# Spectral Efficiency for Massive MIMO Multi-Relay NOMA Systems with CSI errors

Vikalp Mandawaria, Ekant Sharma and Rohit Budhiraja
*Department of Electrical Engineering IIT Kanpur* Kanpur, India
vikalp,ekant,rohitbr@iitk.ac.in

***Abstract*—We consider a multiple-relay-aided massive multi-input multi-output (MIMO) non-orthogonal multiple access (NOMA) system. We practically model this system by considering channel estimation error, and the consequent imperfect successive interference cancellation, which the existing literature has ignored. We derive a novel lower bound for the ergodic spectral efficiency of this multi-relay NOMA system, and characterize its degradation with respect to orthogonal multiple access (OMA), due to the aforementioned artifacts. We crucially show that a multi-relay massive MIMO NOMA system requires *accurate channel information* to outperform OMA.**

***Index Terms*—Massive antennas, relays, spectral efficiency.**

## I. Introduction

Non-orthogonal multiple access (NOMA) is being envisaged to meet the high spectral efficiency (SE) requirements of 5G cellular systems [1]. NOMA is also recently applied to massive multiple-input multiple-output (MIMO) systems [2], [3]. Reference [2] proposed a novel NOMA scheme with shared pilots. Reference [3] derived the delay-limited throughput for NOMA system with both imperfect and perfect successive interference cancellation (SIC).

Massive MIMO and NOMA techniques are being applied to relays to enhance their spectral efficiency (SE), reduce power consumption and improve coverage [4]–[6]. Reference [4] investigated, for a single-relay aided massive MIMO NOMA system, the impact of imperfect BS channel state information (CSI), and consequently the imperfect users SIC, and intra-cluster pilot contamination on the SE. Reference [5], [6] considered, as shown in Fig. 1, a multiple single-antenna-relay-assisted NOMA system with a massive MIMO BS. These relays cooperatively exploit diversity to enhance the system SE. Reference [5] derived the SE, by assuming perfect CSI and SIC, *which is valid only for infinitely large BS antennas.* Reference [6] assumed perfect CSI at the BS and separately optimized BS and relay powers to maximize SE.

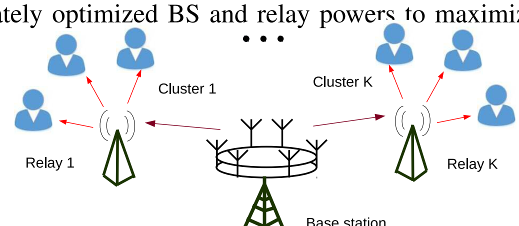



**Fig. 1:** Multiple-relay aided massive MIMO-NOMA.

Motivated by [5], [6], we consider, as shown in Fig. 1, a multi-relay-aided massive MIMO NOMA system. The relays are allocated users closest to them which are then served based on NOMA concept on the same time-frequency resource. In the considered system, while the interference between signal intended for closely-placed users is mitigated with NOMA SIC, the relay deployment itself reduces interference for users located far-away in the cell. Besides this, with the deployment of relays, the users require instantaneous CSI only of the single-antenna relay-to-user links for SIC, instead of massive direct BS-to-user links, reducing the CSI estimation overhead.

The **main contributions** of the current work are as follows

- We derive a closed-form SE lower-bound for a multiple-relay aided massive MIMO NOMA system in Fig. 1 which, *unlike [5] is valid for arbitrary number of antennas.* We consider, different from [6], a practical massive MIMO system with imperfect CSI at the BS, statistical CSI at users, which leads to imperfect SIC at the users.
- We present a novel channel estimation strategy, where BS-to-relay and relay-to-users CSI can be obtained at the BS and at the users by transmitting only $K$ pilots from the relay. This reduces pilot overhead, which increases the SE.
- Our results reveal that NOMA outperforms orthogonal multiple access (OMA) only when accurate CSI is available, otherwise residual interference, due to imperfect SIC, degrades NOMA SE. This study, therefore, captures the effect of channel estimation error on the NOMA performance, which reference [6] did not. We also quantify the SE gain accrued by multi-relay NOMA system over OMA schemes.

