the receiving beamformer to the signal received and also acts as an auxiliary variable in the algorithm, and γ as the corresponding SINR with respect to the decoded signal and e is the *Mean Square Error*(MSE) of the decoded signal. This vital equation can be best explained by a special case of uplink single cell multi user SISO system. The signal received by the only BS is:

$$y_k = h_k \sqrt{p_k} s_k + \sum_{\substack{i=1 \\ i \neq k}}^{N} h_i \sqrt{p_i} s_i + n \tag{16}$$

where h is the channel matrix from Cellular User(CU) to the BS, s and p are the transmit symbol and power from the CU and n is the white Gaussian noise. The SE of the kth CU is:

$$\text{SE}_k^{CU} = C \log(1 + \text{SINR}_k^{CU}) = \frac{\Delta_k}{\Omega_k}, \text{ where}$$
$$\text{SINR}_k^{CU} = \frac{|h_k|^2 p_k}{\sum\limits_{\substack{i=1 \\ i \neq k}}^{N} |h_i|^2 p_i + 1}. \tag{17}$$

The received signal y_k at the BS of the CU k is decoded with a received beamformer u_k. The decoded signal is given as;

$$\hat{s_k} = u_k y_k = u_k h_k \sqrt{p_k} s_k + \sum_{\substack{i=1 \\ i \neq k}}^{N} u_k h_i \sqrt{p_i} s_i + u_k n, \tag{18}$$

the SINR of CU k of the decoded signal is:

$$\gamma_k = \frac{|u_k h_k|^2 p_k}{\sum_{\substack{i=1 \\ i \neq k}}^{N} |u_k h_i|^2 p_i + |u_k|^2}, \tag{19}$$

and the MSE of the decoded signal with the transmit symbol is

$$e_k = E(|\hat{s_k} - s_k|^2) = |1 - u_k h_k \sqrt{p_k}|^2 + \sum_{\substack{i=1 \\ i \neq k}}^{N} |u_k h_i|^2 p_i + |u_k|^2 \tag{20}$$

Upon minimizing the MSE e_k with respect to u_k. The equivalence one can observe is that:

$$\text{SINR}_k^{CU} = \frac{1}{\min\limits_{u_k} e_k} - 1 = \max_{u_k} \frac{1}{e_k} - 1 \tag{21}$$

which follows the equation mentioned earlier (15). In this part of the thesis, this equation serves as to manipulate the non-convex sum-rate utilization into pseudo concave functions for optimizations.

As the model becomes more complex and different, this relationship between MSE and SINR does not hold. Instead, it helps in understanding and providing relationship at a more general version which is provided in the below proposition.

Proposition 1: For any SINR with Δ and Ω as its numerator and denominator, can be reconstructed as,

$$\text{SINR} \;=\; \max_{u} \frac{1}{e} - 1 \text{ where,} \tag{22}$$

$$e \;=\; |1 - u\left(\sqrt{\Delta}\right)|^2 + |u|^2\left(\Omega\right) \tag{23}$$

Reason: If we look carefully the e_k variable in (20), it can be interpreted as,

$$e_k = |1 - u_k \underbrace{h_k\sqrt{p_k}}_{\sqrt{\Delta_k}}|^2 + |u_k|^2 \underbrace{\left(\sum_{\substack{i=1\\ i\neq k}}^{N} |h_i|^2 p_i + 1\right)}_{\Omega_k} \tag{24}$$

For the sum-rate maximization, the WMMSE algorithm uses the auxiliary variables for optimization which is presented in the following proposition.

Proposition 2: For any SINR with Δ and Ω as its numerator and denominator, the SE can be reconstructed as:

$$\text{SE} \;=\; \log\left(1 + \frac{\Delta}{\Omega}\right) \;=\; \underset{w,u}{\text{Maximize}}\; \log(w) - we \text{ where,} \tag{25}$$

$$e \;=\; |1 - u\left(\sqrt{\Delta}\right)|^2 + |u|^2\left(\Omega\right) \text{ and} \tag{26}$$

$$w \;=\; \frac{1}{e}. \tag{27}$$

Proof : Using the proposition 1, the SE can be rewritten as:

$$\text{SE} \;=\; \log\left(1 + \frac{\Delta}{\Omega}\right) \;=\; \log\left(\max_{u} \frac{1}{e}\right) \text{ where,} \tag{28}$$

$$e \;=\; |1 - u\left(\sqrt{\Delta}\right)|^2 + |u|^2\left(\Omega\right) \tag{29}$$

As log is a monotonically increasing function, the log function can be brought inside the max function. Also using a epigrapgh trick on the e variable, the above SE is reconstructed as:

$$\text{SE} \;=\; \log\left(1 + \frac{\Delta}{\Omega}\right) \;=\; \max_{u} \log\left(w\right) \text{ where,} \tag{30}$$

$$w \;\leq\; \frac{1}{e}. \tag{31}$$

As the reconstructed SE is concave function with respect to e variable. The strong Duality holds for the Lagrangian dual problem which is:

$$\mathcal{L}(w, \lambda) \;=\; \log\left(w\right) - \lambda\left(1 - \frac{1}{we}\right) \tag{32}$$

Maximizing the dual problem with respect to w provides the optimal value of λ which is:

$$\lambda* \;=\; we \tag{33}$$

Therefore Applying the optimal $\lambda*$ and the optimal value of w which is $\frac{1}{e}$ to the dual problem the SE can be reconstructed as:

$$\text{SE} = \underset{w,u}{\text{Maximize}} \ \log(w) - we \text{ where,} \tag{34}$$

$$e = |1 - u\left(\sqrt{\Delta}\right)|^2 + |u|^2 (\Omega) \text{ and} \tag{35}$$

$$w = \frac{1}{e}. \tag{36}$$

In the formulation a constant 1 is eliminated as it is the optimization and it has no responsibility in it. Hence the proof.

V. Optimization of SUM SE

The SUM SE is a characterization of the Channel capacity of the homogeneous system. Furthermore, maximizing the SUM SE results in the improved data transmission between the BS and the User. However, it is impossible to ensure that each user will have improved data transfer. In the multi-cell multi-relay NOMA system, the sum SE is maximized by altering the power allocated to each user from the BS. In the sections that follow, we will attempt to maximise the sum SE through power allocation for UATF V-A and Ergodic V-B system models by iteratively optimising using closed-form solutions. *Previous works on NOMA mMIMO systems have not worked on optimizing the sum SE using iterative closed form solutions.* This work can be extended to optimizing GEE and WSEE which is a better characterization and also acts as the trade-off between Channel Capacity and the power consumption in the homogeneous and heterogeneous network respectively.

A. UATF: SUM SE problem formulation and Optimization

The optimization of sum SE consists of series of sub-problems which are solved iteratively with closed-from solutions. The objective function of the sum SE optimzation is:

$$\mathbf{P1}: \quad \underset{\mathbf{P},\mathbf{Q}}{\text{Maximize}} \ R^c_{\text{sum}} \tag{37}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l \ \forall l, \ q_{lk} \leq Q_{lk} \ \forall l, \ \forall k \text{ and } p_{lk,n}, q_{lk} \geq 0 \tag{38}$$

where P_l is the maximum transmit power of the base station of lth cell, Q_{lk} is the maximum power transmitted from relay R_{lk} and $\mathbf{P} = [\mathbf{p}_1, .., \cdots, \mathbf{p}_L] \in \mathbb{R}^{L \times \mathcal{U}_{lk} K}$ and $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_L] \in \mathbb{R}^{K \times L}$ where $\mathbf{p}_l = [p_{l1,1}, \cdots, p_{lK,\mathcal{U}_{lk}}]$ and $\mathbf{q}_l = [q_{l1}, \cdots, q_{lK}]$ and also

$$R^c_{\text{sum}} = \frac{1}{2}\left(\frac{\tau_c - \tau_p}{\tau_c}\right)\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}} \log_2\left(1 + \frac{\overline{\Delta}^c_{lk,n}}{\overline{\Omega}^c_{lk,n}}\right). \tag{39}$$

The constant term $\frac{1}{2}\left(\frac{\tau_c-\tau_p}{\tau_c}\right)$ is eliminated as it is irrelevant in the optimization. Also. the $\log_2$ function is converted into natural algorithm (*ln*) which removes any constant appearing while performing differentiation. The optimization of sum SE comes under the ambit of multi-ratio(sum of log-ratios) fractional programming framework and are therefore challenging-convex problem.

The problem can be solved by novel optimization framework which approximates the non-convex functions into pseudo concave function at a point using the relationship between SINR and *mean square error*(MSE) where the algorithm knowingly termed as WMMSE algorithm. The algorithm converts non-convex hard problem like in **P1** and translates into pseudo concave function, which can be maximized using simple iterative closed form solutions.

For this optimization problem, instead of using $_{lk,n}$ as an optimization variable, we use $\bar{\mu}_{lk}$ as the replacement optimization variable, one can understand that this is an important replacement to attain closed form solutions from the solution for the following sub-problems. As $\hat{\mu}_{lk}$ depends on the power variables too, the constraints for $\bar{\mu}_{lk}$ are iteratively updated to satisfy all the constraints of the optimization. Therefore the problem **P1** is restructured as:

$$\mathbf{P2}: \quad \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Maximize}} \sum_{(l,k)} \sum_{n=1}^{\mathcal{U}_{lk}} \log\left(1+\frac{\overline{\Delta}^c_{lk,n}}{\overline{\Omega}^c_{lk,n}}\right) \tag{40}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ \bar{\mu}_{lk}\leq\hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'}+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k}+1\right)}}. \tag{41}$$

With the restructured problem formulation **P2**, the optimization can proceed with modified WMMSE algorithm. Using Proposition 2, one can equivalently reconstruct the write the SE of the nth user of the cluster $\mathcal{U}_{lk}$ as:

$$\log(1+\text{SINR}^c_{lk,n}) \;=\; \log(1+\frac{\overline{\Delta}^c_{lk,n}}{\overline{\Omega}^c_{lk,n}}) \;=\; \max_{u^c_{lk,n},w^c_{lk,n}} \frac{1}{w^c_{lk,n}} - w^c_{lk,n}e^c_{lk,n} \text{ where,} \tag{42}$$

$$e^c_{lk,n} = |1-u^c_{lk,n}\sqrt{\overline{\Delta}^c_{lk,n}}|^2+|u^c_{lk,n}|^2\overline{\Omega}^c_{lk,n} \tag{43}$$

The important aspect of $e_{lk,n}$ is that the equation is concave in nature with respect to transmit power $\mathbf{P}$ and relay amplitude factor $\boldsymbol{\mu}$. Using the equation (42) to the Problem **P2**, the problem is reconstructed as:

$$\textbf{P3}: \quad \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Maximize}} \sum_{(l,k)} \sum_{n=1}^{\mathcal{U}_{lk}} \underset{u^c_{lk,n}, w^c_{lk,n}}{\text{Maximize}} \ \log\left(w^c_{lk,n}\right) - w^c_{lk,n} e^c_{lk,n} \tag{44}$$

$$s.t \quad \sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l \, \forall l, \ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k} + 1\right)}}.. \tag{45}$$