## II. System Model

We consider, as shown in Fig. 1, a multiple-relay-aided massive MIMO NOMA system with an $N_T$-antenna BS and $K$ single-antenna half-duplex amplify-and-forward relays. The BS communicates via the $k$th relay with a cluster of $U_k$ single-antenna users on the same spectral resource by employing NOMA.[1] The users are associated with a particular relay based on their spatial locations i.e., the users close to a relay forms a cluster [6]. The signal transmission from the BS to the users requires two time slots. In the first time slot, the BS transmits a precoded signal $\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k x_k$ to all the relays, where $\mathbf{w}_k$ is the transmit precoder designed for the $k$th relay, and $x_k = \sum_{n=1}^{U_k} \sqrt{p_{k,n}} s_{k,n}$ is the NOMA signal for the $k$th relay. Here $p_{k,n}$ and $s_{k,n}$ are the transmit power and the message signal for the $n$th user associated with the $k$th relay, respectively. The signal received by the $k$th relay can be expressed as $y_{R,k} = \sum_{k'=1}^{K} \mathbf{h}_k^T \mathbf{w}_{k'} x_{k'} + z_{R,k}$, where $\mathbf{h}_k^T \in \mathbb{C}^{1\times N_T}$ denotes the downlink channel from BS to the $k$th relay, and the scalar $z_{R,k}$ represents the additive white Gaussian noise (AWGN) with unit variance at the $k$th relay.

We note that reference [6] assumed that the BS has perfect downlink CSI. *We consider a practical model where*

[1] To avoid repetition, we assume that $k = 1$ to $K$ throughout this paper.

*BS estimates uplink CSI using the pilots transmitted by the relay*. The CSI estimation is discussed at the end of this section. The BS exploits the time division duplex (TDD) reciprocity to design a zero-forcing (ZF) precoder $\mathbf{w}_k$ to cancel the first-hop multi-relay interference as follows $\mathbf{w}_k = \frac{\mathbf{\Pi}_k \hat{\mathbf{h}}_k^*}{\|\mathbf{\Pi}_k \hat{\mathbf{h}}_k^*\|}$. Here $\mathbf{\Pi}_k = \mathbf{I}_{N_T} - \hat{\mathbf{H}}_k \left(\hat{\mathbf{H}}_k^H \hat{\mathbf{H}}_k\right)^{-1} \hat{\mathbf{H}}_k^H$ with $\hat{\mathbf{H}}_k = \left[\hat{\mathbf{h}}_1, \cdots, \hat{\mathbf{h}}_{k-1}, \hat{\mathbf{h}}_{k+1}, \cdots, \hat{\mathbf{h}}_K\right]^*$, and $\hat{\mathbf{h}}_k$ is estimate of $\mathbf{h}_k$.

In the second time slot, each single-antenna relays amplifies the received signal with fixed-gain $\mu_k$ as $r_k = \mu_k y_{R,k}$ and broadcasts it to all the users. The scalar

$$\mu_k = \sqrt{\frac{q_k}{\mathbb{E}\left[\left|\sum_{k'=1}^{K} \mathbf{h}_k^T \mathbf{w}_{k'} x_{k'} + z_{R,k}\right|^2\right]}}, \tag{1}$$

constrains the average transmit power of the $k$th relay, denoted as $q_k$. The signals transmitted by $K$ relays interfere with each other, and the $n$th user in the $k$th cluster receives a sum-signal $\hat{y}_{k,n} = \sum_{k'=1}^{K} g_{k',k,n} r_{k'} + z_{k,n}$. The scalar $g_{k',k,n}$ is the channel between the $k'$th relay and the $n$th user associated with the $k$th relay, the scalar $z_{k,n}$ is the unit-variance AWGN at the $n$th user associated with the $k$th relay. The $n$th user associated with the $k$th relay uses $f_{k,n}$ to equalize receiver phase as $y_{k,n} = f_{k,n} \hat{y}_{k,n} = f_{k,n}\left(\sum_{k'=1}^{K} g_{k',k,n} r_{k'} + z_{k,n}\right)$. It is designed as $f_{k,n} = \hat{g}_{k,k,n}^* / |\hat{g}_{k,k,n}|$, where $\hat{g}_{k,k,n}$ is the estimate of $g_{k,k,n}$. This is unlike [6], where user perfectly knows $g_{k,k,n}$.

We now express the equalized user signal $y_{k,n}$ to indicate the desired signal and the interference terms as

$$\begin{aligned} y_{k,n} = & \underbrace{f_{k,n} g_{k,k,n} \mu_k \mathbf{h}_k^T \mathbf{w}_k \sqrt{p_{k,n}} s_{k,n}}_{\text{Desired signal}} \\ & + \underbrace{f_{k,n} g_{k,k,n} \mu_k \mathbf{h}_k^T \mathbf{w}_k \sum_{n'=1, n' \neq n}^{U_k} \sqrt{p_{k,n'}} s_{k,n'}}_{\text{Intra-relay interference}} \\ & + \underbrace{\sum_{k' \neq k}^{K} \sum_{n'=1}^{U_{k'}} f_{k,n} g_{k',k,n} \mu_{k'} \mathbf{h}_{k'}^T \mathbf{w}_{k'} \sqrt{p_{k',n'}} s_{k',n'}}_{\text{Inter-relay interference at 2nd hop}} \\ & + \underbrace{\sum_{k'=1}^{K} \sum_{k'' \neq k'}^{K} \sum_{n'=1}^{U_{k''}} f_{k,n} g_{k',k,n} \mu_{k'} \mathbf{h}_{k'}^T \mathbf{w}_{k''} \sqrt{p_{k'',n'}} s_{k'',n'}}_{\text{Inter-relay interference at 1st hop}} \\ & + \underbrace{\sum_{k'=1}^{K} f_{k,n} g_{k',k,n} \mu_{k'} z_{R,k'}}_{\text{Forwarding noise}} + \underbrace{f_{k,n} z_{k,n}}_{\text{Receiving noise}}. \end{aligned} \tag{2}$$