The auxiliary variables $u^c_{lk,n}, w^c_{lk,n}$ can be pushed out of the summation as the each $u^c_{lk,n}$ and $w^c_{lk,n}$ does not have any inter-dependencies for all l,k,n. Therefore the problem **P3** is reformulated as:

$$\textbf{P4}: \quad \underset{\mathbf{P},\mathbf{U},\mathbf{W},\boldsymbol{\mu}}{\text{Maximize}} \sum_{(l,k)} \sum_{n=1}^{\mathcal{U}_{lk}} \log\left(w^c_{lk,n}\right) - w^c_{lk,n} e^c_{lk,n} \tag{46}$$

$$s.t \quad \sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l \, \forall l, \ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k} + 1\right)}}. \tag{47}$$

Here the matrix $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_L] \in \mathbb{R}^{L \times \mathcal{U}_{lk} K}$ where $\mathbf{u}_l = [u^c_{l1,1}, \cdots, u^c_{lK,\mathcal{U}_{lk}}]$ and $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_L] \in \mathbb{R}^{L \times \mathcal{U}_{lk} K}$ where $\mathbf{w}_l = [w^c_{l1,1}, \cdots, w^c_{lK,\mathcal{U}_{lk}}]$. Now, expanding the auxiliary variable $e^t_{lk,n}$ by expanding the Numerator and Denominator of the SINR becomes:

$$\begin{aligned} e^c_{lk,n} &= 1 + |u^c_{lk,n}|^2 A_{lk,n} \bar{\mu}^2_{lk} p_{lk,n} - 2(u^c_{lk,n} \sqrt{A_{lk,n} \bar{\mu}^2_{lk} p_{lk,n}}) \\ &+ |u^c_{lk,n}|^2 \Bigg(C^{(0)}_{lk,n} \bar{\mu}^2_{lk} p_{lk,n} + \sum_{n'=1}^{n-1} C^{(1)}_{lk,n,n'} \bar{\mu}^2_{lk} p_{lk,n'} + \sum_{n'=n+1}^{\mathcal{U}_{lk}} C^{(2)}_{lk,n,n'} p_{lk,n'} \bar{\mu}^2_{lk} + \sum_{(l',k') \neq (l,k)} \sum_{n'=1}^{\mathcal{U}_{l'k'}} C^{(3)}_{l'k',lk,n} p_{l'k',n'} \mu^2_{l'k'} \\ &+ \sum_{(l',k')} \sum_{(l'',k'') \neq (l',k')} \sum_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)}_{\substack{l''k'',l'k',\\ lk,n}} p_{l''k'',n'} \bar{\mu}^2_{l'k'} + \sum_{l'=1}^{L} \sum_{k'=1}^{K} C^{(5)}_{l'k',lk,n} \bar{\mu}^2_{l'k'} + 1 \Bigg) \end{aligned} \tag{48}$$

Grouping the above equation transmit power-wise and multiplying with $w^t_{lk,n}$, the equation becomes:

$$w^c_{lk,n} e^c_{lk,n} = const(w^c_{lk,n}) + \sum_{(l',k')} \sum_{n'=1}^{\mathbb{U}_{lk}} \alpha^c_{l'k',n',lk,n} p_{l'k',n'} + \beta^c_{lk,n} \sqrt{p_{lk,n}}. \tag{49}$$

All the constants including the amplitude factor $\boldsymbol{\mu}$ are combined and termed as α and β and are termed as:

$$\alpha^c_{l'k',n',lk,n} = w^c_{lk,n}|u^c_{lk,n}|^2 * \begin{cases} A_{lk,n}\overline{\mu}^2_{lk} + C^{(0)}_{lk,n}\overline{\mu}^2_{lk} + \sum\limits_{l''k''\neq lk} C^{(4)}_{\substack{lk,l''k'',\\ lk,n}}\overline{\mu}^2_{l''k''} & (l'k',n') = (l,k,n) \\ C^{(1)}_{lk,n,n'}\overline{\mu}^2_{lk} + \sum\limits_{l''k''\neq lk} C^{(4)}_{\substack{lk,l''k'',\\ lk,n}}\overline{\mu}^2_{l''k''} & (l'k') = (l,k)\ \&\ n' \leq n-1 \\ C^{(2)}_{lk,n,n'}\overline{\mu}^2_{lk} + \sum\limits_{l''k''\neq lk} C^{(4)}_{\substack{lk,l''k'',\\ lk,n}}\overline{\mu}^2_{l''k''} & (l'k') = (l,k)\ \&\ n' \geq n+1 \\ C^{(3)}_{l'k',lk,n}\overline{\mu}^2_{lk} + \sum\limits_{l''k''\neq l'k'} C^{(4)}_{\substack{l'k',l''k'',\\ lk,n}}\overline{\mu}^2_{l''k''} & (l'k') \neq (l,k) \end{cases} \tag{50}$$

$$\beta^c_{lk,n} = -2w^c_{lk,n}|u^c_{lk,n}|\sqrt{A_{lk,n}\overline{\mu}^2_{lk}} \tag{51}$$

Therefore, the objective function of problem **P4** becomes:

$$\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}} \log(w^c_{lk,n}) - w^c_{lk,n}e^c_{lk,n} =$$

$$\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}}\left(const(\mathbf{W}) - \sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}} \alpha^c_{l'k',n',lk,n}p_{l'k',n'} + \beta^c_{lk,n}\sqrt{p_{lk,n}}\right) \tag{52}$$

Therefore the problem formulation **P4** is reconstructed as:

$$\mathbf{P5}: \quad \underset{\mathbf{P},\mathbf{U},\mathbf{W},\boldsymbol{\mu}}{\text{Maximize}} \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}}\left(const(\mathbf{W}) - \sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}} \alpha^c_{l'k',n',lk,n}p_{l'k',n'} + \beta^c_{lk,n}\sqrt{p_{lk,n}}\right) \tag{53}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'} + \sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k} + 1\right)}}. \tag{54}$$

Now, after formulating the objective function **P5**, all the optimizing variables $\mathbf{P}, \mathbf{U}, \mathbf{W}, \boldsymbol{\mu}$ are iteratively optimized. Firstly, the auxiliary variable $u^c_{lk,n}$ is optimized by first order differentiation of $e^c_{lk,n}$ and equating to zero which is:

$${u^c_{lk,n}}^* = \frac{\sqrt{\overline{\Delta}^c_{lk,n}}}{\overline{\Omega}^c_{lk,n}} \tag{55}$$

The optimization of the variable $w^c_{lk,n}$ can be easily known as it acts as the auxiliary variable for $e_{lk,n}$. Therefore, the optimal $w^c_{lk,n}$ is:

$$w^{c*}_{lk,n} = \frac{1}{e^{c*}_{lk,n}} \tag{56}$$

The only optimizing variables left are the transmit power and relay amplitude factor. As mentioned earlier that $e^c_{lk,n}$ is concave in nature with respect to the optimizing variables. *The variable*

$e^c_{lk,n}$ *couldn't have been concave in nature, if continued with variables* $\mathbf{P}$ *and* $\mathbf{Q}$. Therefore, with the auxiliary variables $w^{c*}_{lk,n}$ and $u^{c*}_{lk,n}$ acting as fixed point equations, the problem **P5** acts as a pseudo concave optimization with respect to the $\mathbf{P}$ and $\boldsymbol{\mu}$ and provides a optimal value. Therefore, the optimal $p_{lk,n}$ are obtained through first-order differentiation of the objective function **P5** and equating to zero which provides:

$$p^*_{lk,n} = \left\{ \frac{\boldsymbol{\beta}^c_{lk,n}}{\left(\sum\limits_{(l',k')} \sum\limits_{n'=1}^{\mathbb{U}_{lk}} \boldsymbol{\alpha}^c_{lk,n,l'k',n'} \right) + \lambda^*_l} \right\}^2 \tag{57}$$

where λ^*_l is an internal auxiliary variable which can optimized through bisection algorithm which helps to satisfy $\sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} p^*_{lk,n} \leq P_l$. Similarly, if the expanding the auxiliary variable $e^t_{lk,n}$ by expanding the Numerator and Denominator of the SINR and group with respect to the μ_{lk} and multiplying by $w^c_{lk,n}$ becomes:

$$w^c_{lk,n} e^c_{lk,n} = const(w^c_{lk,n}) + \sum_{(l',k')} \sum_{n'=1}^{\mathcal{U}_{l'k'}} \gamma^c_{l'k'n',lk} \mu^2_{l'k'} + \omega^t_{lk} \mu_{lk}. \tag{58}$$

All the constants including the transmit power $\mathbf{P}$ are combined and termed as ω and γ and are termed as:

$$\gamma^c_{l'k'n',lk} = w^c_{lk,n} |u^c_{lk,n}|^2 * \begin{cases} A_{lk,n} p_{lk,n} + C^{(0)}_{lk,n} p_{lk,n} + C^{(5)}_{lk,lk,n} + \sum\limits_{l''k'' \neq lk} \sum\limits_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)}_{\substack{l''k'',lk,\\ lk,n}} p_{l''k'',n'} & (l'k',n') = (l,k,n) \\ C^{(1)}_{lk,n,n'} p_{lk,n'} + C^{(5)}_{lk,lk,n} + \sum\limits_{l''k'' \neq lk} \sum\limits_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)}_{\substack{l''k'',lk,\\ lk,n}} p_{l''k'',n'} & \substack{(l'k')=(l,k)\\ n' \leq n-1} \\ C^{(2)}_{lk,n,n'} p_{lk,n'} + C^{(5)}_{lk,lk,n} + \sum\limits_{l''k'' \neq lk} \sum\limits_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)}_{\substack{l''k'',lk,\\ lk,n}} p_{l''k'',n'} & \substack{(l'k')=(l,k)\\ n' \geq n+1} \\ \sum\limits_{n'=1}^{\mathcal{U}_{l'k'}} C^{(3)}_{l'k',lk,n} p_{lk,n'} + C^{(5)}_{l'k',lk,n} + \sum\limits_{l''k'' \neq l'k'} \sum\limits_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)}_{\substack{l''k'',l'k',\\ lk,n}} p_{l''k'',n'} & (l'k') \neq (l,k) \end{cases} \tag{59}$$

$$\omega^c_{lk} = -2 w^c_{lk,n} |u^c_{lk,n}| \sqrt{A_{lk,n} p_{lk,n}} \tag{60}$$

Therefore the problem formulation **P4** is reconstructed as:

$$\textbf{P6}: \quad \underset{\mathbf{P},\mathbf{U},\mathbf{W},\boldsymbol{\mu}}{\text{Maximize}} \sum_{(l,k)} \sum_{n=1}^{\mathbb{U}_{lk}} \left(const(\mathbf{W}) - \sum_{(l',k')} \gamma^c_{l'k',lk} \mu^2_{lk} + \omega^t_{lk} \mu_{lk} \right) \tag{61}$$

$$s.t \quad \sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l \, \forall l, \;\; \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k} + 1 \right)}}. \tag{62}$$

Using the problem **P6**, the $\mu_{lk,n}$ is optimized through first-order differentiation and equated to zero as:

$$\bar{\mu}^*_{lk} = \min\left\{ \frac{\boldsymbol{\omega}^c_{lk}}{\sum\limits_{(l',k')} \sum\limits_{n'=1}^{\mathcal{U}_{lk}} \gamma^c_{lkn,l'k'}}, \hat{\mu_{lk}} \right\} \tag{63}$$

where,

$$\hat{\mu_{lk}} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p^*_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p^*_{l'k} + 1 \right)}}. \tag{64}$$

We observe that (57) and (63) are fixed-point equations, where the RHS expression depend themselves on $p_{lk,n}$ and μ_{lk}. The constraints $\hat{\mu_{lk}}$ are iteratively updated with the optimal power values. We therefore develop an iterative algorithm to solve the problem **P1** by starting from a feasible transmit power and relay transmit power and iteratively updating the auxiliary variables and transmit powers and relay amplitude factors with the solutions provided below. The resulting formal procedure to solve **P1** in (37) is provided in Algorithm 1.

Algorithm 1: sum SE maximization using Deterministic WMMSE approach

Input: Given $\epsilon > 0$, the max iterations N and max power P_l for UE U_{lk} and max power Q_{lk} for the relay. Calculate the initial values $p_{lk,n}, \mu_{lk}$ with random power allocation for all relay and users i.e., $\mathbf{p}^{(1)}_{lk,n} \sim \mathbb{U}[0, P^{max}_l]$ and $\mu^{(1)}_{lk} \sim \mathbb{U}[0, Q^{lk}_{max}]$

Output: $p^*_{lk,n}$ and μ^*_{lk}.

for $n \leftarrow 1$ ***to*** N **do**

  Given a feasible $p^{(i)}_{lk,n}$ and $\mu^{(i)}_{lk}$, update auxiliary variable $u_{lk,n}$ using (55)

  Update the auxiliary variable $w_{lk,n}$ using (56)

  Compute $p^{(n+1)}_{jk}$ using (57)

  Update the $\hat{\mu}_{lk}$ constraint variable using (64)

  Compute $\mu^{(i+1)}_{lk}$ using (63)

  Do until convergence **if** $\sum\limits_{l=1}^{L} \sum\limits_{l=1}^{K} \sum\limits_{n=1}^{\mathbb{U}_{lk}} \left(p^{(i)}_{lk,n} - p^{(i-1)}_{lk,n} \right) < \epsilon$ **then**

  break.

return $\mathbf{p}^*, \mathbf{mu}^*$

B. *Ergodic: SUM SE problem formulation and optimization*

In the previous section of this chapter, we solved a deterministic SE issue by making use of a UATF SE expression. Although calculating these expectations in closed-form for MRC is a simple process, doing so for the ZF and MMSE combining schemes is a non-trivial endeavour. Inorder to solve it, with using deterministic algorithm, the optimum power allocation scheme for ZF and MMSE schemes makes use of statistical averages (see equation(15) of [4]), which necessitates the gathering of a large number of random channel realisations prior to the updating of the transmit powers. The deterministic sum SE maximisation has a larger computational complexity, and as a result, it calls for a greater amount of memory, as well as longer time to store the samples. We are going to now recast the sum SE problem that was in P1 as a stochastic optimization problem and then optimize it using a low-complexity stochastic modified WMMSE framework. The optimization process uses both the stochastic sequential upper-bound minimization technique (SSUM) algorithm [26] and the weighted minimum mean squared error (WMMSE) algorithm. This will help us lower the memory required as well as the computing complexity. The summary of this section is that, after every realization, the surrogate function is formed with instantaneous sum SE and the optimizing variables are updated. The problem formulation follows a pattern quite similar to that of the section before it. The following is the objective function for the optimization of the ergodic sum SE:

$$\mathbf{P1}_{sto}: \qquad \underset{\mathbf{P},\mathbf{Q}}{\text{Maximize}}\ \mathbf{E}\left[g(\mathbf{P},\mathbf{Q},\mathcal{F})\right] \triangleq R^{e}_{sum} \tag{65}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ q_{lk} \leq Q_{lk}\ \forall l,\ \forall k \text{ and } p_{lk,n}, q_{lk} \geq 0. \tag{66}$$

Here $g(\mathbf{P},\mathbf{Q},\mathcal{F})$ denotes the instantaneous sum SE of the system, with $\mathcal{F} = \{\mathbf{h}^{l'}_{lk}, g^{l'k'}_{lk,n}\}, \forall l, k, l', k', n$. It is defined as

$$g(\mathbf{P},\mathbf{Q},\mathcal{F}) = \frac{1}{2}\left(\frac{\tau_c - \tau_p}{\tau_c}\right)\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}} \log_2\left(1 + \frac{\overline{\Delta}^{i}_{lk,n}(\mathcal{F})}{\overline{\Omega}^{i}_{lk,n}(\mathcal{F})}\right) \tag{67}$$

where $\overline{\Delta}^{i}_{lk,n}(\mathcal{F}), \overline{\Omega}^{i}_{lk,n}(\mathcal{F})$ are instantaneous SINR, Numerator of SINR, Denominator of the SINR of the nth user in $\mathcal{U}_{lk}$. The expectation is due to the random channels $\mathcal{F}$ generated. While reconstructing **P2** of the deterministic optimization it is mentioned that the optimization variable q_{lk} is replaced by $\bar{\mu}_{lk}$ inorder to attain the concavity of the sub-objective functions. And the constraint $\hat{\mu}_{lk}$ is iteratively updated to satisfy the constraint of the relay transmit power. Considering the aforementioned reasons, and also eliminating the constants and converting $\log_2$ into log function the problem is reconstructed as:

$$\mathbf{P2}_{sto}: \qquad \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Maximize}}\ \mathbf{E}\left[\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}} \log_2\left(1 + \frac{\overline{\Delta}^{i}_{lk,n}((\mathcal{F})}{\overline{\Omega}^{i}_{lk,n}(\mathcal{F})}\right)\right] \tag{68}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk}p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk}p_{l'k} + 1\right)}}. \tag{69}$$

To solve this stochastic non-convex optimization problem we propose a modified SSUM-WMMSE algorithm to solve $\mathbf{P2}_{sto}$. Using Proposition 2, we can reconstruct the instantaneous SE and therefore the problem $\mathbf{P2}_{sto}$ as:

$$\mathbf{P3}_{sto}: \quad \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Maximize}} \lim_{T\Rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}\left[\underset{\mathbf{U}^t,\mathbf{W}^t}{\text{Maximize}} \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} w^t_{lk,n}e^t_{lk,n} - \log(w^t_{lk,n})\right] \tag{70}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'} + \sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k}+1\right)}}. \tag{71}$$

where,

$$e^t_{lk,n} = |1 - u^t_{lk,n}\sqrt{\overline{\Delta}^i_{lk,n}(\mathcal{F}^t)}|^2 + |u^t_{lk,n}|^2\,\overline{\Omega}^i_{lk,n}(\mathcal{F}^t) \tag{72}$$

As there are no inter-dependence in the auxiliary variables $w^t_{lk,n}$ and $u^t_{lk,n}$ they are grouped together into matrices $\mathbf{W},\mathbf{U}$ respectively. As we know that $w^t_{lk,n}$ is the auxiliary variable to epigraph the $e^i_{lk,n}$ variable its optimal value is expressed as:

$$w^{t*}_{lk,n} = \frac{1}{e^{t*}_{lk,n}} \tag{73}$$

Minimizing $e^t_{lk,n}$ auxiliary variable under $u^t_{lk,n}$ through first order differentiation and equating to zero provides the optimal value of:

$$u^{t*}_{lk,n} = \frac{\sqrt{\overline{\Delta}^i_{lk,n}(\mathcal{F}^t)}}{\overline{\Omega}^i_{lk,n}(\mathcal{F}^t)} \tag{74}$$

After maximizing the inner optimization $(\mathbf{U}^t,\mathbf{W}^t)$ for each instantaneous SE, the problem $\mathbf{P3}_{sto}$ is deconstructed as:

$$\mathbf{P4}_{sto}: \quad \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Maximize}} \lim_{T\Rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \log(w^{t**}_{lk,n}) - w^{t**}_{lk,n}e^{t**}_{lk,n}\right] \tag{75}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'} + \sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k}+1\right)}}. \tag{76}$$

where,

$$e^{t**}_{lk,n} = |1 - u^{t**}_{lk,n}\sqrt{\overline{\Delta}^i_{lk,n}(\mathcal{F}^t)}|^2 + |u^{t**}_{lk,n}|^2\,\overline{\Omega}^i_{lk,n}(\mathcal{F}^t) \tag{77}$$

and the auxiliary variables $u^{t**}_{lk,n}, w^{t**}_{lk,n}$ are the optimized values after inner optimization. The problem $\mathbf{P4}_{sto}$ is the outer optimization $(\mathbf{P},\boldsymbol{\mu})$ which is a case of stochastic optimization problem and is solved by the SSUM algorithm. The algorithm is summarized below.

Summary of the SSUM Algorithm: In the paper [26], the proposed SSUM algorithm is that, at each iteration t, a new realization of channel $\mathcal{F}^{\sqcup} = \mathbf{H}^t$,$\mathbf{g}^t$ are obtained and the surrogate function is update to tth realization $(\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^t, \boldsymbol{\mu}^t))$. Using this surrogate function, the optimization variables are updated which is expressed as:

$$\mathbf{P5}_{sto} : [\mathbf{P}^{t+1}, \boldsymbol{\mu}^{t+1}] \Leftarrow \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Argmax}} \ \frac{1}{t}\left[\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^t, \boldsymbol{\mu}^t)\right] \tag{78}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} p_{lk,n} \leq P_l \, \forall l, \ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k} + 1\right)}}. \tag{79}$$

Formation of Surrogate function: At each iteration t, a new realization of channel $\mathcal{F}^{\sqcup} = \mathbf{H}^t$,$\mathbf{g}^t$ are obtained and the surrogate function upto $t-1$th realization ($\mathbf{R}^{1:t-1}(\mathbf{P}, \boldsymbol{\mu})$) are updated with the convex approximate function ($\tilde{R}^i_{sum}(\mathbf{P}, \boldsymbol{\mu}, \mathcal{F}^t)$) of the instantaneous sum SE to surrogate function upto tth realization ($\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu})$) through recursive surrogate function from [27], which also guarantee the convergence of the algorithm and expressed as:

$$\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^t, \boldsymbol{\mu}^t) = \frac{1}{t}\left[\tilde{R}^i_{sum}(\mathbf{P}, \boldsymbol{\mu}, \mathcal{F}^t)\right] + \frac{t-1}{t}\left[\mathbf{R}^{1:t-1}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^{t-1}, \boldsymbol{\mu}^{t-1})\right] \tag{80}$$

Coming back to the stochastic optimization of Ergodic sum SE, we need to find the surrogate function and the convex approximate at the t realization and is found using the following proposition.