We note from (2) that the i) intra-relay interference is among the users within the same cluster; ii) 2nd hop inter-relay interference is from the $k$th relay to the users in other clusters; and iii) 1st hop inter-relay interference is due to the ZF precoder designed with the estimated channel at the BS; with perfect CSI, this term will become zero.

To mitigate intra-relay interference, users within a cluster perform SIC for which we assume, similar to [6], that the users associated with the $k$th relay are ordered in descending order of their respective path-losses. Among the users associated with the $k$th relay, the $m$th user cancels intra-relay interference from the $l$th user ($\forall\ l > m$) and finally detects its own signal while treating signal from the $n$th user ($\forall\ n < m$) as inherent intra-cluster interference.

For SIC, users require CSI of both BS-to-relay and relay-to-user links. For the CSI of the BS-to-relay link, users rely on the statistical value. This is justified as the first hop massive MIMO channel hardens for large number of BS antennas i.e., the effective BS-to-relay sub-channel can be approximated as $\mathbb{E}\left[\mathbf{h}_k^T \mathbf{w}_k\right]$. After SIC, referred to as *partially-statistical SIC*, the signal at the $n$th user associated with the $k$th relay contains the following intra-relay interference:

$$\begin{aligned} & \underbrace{f_{k,n} g_{k,k,n} \mu_k \mathbf{h}_k^T \mathbf{w}_k \sum_{n'=1}^{n-1} \sqrt{p_{k,n'}} s_{k,n'}}_{\text{Inherent intra-relay interference}} + \\ & \underbrace{\sum_{n'=n+1}^{U_k} \mu_k \left[f_{k,n} g_{k,k,n} \mathbf{h}_k^T \mathbf{w}_k - f_{k,n} \hat{g}_{k,k,n} \mathbb{E}\left[\mathbf{h}_k^T \mathbf{w}_k\right]\right] \sqrt{p_{k,n'}} s_{k,n'}}_{\text{Residual intra-relay interference due to imperfect SIC}}, \end{aligned} \tag{3}$$

where the first-term is the inherent intra-relay interference and the second-term is the residual intra-relay interference due to imperfect SIC. After SIC, the post-processed signal at the $n$th user associated with the $k$th relay, $\bar{y}_{k,n}$ is given as

$$\begin{aligned} \bar{y}_{k,n} = & f_{k,n} g_{k,k,n} \mu_k \mathbf{h}_k^T \mathbf{w}_k \sqrt{p_{k,n}} s_{k,n} \\ & + \sum_{n'=n+1}^{U_k} \mu_k \left[f_{k,n} g_{k,k,n} \mathbf{h}_k^T \mathbf{w}_k - f_{k,n} \hat{g}_{k,k,n} \mathbb{E}\left[\mathbf{h}_k^T \mathbf{w}_k\right]\right] \sqrt{p_{k,n'}} s_{k,n'} \\ & + f_{k,n} g_{k,k,n} \mu_k \mathbf{h}_k^T \mathbf{w}_k \sum_{n'=1}^{n-1} \sqrt{p_{k,n'}} s_{k,n'} \\ & + \sum_{k' \neq k}^{K} \sum_{n'=1}^{U_{k'}} f_{k,n} g_{k',k,n} \mu_{k'} \mathbf{h}_{k'}^T \mathbf{w}_{k'} \sqrt{p_{k',n'}} s_{k',n'} \\ & + \sum_{k'=1}^{K} \sum_{k'' \neq k}^{K} \sum_{n'=1}^{U_{k''}} f_{k,n} g_{k',k,n} \mu_{k'} \mathbf{h}_{k'}^T \mathbf{w}_{k''} \sqrt{p_{k'',n'}} s_{k'',n'} \\ & + \sum_{k'=1}^{K} f_{k,n} g_{k',k,n} \mu_{k'} z_{R,k'} + f_{k,n} z_{k,n}. \end{aligned} \tag{4}$$