Proposition 3: The objective function in $\mathbf{P4}_{sto}$ can act as a convex approximate for the recursive surrogate function. In mathematical way,

$$\tilde{R}^i_{sum}(\mathbf{P}, \boldsymbol{\mu}, \mathcal{F}^t) = \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \log(w^{t**}_{lk,n}) - w^{t**}_{lk,n} e^{t**}_{lk,n} \tag{81}$$

Proof: Expanding the auxiliary variable $e^{t**}_{lk,n}$ by expanding the Numerator and Denominator of the SINR becomes:

$$\begin{aligned} e^t_{lk,n} &= 1 + |u^{t**}_{lk,n}|^2 \hat{A}^t_{lk,n} \tilde{\mu}^2_{lk} p_{lk,n} - 2Re(u^{t**}_{lk,n}\sqrt{\hat{A}^t_{lk,n}\tilde{\mu}^2_{lk} p_{lk,n}}) \\ &+ |u^{t**}_{lk,n}|^2 \Bigg(\sum_{n'=1}^{n-1} \hat{C}^{(1)^t}_{lk,n,n'} \tilde{\mu}^2_{lk} p_{lk,n'} + \sum_{n'=n+1}^{\mathcal{U}_{lk}} \hat{C}^{(2)^t}_{lk,n,n'} p_{lk,n'} \tilde{\mu}^2_{lk} + \sum_{(l',k')\neq(l,k)} \sum_{n'=1}^{\mathcal{U}_{l'k'}} \hat{C}^{(3)^t}_{l'k',lk,n} p_{l'k',n'} \mu^2_{l'k'} \\ &\qquad + \sum_{(l',k')} \sum_{(l'',k'')\neq(l',k')} \sum_{n'=1}^{\mathcal{U}_{l''k''}} C^{(4)^t}_{\substack{l''k'',l'k',\\ lk,n}} p_{l''k'',n'} \tilde{\mu}^2_{l'k'} + \sum_{l'=1}^{L}\sum_{k'=1}^{K} \hat{C}^{(5)^t}_{l'k',lk,n} \tilde{\mu}^2_{l'k'} + 1 \Bigg) \end{aligned} \tag{82}$$

Grouping the above equation transmit power-wise and multiplying with $w^{t**}_{lk,n}$, the equation becomes:

$$w^{t**}_{lk,n}e^{t**}_{lk,n} = const(w^{t**}_{lk,n}) + \sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}} \alpha^{t}_{l'k',n',lk,n}p_{l'k',n'} + \beta^{t}_{lk,n}\sqrt{p_{lk,n}}. \tag{83}$$

All the constants including the amplitude factor $\boldsymbol{\mu}$ are combined and termed as α and β and are similar to (59). Therefore, the objective function of problem $\mathbf{P4}_{sto}$ becomes:

$$\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \log(w^{t**}_{lk,n}) - w^{t**}_{lk,n}e^{t**}_{lk,n} =$$

$$\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \left(const(\mathbf{W}) - \sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}} \alpha^{t}_{l'k',n',lk,n}p_{l'k',n'} + \beta^{t}_{lk,n}\sqrt{p_{lk,n}} \right) \tag{84}$$

If the equation is carefully noticed, the function is concave in transmit power variables $\mathbf{P}$ and also the $t+1$th channel realization will also have the exact objective sum SE function. *The aforementioned reasons has proved that the instantaneous this objective function can be considered as the instantaneous convex approximate for the surrogate function.*

Hence the proof.

By induction method one can understand that the surrogate function upto tth realization $\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^t, \boldsymbol{\mu}^t)$ will have similar kind of constants with them as in α and β. Therefore the surrogate function upto tth realization is written as:

$$\mathbf{R}^{1:t}(\mathbf{P}, \boldsymbol{\mu}, \mathbf{P}^t, \boldsymbol{\mu}^t) = \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \left(const(\mathbf{W}) - \sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}} \boldsymbol{\alpha}^{1:t}_{l'k',n',lk,n}p_{l'k',n'} + \boldsymbol{\beta}^{\mathbf{1:t}}{}_{lk,n}\sqrt{p_{lk,n}} \right) \tag{85}$$

In order to update the constants of surrogate function $\boldsymbol{\alpha}^{1:t}_{l'k',n',lk,n}$ and $\boldsymbol{\beta}^{lk,n}_{1:t}$for the tth iteration, they are updated as:

$$\boldsymbol{\alpha}^{1:t}_{l'k',n',lk,n} = \frac{1}{t}\alpha^{t}_{l'k',n',lk,n} + \frac{t-1}{t}\boldsymbol{\alpha}^{1:t-1}_{l'k',n',lk,n} \tag{86}$$
$$\boldsymbol{\beta}^{1:t}_{lk,n} = \frac{1}{t}\beta^{t}_{lk,n} + \frac{t-1}{t}\boldsymbol{\beta}^{1:t-1}_{lk,n}$$

The surrogate functions can be viewed as a pseudo concave function, in hindsight, it can be said that WMMSE algorithm provides pseudo concave functions. At each iteration t, the optimal solution $p^{t+1}_{lk,n}$ and μ^{t+1}_{lk} is obtained.

For this optimization, it follows similar pattern of maximizing of deterministic sum SE but with the history of the previous terms. With the problem formulation $\mathbf{P5}_{sto}$ when update with the constants of surrogate function becomes

$$\mathbf{P6}_{sto}:[\mathbf{P}^{t+1},\boldsymbol{\mu}^{t+1}]\Leftarrow\ \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Argmax}}\ -\frac{1}{t}\left[\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}}\left(\sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}}\boldsymbol{\alpha}^{1:t}_{l'k',n',lk,n}p_{l'k',n'}+\boldsymbol{\beta}^{\mathbf{1:t}}{}_{lk,n}\sqrt{p_{lk,n}}\right)\right] \tag{87}$$

$$s.t\quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}}p_{lk,n}\leq P_l\,\forall l,\ \bar{\mu}_{lk}\leq\hat{\mu}_{lk}\quad \text{where,}$$

$$\hat{\mu}_{lk}=\sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'}+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k}+1\right)}}. \tag{88}$$

With the above objective function, the optimal $p_{lk,n}$ is provided through first order maximization of the objective function and equating to zero which is expressed as:

$$p^{t+1}_{lk,n}=\left\{\frac{\boldsymbol{\beta}^{1:t}_{lk,n}}{\left(\sum\limits_{(l',k')}\sum\limits_{n'=1}^{\mathbb{U}_{lk}}\boldsymbol{\alpha}^{1:t}_{lk,n,l'k',n'}\right)+\lambda^*_l}\right\}^2 \tag{89}$$

where λ^*_l found through bisection algorithm which satisfies $\sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}}p^{t+1}_{lk,n}\leq P_l$.

Similarly, instead of grouping the variable **P** wise, if the equation (116) is grouped with respect to $\boldsymbol{\mu}$ wise and multiplied by $w^{t**}_{lk,n}$ the term becomes:

$$w^{t**}_{lk,n}e^{t**}_{lk,n}=const(w^{t**}_{lk,n})+\sum_{(l',k')}\sum_{n'=1}^{\mathcal{U}_{l'k'}}\gamma^{t}_{l'k'n',lk}\mu^2_{lk}+\omega^t_{lk}\mu_{lk}. \tag{90}$$

Using the similar steps of updating the surrogate function with the surrogate constants $\boldsymbol{\gamma}^{1:t}_{lk,l'k'}$ and $\boldsymbol{\omega}^{1:t}_{lk,l'k'}$ forming the objective function as in $\mathbf{P6}_{sto}$, as is constructed as:

$$\mathbf{P7}_{sto}:[\mathbf{P}^{t+1},\boldsymbol{\mu}^{t+1}]\Leftarrow\ \underset{\mathbf{P},\boldsymbol{\mu}}{\text{Argmax}}\ -\frac{1}{t}\left[\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}}\left(\sum_{(l',k')}\sum_{n'=1}^{\mathbb{U}_{lk}}\boldsymbol{\gamma}^{1:t}_{l'k'n',lk}\overline{\mu}^2_{l'k'}+\boldsymbol{\gamma}^{\mathbf{1:t}}{}_{lk}\overline{\mu}_{lk}\right)\right] \tag{91}$$

$$s.t\quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}}p_{lk,n}\leq P_l\,\forall l,\ \bar{\mu}_{lk}\leq\hat{\mu}_{lk}\quad \text{where,}$$

$$\hat{\mu}_{lk}=\sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}p_{l'k'}+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}p_{l'k}+1\right)}}. \tag{92}$$

With the above objective function, the μ_{lk} is optimized through first-order differentiation and equated to zero as:

$$\mu_{lk}^{t+1} = \min\left\{ \frac{\boldsymbol{\omega}_{lk}^{1:t}}{\sum\limits_{(l',k')} \sum_{n'=1}^{\mathcal{U}_{lk}} \boldsymbol{\gamma}_{lkn,l'k'}^{1:t}}, \hat{\mu_{lk}} \right\} \quad \text{where,}$$

$$\hat{\mu_{lk}} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} p_{l'k'}^{t+1} + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} p_{l'k}^{t+1} + 1\right)}}. \tag{93}$$

Overall, the modified stochastic WMMSE algorithm is summarized in Algorithm 2.

Algorithm 2: sum SE maximization using Stochastic WMMSE approach

Input: Given $\epsilon > 0$, the max iterations N and max power P_l for UE U_{lk} and max power Q_{lk} for the relay. Calculate the initial values $p_{lk,n}, \mu_{lk}$ with random power allocation for all relay and users i.e., $\mathbf{p}_{lk,n}^{(1)} \sim \mathbb{U}[0, P_l^{max}]$ and $\mu_{lk}^{(1)} \sim \mathbb{U}[0, Q_{lk}]$

Output: $p_{lk,n}^*$ and μ_{lk}^*.

- **for** $i \leftarrow 1$ ***to*** N **do**
  - Obtain the ith channel realization.
  - **for** $j \leftarrow 1$ ***to*** N **do**
    - Given a feasible $p_{lk,n}^i$ and μ_{lk}^i, update auxiliary variable $u_{lk,n}^{ij}$ using (74)
    - Update the auxiliary variable $w_{lk,n}^{ij}$ using (73) Do until convergence **if** $\left(\hat{R}_{\mathit{sum}}^i(p,\mu) - \left\{ \sum\limits_{(l,k)} \sum\limits_{n=1}^{\mathbb{U}_{lk}} w_{lk,n}^{ij+1} e_{lk,n}^{ij+1} - \log(w_{lk,n}^{ij+1}) \right\}\right) < \epsilon$ **then**
    - break.
    - **Output:** $w_{lk,n}^{i**}$ and w_{lk}^{i**}.
  - Compute p_{jk}^{i+1} using (57)
  - Update the $\hat{\mu}_{lk}$ constraint variable using (64)
  - Compute μ_{lk}^{i+1} using (63)
  - Update $\bar{\alpha}^i, \bar{\beta}^i, \bar{\gamma}^i, \bar{\omega}^i$ using equations
  - Do until convergence **if** $\sum\limits_{l=1}^{L} \sum\limits_{l=1}^{K} \sum\limits_{n=1}^{\mathbb{U}_{lk}} \left(p_{lk,n}^{(i)} - p_{lk,n}^{(i-1)}\right) < \epsilon$ **then**
  - break.
- **return** $\mathbf{p}^*, \mu^*$

VI. SUM SE MAXIMIZATION WITH OPTIMAL ALLOCATION OF PRECODER

In the previous sections of the multi-cell multi-relay downlink NOMA systems, we solved the optimization of sum SE with optimal power allocation deterministically and stochastically with different precoding combiners i.e MRC, Ia-ZF, Ia-MMSE. The precoding combiners with this approach depends on the instantaneous channel realizations of $\mathbf{H}$ (10). And especially precoding combiners such as Ia-ZF or Ia-MMSE requires a considerably large computation complexity and also requires a greater amount of memory. Therefore, the question arising that instead of using instantaneous percoder schemes, what if we observe deterministic precoder based on second-order statistics with samples of instantaneous channel realization. This section tries to understand the transmit signal representation, SE representation using the above transmitted signal precoders, the sum SE problem formulation and solving it all while optimally allocating precoders deterministacally with statistics from $\mathbf{H}$ and $\mathbf{g}$. *This is section is a initial study of precoder application and requires lot of further study into it. Previous works have never worked on stochastic sum SE precoder allocation.*

1) Data Transmission BS-Relay: In the data transmission mentioned in the above system model, it is considered that the precoder is unit norm and is used for decoding at relay w_{lk}. Also the transmitted signal is multiplied with the square root of power transmitter $p_{lk,n}$ of the users in the clusters (5).

In this section, we use the precoders which are for the the users $w_{lk,n}$ and we also eliminate the power variable $p_{lk,n}$ in the data transmission model by constraining the precoders with transmit powers. The precoded NOMA transmit signal with $s_{lk,n}$ being the signal of the nth user in cluster $\mathcal{U}_{lk}$ where broadcasted by the lth BS is

$$\mathbf{x}^l = \sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{jk}} \mathbf{w}_{lk,n} s_{lk,n} \tag{94}$$

Here $\mathbf{w}_{lk,n} \in \mathbb{C}^{N\times 1}$ is the precoder for the NOMA signal for the users in cluster $\mathcal{U}_{jk}$.

2) Ergodic SE analysis: Using a simple traditional comparison method, it is possible to get the SINR and the SE of the user and of the system. Looking at the equations (5) and (94), the former can be converted to later by eliminating $\mathbf{w}_{lk}$ and replacing $\sqrt{p_{lk,n}}$ by $\mathbf{w}_{lk,n}$ which is given in the following equation:

$$\mathbf{x}^l = \sum_{k=1}^{K} \cancel{\mathbf{w}_{lk}} \sum_{n=1}^{\mathcal{U}_{jk}} \frac{\mathbf{w}_{lk,n}}{\cancel{\sqrt{p_{lk,n}}}} s_{lk,n} \tag{95}$$

Following the similar trend of eliminating and replacing the precoders in the (**??**), the Genie-Bounded Ergodic Sum SE of the user is obtained as:

$$\overline{R}^p_{lk,n\text{sum}} = \sum_{(l,k)} \sum_{n=1}^{\mathbb{U}_{lk}} \mathbb{E}\left[\frac{1}{2}\left(1-\frac{\tau}{\tau_c}\right)\log_2\left(1+\frac{\overline{\Delta}^p_{lk,n}}{\overline{\Omega}^p_{lk,n}}\right)\right], \text{ where } \overline{\Omega}^p_{lk,n} = \sum_{m=0}^{7} \overline{I}^{(m)^t}_{lk,n} + 1, \tag{96}$$

$$\overline{\Delta}^{p}_{lk,n}=|\overline{A}_{lk,n}\tilde{\mu}_{lk}\mathbf{w}_{lk,n}|^2,\ \overline{I}^{(1)}_{lk,n}=\sum_{n'=1}^{n-1}|\overline{C}^{(1)}_{lk,n,n'}\tilde{\mu}_{lk}\mathbf{w}_{lk,n'}|^2,\ \overline{I}^{(2)}_{lk,n}=\sum_{n'=n+1}^{\mathcal{U}_{lk}}|\overline{C}^{(2)}_{lk,n,n'}\mathbf{w}_{lk,n'}\tilde{\mu}_{lk}|^2,$$

$$\sum_{m=3}^{6}\overline{I}^{(m)}_{lk,n}=\sum_{(l',k')\neq(l,k)}\sum_{n'=1}^{\mathcal{U}_{l'k'}}|\overline{C}^{(3)}_{l'k',lk,n}\mathbf{w}_{l'k',n'}\mu_{l'k'}|^2+\sum_{(l',k')}\sum_{(l'',k'')\neq(l',k')}\sum_{n'=1}^{\mathcal{U}_{l''k''}}|\overline{C}^{(4)}_{\substack{l''k'',l'k',\\ lk,n}}\mathbf{w}_{l''k'',n'}\tilde{\mu}_{l'k'}|^2,$$

$$\overline{I}^{(7)}_{k,n}=\sum_{l'=1}^{L}\sum_{k'=1}^{K}\overline{C}^{(5)}_{l'k',lk,n}\tilde{\mu}^2_{l'k'},\ \text{and}$$

$$\overline{\mu}_{lk}=\sqrt{\frac{q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}\sum_{n'=1}^{\mathcal{U}_{l'k'}}|\mathbf{w}_{l'k',n'}|^2+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}\sum_{n'=1}^{\mathcal{U}_{l'k}}|\mathbf{w}_{l'k,n'}|^2+1\right)}}. \tag{97}$$

where the terms $\overline{A}_{lk,n}$, $\overline{C}^{(1)}_{lk,n}$, $\overline{C}^{(2)}_{lk,n}$, $\overline{C}^{(3)}_{l'k',lk,n}$, $\overline{C}^{(4)}_{\substack{l''k'',l'k',\\ lk,n}}$ and $\overline{C}^{(5)}_{l'k',lk,n}$ are functions of instantaneous channel realizations which are given as

$$\overline{A}_{lk,n}=f_{lk,n}g^{lk}_{lk,n}\mathbf{h}^{lH}_{lk},\ \overline{C}^{(1)}_{lk,n}=f_{lk,n}g^{lk}_{lk,n}\mathbf{h}^{lH}_{lk},\ \overline{C}^{(2)}_{lk,n}=f_{lk,n}g^{lk}_{lk,n}\mathbf{h}^{lH}_{lk}-f_{lk,n}\hat{g}^{lk}_{lk,n}\hat{\mathbf{h}}^{lH}_{lk},$$

$$\overline{C}^{(3)}_{l'k',lk,n}=f_{lk,n}g^{l'k'}_{lk,n}\mathbf{h}^{l'H}_{l'k'},\ \overline{C}^{(4)}_{\substack{l''k'',l'k',\\ lk,n}}=f_{lk,n}g^{l'k'}_{lk,n}\mathbf{h}^{l''H}_{l'k'},\ \overline{C}^{(5)}_{l'k',lk,n}=|f_{lk,n}g^{l'k'}_{lk,n}|^2 \tag{98}$$

3) Sum SE formulation and Optimization: In this section, we solve the stochastic SE optimization by optimizing the precoder allocation, following the similar trend of the previous section. We use SSUM and WMMSE algorithm for optimization where at every realization the instantaneous sum SE is learnt through recursive surrogate function (SSUM) and the optimizing variables along with auxiliary variables are updated regularly. Before formulating the objective function, we also know that, inorder to attain the closed form solution, we replace the transmit relay power q_{lk} to amplitude factor μ_{lk} as the optimization variable to attain iterative closed form solutions. We also know that the stochastic optimization is based on the samples of the channel realization and the practical mean. Using all the known factors, the objective function for the optimization of the Ergodic sum SE is constructed as:

$$\mathbf{P1}_{pre}:\quad \underset{\overline{\mathbf{W}},\boldsymbol{\mu}}{\text{Maximize}}\ \lim_{T\Rightarrow\infty}\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{(l,k)}\sum_{n=1}^{\mathcal{U}_{lk}}\log_2\left(1+\frac{\overline{\Delta}^{p}_{lk,n}(\mathcal{F}^t)}{\overline{\Omega}^{p}_{lk,n}(\mathcal{F}^t)}\right)\right] \tag{99}$$

$$s.t\quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}}|\mathbf{w}_{lk,n}|^2\le P_l\,\forall l,\ \bar{\mu}_{lk}\le\hat{\mu}_{lk}\quad \text{where,}$$

$$\hat{\mu}_{lk}=\sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')}\kappa_{l'k',lk}\sum_{n'=1}^{\mathcal{U}_{l'k'}}|\mathbf{w}_{l'k',n'}|^2+\sum\limits_{l'=1}^{L}\rho_{l'k,lk}\sum_{n'=1}^{\mathcal{U}_{l'k}}|\mathbf{w}_{l'k,n'}|^2+1\right)}}. \tag{100}$$

where $\mathcal{F}=\{\mathbf{h}^{l'}_{lk},g^{l'k'}_{lk,n}\},\forall l,k,l',k',n$. Using the methods followed to construct double optimization using auxiliary variables in $\mathbf{P4}_{sto}$ from $\mathbf{P3}_{sto}$, the above problem formulation $\mathbf{P1}_{pre}$ can reconstructed with auxiliary variables as:

$$\mathbf{P2}_{pre}: \quad \underset{\overline{\mathbf{W}},\boldsymbol{\mu}}{\text{Maximize}} \lim_{T\Rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}\left[\underset{\mathbf{U}^t,\mathbf{W}^t}{\text{Maximize}} \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} w^t_{lk,n}e^t_{lk,n} - \log(w^t_{lk,n})\right] \tag{101}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} |\mathbf{w}_{lk,n}|^2 \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk}\sum_{n'=1}^{\mathcal{U}_{l'k'}} |\mathbf{w}_{l'k',n'}|^2 + \sum\limits_{l'=1}^{L}\rho_{l'k,lk}\sum_{n'=1}^{\mathcal{U}_{l'k}} |\mathbf{w}_{l'k,n'}|^2 + 1\right)}}. \tag{102}$$

where the auxiliary variables $w^t_{lk,n}$ and $u^t_{lk,n}$ are grouped into matrix $\mathbf{W}$,$\mathbf{U}^t$ respectively. The auxiliary variable $e_{lk,n}$ is defined as:

$$e^t_{lk,n} = |1 - u^t_{lk,n}\sqrt{\overline{\Delta}^p_{lk,n}(\mathcal{F}^t)}|^2 + |u^t_{lk,n}|^2\,\overline{\Omega}^p_{lk,n}(\mathcal{F}^t), \forall\, l,k,n.$$

And the optimal values of $w^t_{lk,n}$ and $u^t_{lk,n}$ are given as:

$$u^{t*}_{lk,n} = \frac{\sqrt{\overline{\Delta}^i_{lk,n}(\mathcal{F}^t)}}{\overline{\Omega}^i_{lk,n}(\mathcal{F}^t)} \tag{103}$$

$$w^{t*}_{lk,n} = \frac{1}{e^{t*}_{lk,n}} \text{ where,} \tag{104}$$

$$e^{t*}_{lk,n} = |1 - u^{t*}_{lk,n}\sqrt{\overline{\Delta}^p_{lk,n}(\mathcal{F}^t)}|^2 + |u^{t*}_{lk,n}|^2\,\overline{\Omega}^p_{lk,n}(\mathcal{F}^t)$$

Once the inner optimization (i.e optimizing the matrices $\mathbf{W}^t,\mathbf{U}^t$) is completed. With the optimal auxiliary variables $(\mathbf{W}^{t**},\mathbf{U}^{t**})$ the problem formulation $\mathbf{P2}_{pre}$ is deconstructed as:

$$\mathbf{P3}_{pre}: \quad \underset{\overline{\mathbf{W}},\boldsymbol{\mu}}{\text{Maximize}} \lim_{T\Rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} w^{t**}_{lk,n}e^{t**}_{lk,n} - \log(w^{t**}_{lk,n})\right] \tag{105}$$

$$s.t \quad \sum_{k=1}^{K}\sum_{n=1}^{\mathcal{U}_{lk}} |\mathbf{w}_{lk,n}|^2 \leq P_l\,\forall l,\ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk}\sum_{n'=1}^{\mathcal{U}_{l'k'}} |\mathbf{w}_{l'k',n'}|^2 + \sum\limits_{l'=1}^{L}\rho_{l'k,lk}\sum_{n'=1}^{\mathcal{U}_{l'k}} |\mathbf{w}_{l'k,n'}|^2 + 1\right)}}. \tag{106}$$

where the variables $e^{t**}_{lk,n}$, $w^{t**}_{lk,n}$ and $u^{t**}_{lk,n}$ are the optimal variables after several iterations. As the above objective function (105) is pseudo concave function from the proposition 2. Therefore the problem $\mathbf{P3}_{pre}$ becomes stochastic convex optimization problem. Inorder to apply the SSUM algorithm, we require a convex approximation of the objective function of the tth realization which helps in forming the surrogate function. From the proposition 2, that:

$$\tilde{R}_{\text{sum}}(\overline{\mathbf{W}},\boldsymbol{\mu},\mathcal{F}^t) = \sum_{(l,k)}\sum_{n=1}^{\mathbb{U}_{lk}} \log(w^{t**}_{lk,n}) - w^{t**}_{lk,n}e^{t**}_{lk,n} \tag{107}$$

is a valid convex approximate of the objective function of $\mathbf{P2}_{pre}$ only at the t realization. Using the recursive surrogate function [], the surrogate function $\tilde{\mathbb{R}}^{1:t}_{\text{sum}}$ is expressed as:

$$\mathbf{R}^{1:t}(\bar{\mathbf{W}}, \boldsymbol{\mu}) = \frac{1}{t}\left[\mathbf{R}^{1:t-1}(\bar{\mathbf{W}}, \boldsymbol{\mu})\right] + \frac{t-1}{t}\left[\tilde{R}_{\text{sum}}(\overline{\mathbf{W}}, \boldsymbol{\mu}, \mathcal{F}^t)\right] \tag{108}$$

To make the above equation simpler, we try to group the auxiliary variable $e^{t**}_{lk,n}$ present in the equation (110), based on percoder $\mathbf{w}_{lk,n}$ which is presented as:

$$w^{t**}_{lk,n} e^{t**}_{lk,n} = cons(w^{t**}_{lk,n}) + \sum_{(l',k')} \sum_{n'=1}^{\mathbb{U}_{lk}} \mathbf{w}^H_{l'k',n'} \boldsymbol{\alpha}^t_{l'k',n',lk,n} \mathbf{w}_{l'k',n'} + \boldsymbol{\beta}^t_{lk,n} \mathbf{w}_{lk,n}. \tag{109}$$

All the constants including the amplitude factor $\boldsymbol{\mu}$ are combined and termed as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. With the above grouping the convex approximate function $\tilde{R}_{\text{sum}}(\overline{\mathbf{W}}, \boldsymbol{\mu}, \mathcal{F}^t)$ is written as:

$$\tilde{R}_{\text{sum}}(\overline{\mathbf{W}}, \boldsymbol{\mu}, \mathcal{F}^t) = \sum_{(l,k)} \sum_{n=1}^{\mathbb{U}_{lk}} \left(const(w) - \sum_{(l',k')} \sum_{n'=1}^{\mathbb{U}_{lk}} \mathbf{w}^H_{l'k',n'} \boldsymbol{\alpha}^t_{l'k',n',lk,n} \mathbf{w}_{l'k',n'} + \boldsymbol{\beta}^t_{lk,n} \mathbf{w}_{lk,n} \right) \tag{110}$$

As one can understand that for any t realization, the convex approximate tend to be the same. Therefore using the mathematical induction and keeping the amplitude factor $\boldsymbol{\mu}$ as fixed and using the grouped auxiliary variable, one can simplify the surrogate function to be:

$$\mathbf{R}^{1:t}(\bar{\mathbf{W}}, \boldsymbol{\mu}) = \sum_{(l,k)} \sum_{n=1}^{\mathbb{U}_{lk}} \left(const(w) - \sum_{(l',k')} \sum_{n'=1}^{\mathbb{U}_{lk}} \mathbf{w}^H_{l'k',n'} \breve{\boldsymbol{\alpha}}^{1:t}_{l'k',n',lk,n} \mathbf{w}_{l'k',n'} + \breve{\boldsymbol{\beta}}^{lk,n}_{1:t} \mathbf{w}_{lk,n} \right) \text{ where,} \tag{111}$$

$$\breve{\boldsymbol{\alpha}}^{1:t}_{l'k',n',lk,n} = \frac{1}{t} \boldsymbol{\alpha}^t_{l'k',n',lk,n} + \frac{t-1}{t} \breve{\boldsymbol{\alpha}}^{1:t-1}_{l'k',n',lk,n} \tag{112}$$

$$\breve{\boldsymbol{\beta}}^{1:t}_{lk,n} = \frac{1}{t} \boldsymbol{\beta}^t_{lk,n} + \frac{t-1}{t} \breve{\boldsymbol{\beta}}^{1:t-1}_{lk,n} \tag{113}$$

Using the construction of the objective function mentioned in $\mathbf{P6}_{sto}$, the problem formulation for the stochastic optimization with the surrogate function is constructed as:

$$\mathbf{P4}_{pre} : [\overline{\mathbf{W}}^{t+1}, \boldsymbol{\mu}^{t+1}] \Leftarrow \underset{\overline{\mathbf{W}}, \boldsymbol{\mu}}{\text{Argmax}} \ \frac{1}{t}\left[\mathbf{R}^{1:t-1}(\bar{\mathbf{W}}, \boldsymbol{\mu})\right] + \frac{t-1}{t}\left[\tilde{R}_{\text{sum}}(\overline{\mathbf{W}}, \boldsymbol{\mu}, \mathcal{F}^t)\right] \triangleq \mathbf{R}^{1:t}(\bar{\mathbf{W}}, \boldsymbol{\mu}) \tag{114}$$

$$s.t \quad \sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} |\mathbf{w}_{lk,n}|^2 \leq P_l \, \forall l, \ \bar{\mu}_{lk} \leq \hat{\mu}_{lk} \quad \text{where,}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} \sum_{n'=1}^{\mathcal{U}_{l'k'}} |\mathbf{w}_{l'k',n'}|^2 + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} \sum_{n'=1}^{\mathcal{U}_{l'k}} |\mathbf{w}_{l'k,n'}|^2 + 1 \right)}}.$$

the optimized $\mathbf{w}_{lk,n}$ is provided through the first order differentiation of the objective function of $\mathbf{P4}_{pre}$, which surprisingly follows the similar pattern to that of optimizing the Ergodic sum

SE and is given below:

$$\mathbf{w}_{lk,n}^{t+1} = \left\{ \left(\sum_{(l',k')} \sum_{n'=1}^{\mathbb{U}_{lk}} \breve{\boldsymbol{\alpha}}_{lk,n,l'k',n'}^{1:t} \right) + \lambda_l^* I_N \right\}^{-1} \left\{ \left(\boldsymbol{\beta}_{lk,n}^{1:t} \right) \right\} \tag{115}$$

where λ_l^* found through bisection algorithm which satisfies $\sum_{k=1}^{K} \sum_{n=1}^{\mathcal{U}_{lk}} |\mathbf{w}_{lk,n}^{t+1}|^2 \leq P_l$.

Similarly, instead of grouping the variable $\overline{\mathbf{W}}$ wise, if the equation (109) is grouped with respect to $\boldsymbol{\mu}$ wise, the term becomes:

$$w_{lk,n}^{t**} e_{lk,n}^{t**} = const(w) + \sum_{(l',k')} \gamma_{l'k',lk}^{t} \mu_{lk}^2 + \omega_{lk}^{t} \mu_{lk}. \tag{116}$$

Keeping the updated precoder fixed and following the similar steps from problem formulation $\mathbf{P3}_{pre}$ to $\mathbf{P4}_{pre}$ and by forming the simpler surrogate function (111) and updation of the surrogate variables (112), the $\mu_{lk,n}$ is optimized through first-order differentiation of new surrogate variable keeping precoder constant and equating to zeros is expressed as:

$$\overline{\mu}_{lk}^{t+1} = \min \left\{ \frac{\omega_{lk}^{1:t}}{\sum\limits_{(l',k')} \gamma_{lk,l'k'}^{1:t}}, \hat{\mu_{lk}} \right\} \text{ where,} \tag{117}$$

$$\hat{\mu}_{lk} = \sqrt{\frac{Q_{lk}}{\left(\sum\limits_{(l',k')} \kappa_{l'k',lk} \sum_{n'=1}^{\mathcal{U}_{l'k'}} |\mathbf{w}_{l'k',n'}|^2 + \sum\limits_{l'=1}^{L} \rho_{l'k,lk} \sum_{n'=1}^{\mathcal{U}_{l'k}} |\mathbf{w}_{l'k,n'}|^2 + 1 \right)}}. \tag{118}$$

The overall Algorithm, the modified stochastic successive convex approximation is summarized in Algorithm 3.

Algorithm 3: sum SE maximization with Precoder allocation

Input: Given $\epsilon > 0$, the max iterations N and max power P_l for UE U_{lk} and max power Q_{lk} for the relay. Calculate the initial values $\mathbf{w}_{lk,n}, \mu_{lk}$ with random precoder allocation for all relay and users i.e., $|\mathbf{w}_{lk,n}^{(1)}|^2 \sim \mathbb{U}[0, P_l^{max}]$ and $\mu_{lk}^{(1)} \sim \mathbb{U}[0, Q_{lk}]$

Output: $\mathbf{w}_{lk,n}^*$ and μ_{lk}^*.

for $i \leftarrow 1$ ***to*** N **do**
- Obtain the ith channel realization.
- **for** $j \leftarrow 1$ ***to*** N **do**
  - Given a feasible $\mathbf{w}_{lk,n}^{i}$ and μ_{lk}^{i}, update auxiliary variable $u_{lk,n}^{ij}$ using (**??**)
  - Update the auxiliary variable $w_{lk,n}^{ij}$ using (104) Do until convergence **if** $\left\{ \sum\limits_{(l,k)} \sum\limits_{n=1}^{\mathbb{U}_{lk}} \left(w_{lk,n}^{ij} e_{lk,n}^{ij} - \log(w_{lk,n}^{ij})) \right) - \left(w_{lk,n}^{ij+1} e_{lk,n}^{ij+1} - \log(w_{lk,n}^{ij+1}) \right) \right\} < \epsilon$ **then**
    - break.
  - **Output:** $w_{lk,n}^{i**}$ and u_{lk}^{i**}.
- Compute p_{jk}^{i+1} using (115)
- Update the $\hat{\mu}_{lk}$ constraint variable using (118)
- Compute μ_{lk}^{i+1} using (117)
- Update $\bar{\alpha}^i, \bar{\beta}^i, \bar{\gamma}^i, \bar{\omega}^i$ using equations (112)
- Do until convergence **if** $\sum\limits_{l=1}^{L} \sum\limits_{l=1}^{K} \sum\limits_{n=1}^{\mathbb{U}_{lk}} \left(|\mathbf{w}_{lk,n}^{(i)}|^2 - |\mathbf{w}_{lk,n}^{(i-1)}|^2 \right) < \epsilon$ **then**
  - break.

return $\mathbf{w}^*$,μ^*

VII. SIMULATION RESULTS

In this part, we assess the multi-cell multi-relay mMIMO NOMA Downlink system's optimizations' evaluation, where the BS and users estimate the local CSI and the users deploy imperfect SIC. Consider a 20-MHz system with 4 cell, 60 BS antennas, 3 relays, 12 users with each relay of 4 users and remains throughout the simulation unless its mentioned. Also each coherence period has 200 symbols and a K-symbol pilot transmission interval with $p_p = 20dBm$. The system model model is simulated such that the BS is at the center and the relays are uniformly distributed around 200m circle and the users are randomly distributed around 100m circle. We model the large scale fading coefficient from the l'th BS to k th relay in lth cell and k' th relay in l'th cell to nth user in cluster $\mathcal{U}_{lk}$ as $\beta_{lk}^{l\prime}[\mathrm{dB}] = \Upsilon - 10\alpha\log_{10}\left(d_{lk}^{l''}\right) + F_{lk}^{l''}$ and $\beta_{lk,n}^{l'k'}[\mathrm{dB}] = \Upsilon - 10\alpha\log_{10}\left(d_{lk,n}^{l'k'}\right) + F_{lk,n}^{l'k'}$. Here Υ denotes the median channel gain at a reference distance of 30.18, α is the path loss exponent of 2.6, $d_{\underset{lk}{l\prime}}$ and $d_{lk,n}^{l'k'}$ is the separation distance between the BS-and the relay and relay and user in metres, and the scalars $F_{lk}^{l''}$ and $F_{lk,n}^{l'k'}$ are the shadow fading terms which models the log-normal random variations. The Rician factors of the random channels are modelled as $K_{lk}^{l'}$ and $K_{lk,n}^{l'k'}$ can be calculated as $K_{lk}^{l'}[\mathrm{dB}] = 13 - 0.03d_{lk}^{l'}$ and $K_{lk,n}^{l'k'}[\mathrm{dB}] = 13 - 0.03d_{lk,n}^{l'k'}$ respectively.The other parameters is the assumption of ULA at the base station and the ASD of 10^o.

A. Validation of optimization techniques

As the work of the paper is of optimization techniques, it is necessary to understand that the optimal value has been achieved. The ϵ is set to be as 0.001 and the framework converges smoothly. In the Fig.1, we plot the sum SE (UATF and Ergodic) of users vs the maximum transmit power allocation to the BS P_T and to the relay Q_T. The plots are compared between Optimized Power Allocation (OPA) and Full Power Allocation (FPA) where the users are allocated power equally to maximum. $P_T/(K * \mathcal{U}_{lk})$. The Fig.1a, provides the OPA plot of the converged value of deterministic WMMSE optimization of the UATF sum SE. Similarly, the Figs.1b-d provides the FPA plot of the converged value of stochastic WMMSE optimization of the Ergodic sum SE with the precoders of MRC, ZF,MMSE respectively. At higher powers, you can clearly see that MRC precoders can maximize the sum SE much better that the MMSE precoder, which should be true because the MMSE precoder is built based on maximizing the SINR at FPA. Thus strengthening the work. The Fig.1e, provides the OPA plot of the converged value of stochastic WMMSE optimization of the Ergodic sum SE with optimal precoder allocation. The FPA of this plot is the random precoder which satisfies the power constraint and is optimized. Inorder to plot the OPA, the paper has taken different channel realizations than the one used for optimizing, thus proving that the optimized precoder is solely based on second-order statistics of the channel. The paper has taken 10 User Setups and have averaged it for the uniformity between optimization techniques.

B. Algorithm Time Complexity vs Convergence

The base paper for this work [4], provides an exciting optimizing work which uses the modified AMM approach with QT and LDT. In the Fig.2 we plot the CPU time while optimizing the UATF sum SE at each iterations with maximum power allocation at the BS and at the relays to be $20dBm$. The setup is compared in a Intel® Core™ i5-8250U CPU @ 1.60GHz $\times$8 with a 8GB Ram. Inorder to provide a perspective, the no. of cell taken to be 1 to reduce the convergence time of the existing work. In this plot, we compare the time taken to converge, with an best, low-complexity, existing framework (AMM) [4] and the current framework (Deterministic WMMSE approach) by optimizing the UATF sum SE and the iterations for former is 1050 and the later is 60. The time scale is a log scale inorder to provide better information from the plot. As you can see, the current optimization technique is way faster than the existing best technique by a large difference as the existing framework uses cvx for optimization. Also the current framework does take large iterations to converge at a inconsiderable time. With the faster convergence there is always a trade-off, here its the sub-optimality of the current framework as it does harsher convex approximation to bring iterative closed-form solution. As the famous saying goes *"You live by the sword, you'll die by the sword."*

C. Inference from the Optimization Techniques

Inorder to comparing the optimization techniques, we need to make sure the initial values are equal for all the optimizations. In the Fig.3, we plot the Ergodic sum SE vs the maximum power allocation of the BS and the relay. We compare how the optimal precoder allocation has fared with the MR and ZF precoders. From the Fig.3a, which is optimized power/precoder allocation, one can infer that at low SNR, all the three are closely converged and at high SNR you can clearly see that the OPA of ZF $\geq$ Precoder Allocation $\geq$ MRC. But Fig.3b, at high SNR, shows that the FPA of ZF $\geq$ MRC $\geq$ Precoder Allocation. The above observation provides the details that the optimal precoder allocation has better optimized which is due the random

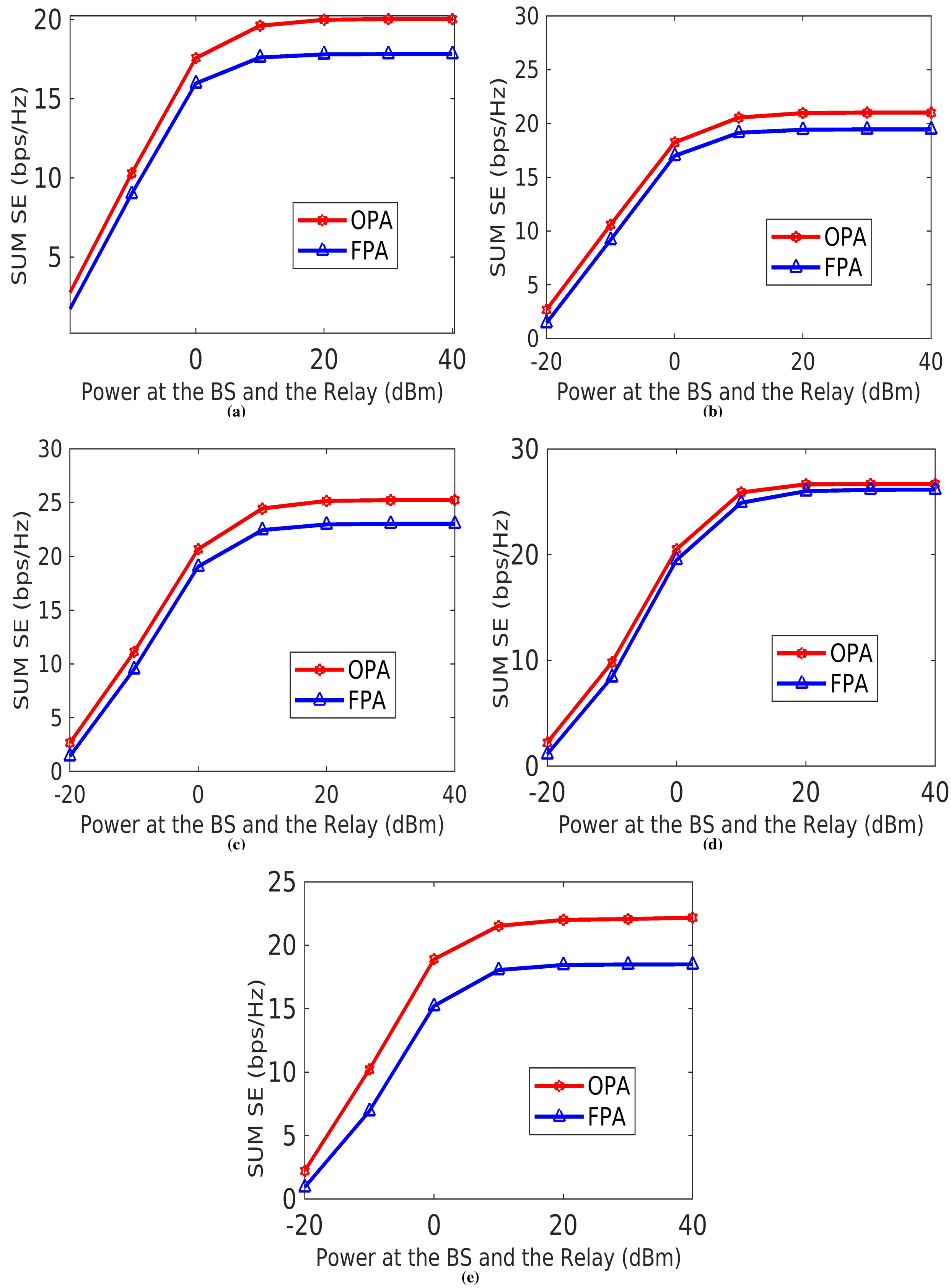

Fig. 1: a) UATF SUM SE versus BS transmit power P_T Ergodic Sum SE with precoder b) MRC; c) ZF and; d) MMSE e)Ergodic Sum SE with precoder allocation

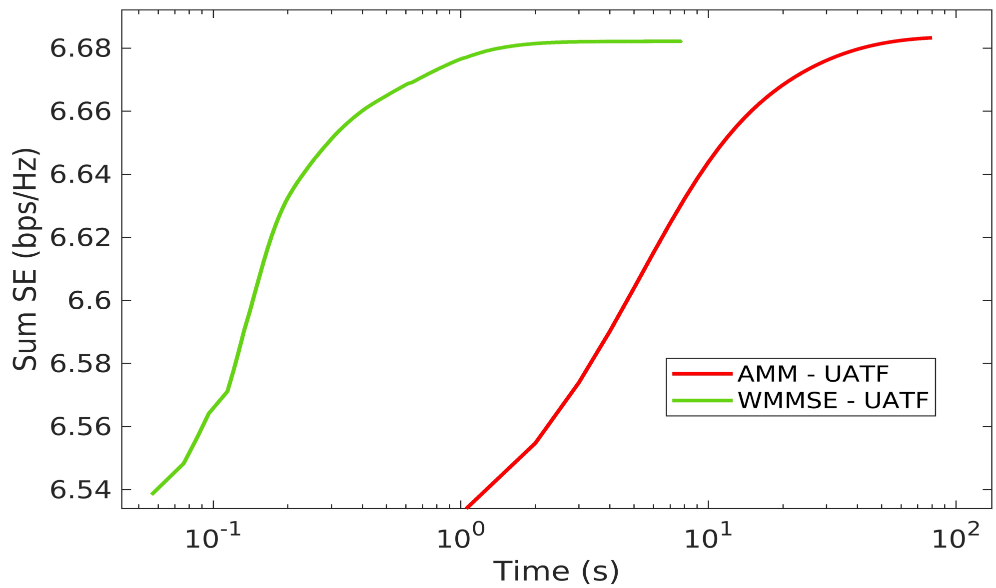

Fig. 2: UATF Sum SE vs Time.

initialization of the precoder with the power constraints. It also gives a interesting fact that, being a deterministic precoder, it fares better than MRC but not to ZF where these have precoders that are instantaneous. It is an exciting future prospect to move towards optimal precoder allocation for two reasons: 1) Precoders will be deterministic, reduction in the time complexity in the formation of the precoders. 2) Reduction in cost of both optimizing power and formation of precoder. These aforementioned reasons will also improves the throughput of the system which is need of the hour.

VIII. Conclusion

The paper has worked on a existing system model, to obtain three optimization frameworks with the WMMSE approach at a very low cost, which was observed with the time complexity plot. The paper has provided Deterministic WMMSE approach, Stochastic WMMSE approach. The paper also extended its work to precoder allocation, a new way of inexpensive stochastic optimizations with a very insightful future prospect to it.

References

[1] Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, and X. You, "Signal processing for MIMO-NOMA: present and future challenges," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 32–38, 2018.

[2] A. S. de Sena, F. R. M. Lima, D. B. da Costa, Z. Ding, P. H. J. Nardelli, U. S. Dias, and C. B. Papadias, "Massive MIMO-NOMA networks with imperfect SIC: design and fairness enhancement," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6100–6115, 2020.

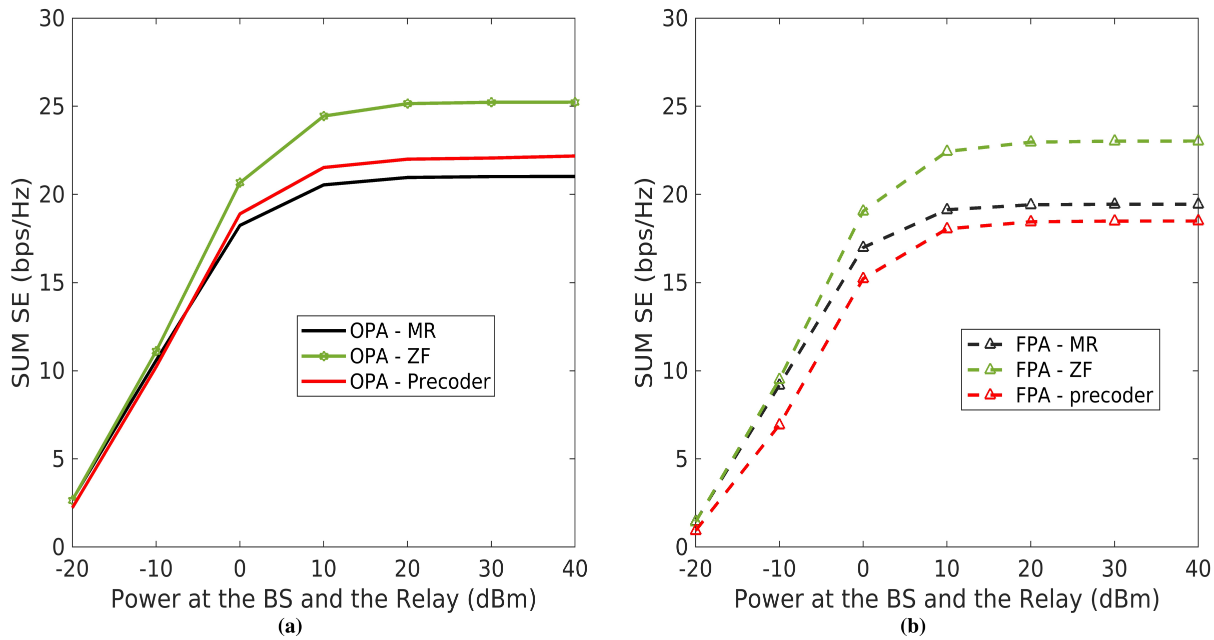

Fig. 3: Ergodic Sum SE versus BS transmit power P_T with a) OPA b) FPA.

[3] L. Liu, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving MIMO-NOMA: Iterative LMMSE detection," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1758–1773, 2019.

[4] D. N. Amudala, B. Kumar, and R. Budhiraja, "Spatially-correlated rician-faded multi-relay multi-cell massive mimo noma systems," *IEEE Transactions on Communications*, pp. 1–1, 2022.

[5] V. Mandawaria, E. Sharma, and R. Budhiraja, "WSEE maximization of mmwave NOMA systems," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1413 – 1417, 2019.

[6] D. Kudathanthirige and G. A. A. Baduge, "NOMA-aided multicell downlink massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 612–627, 2019.

[7] Y. Li and G. A. A. Baduge, "NOMA-aided cell-free massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 950–953, 2018.

[8] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, L. Hanzo, and P. Xiao, "On the performance of cell-free massive MIMO relying on adaptive NOMA/OMA mode-switching," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 792–810, 2020.

[9] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 5, no. 4, pp. 900–919, 2019.

[10] D. Zhang, Y. Liu, Z. Ding, Z. Zhou, A. Nallanathan, and T. Sato, "Performance analysis of non-regenerative massive-MIMO-NOMA relay systems for 5G," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4777–4790, 2017.

[11] X. Chen, R. Jia, and D. W. K. Ng, "The application of relay to massive non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5168–5180, 2018.

[12] Mandawaria, E. Sharma, and R. Budhiraja, "Energy-efficient massive MIMO multi-relay NOMA systems with CSI errors," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7410–7428, 2020.

[13] Y. Li and G. Amarasuriya, "Multiple relay-aided massive MIMO NOMA," in *2019 IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[14] Y. Li and G. A. A. Baduge, "Relay-aided downlink massive MIMO NOMA with estimated CSI," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2258–2271, 2021.

[15] D. Zhang, Y. Liu, Z. Ding, Z. Zhou, A. Nallanathan, and T. Sato, "Performance analysis of non-regenerative massive-MIMO-NOMA relay systems for 5G," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4777–4790, 2017.

[16] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[17] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, 2020.

[18] H. Yang, K. Zhang, K. Zheng, and Y. Qian, "Joint frame design and resource allocation for ultra-reliable and low-latency vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3607–3622, 2020.

[19] M. Monemi and H. Tabassum, "Performance of UAV-assisted D2D networks in the finite block-length regime," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7270–7285, 2020.

[20] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing resource allocation in the short blocklength regime for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 402–415, 2019.

[21] K. Singh, M.-L. Ku, and M. F. Flanagan, "Energy-efficient precoder design for downlink multi-user MISO networks with finite blocklength codes," *IEEE Trans. on Green Commun. and Netw.*, vol. 5, no. 1, pp. 160–173, 2021.

[22] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4189–4203, 2019.

[23] Q.-D. Vu, L.-N. Tran, and M. Juntti, "On spectral efficiency for multiuser MISO systems under imperfect channel information," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1946–1951, 2021.

[24] H. Zhang, F. Fang, J. Cheng, K. Long *et al.*, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 48–53, April 2018.

[25] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth Dover printing ed. New York: Dover, 1964.

[26] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," 2013. [Online]. Available: https://arxiv.org/abs/1307.4457

[27] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 16, pp. 4189–4203, 2019.