**Channel estimation:** In the proposed method, the BS-to-relay and relay-to-users CSI can be obtained at the BS and at the users by transmitting only $K$ pilots from the relay. For the BS to estimate the uplink channel from the $k$th relay to the BS i.e., $\mathbf{h}_k$, and the users to estimate the downlink channel from the $k'$th relay to the $n$th user associated with the $k$th relay i.e. $g_{k',k,n}$, the $k$th relay transmits $\tau$-length mutually orthogonal pilot sequences $\boldsymbol{\psi}_k$ to the BS and the users. The pilot sequence transmitted by the $k$th relay $\boldsymbol{\psi}_k \in \mathbb{C}^{\tau \times 1}$ satisfies $\|\boldsymbol{\psi}_k\|^2 = 1$, $\boldsymbol{\psi}_k^H \boldsymbol{\psi}_{k'} = 0$ for $k \neq k'$ and $\tau \geq K$.

*BS-to-relay channel state acquisition:* We express the channel from the $k$th relay to the BS, $\mathbf{h}_k$ as $\mathbf{h}_k = \sqrt{\alpha_k} \bar{\mathbf{h}}_k$ where the vector $\bar{\mathbf{h}}_k$ represents small scale fading with i.i.d. $\mathcal{CN}(0,1)$ elements and the scalar $\alpha_k$ represents the large-scale fading coefficient between the BS and the $k$th relay [7]. The pilot signal received at the BS, $\mathbf{Y}_K^{\mathrm{BP}} \in \mathbb{C}^{N_T \times \tau}$ is

$$\mathbf{Y}^{\mathrm{BP}} = \sqrt{\tau p_p} \sum_{k=1} \mathbf{h}_k \boldsymbol{\psi}_k^H + \mathbf{N}^{\mathrm{BP}}, \tag{5}$$

where $p_p$ is the pilot transmit power, and $\mathbf{N}^{\mathrm{BP}} \sim$

$\mathcal{CN}_{N_T\times\tau}\left(\mathbf{0}_{N_T\times\tau}, \mathbf{I}_{N_T}\otimes\mathbf{I}_\tau\right)$ represents the AWGN. The channel $\mathbf{h}_k$ is estimated at the BS by projecting $\boldsymbol{\psi}_k$ onto $\mathbf{Y}^{\mathrm{BP}}$ as

$$\tilde{\mathbf{y}}_k^{\mathrm{BP}} = \mathbf{Y}^{\mathrm{BP}}\boldsymbol{\psi}_k = \sqrt{\tau p_p}\mathbf{h}_k + \mathbf{N}^{\mathrm{BP}}\boldsymbol{\psi}_k. \quad (6)$$

The channel estimation error for the MMSE estimation is [7]

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \boldsymbol{\varepsilon}_k, \text{where } \hat{\mathbf{h}}_k = \frac{\sqrt{\tau p_p}\alpha_k}{1+\tau p_p\alpha_k}\tilde{\mathbf{y}}_k^{\mathrm{BP}} \triangleq \sqrt{\eta_k}\mathbf{v}_k. \quad (7)$$

Here $\hat{\mathbf{h}}_k$ is the MMSE estimate of $\mathbf{h}_k$ [7], and $\eta_k$ and $\mathbf{v}_k$ are the large-scale and small-scale fading coefficients in $\hat{\mathbf{h}}_k$ respectively and are given as

$$\eta_k = \mathbb{E}\left[\|\hat{\mathbf{h}}_k\|^2\right] = \frac{\tau p_p\alpha_k^2}{1+\tau p_p\alpha_k} \text{ and } \mathbf{v}_k = \frac{\tilde{\mathbf{y}}^{\mathrm{BP}}}{\sqrt{1+\tau p_p\alpha_k}}. \quad (8)$$

We note that $\eta_k$ is independent of the BS antenna indices and $\mathbf{v}_k \sim \mathcal{CN}(0, \mathbf{I}_{N_T})$. We also infer from (7) that the variance of the channel estimation error is $\mathbb{E}\left[\|\boldsymbol{\varepsilon}_k\|^2\right] = (\alpha_k - \eta_k)N_T$.

*Relay-to-users channel state acquisition:* We express the channel between the $k'$th relay and the $n$th user associated with the $k$th relay as $g_{k',k,n} = \sqrt{\beta_{k',k,n}}\bar{g}_{k',k,n}$ where $\bar{g}_{k',k,n}$ represents the i.i.d. small-scale fading with pdf $\mathcal{CN}(0,1)$ and $\beta_{k',k,n}$ is the large-scale fading coefficient. The pilot signal received by the $n$th user associated with the $k$th relay is

$$\mathbf{y}_{k,n}^p = \sqrt{\tau p_p}\sum_{k'=1}^{K} g_{k',k,n}\boldsymbol{\psi}_{k'}^H + \mathbf{n}_{k,n}^{\mathrm{P}}, \quad (9)$$

where $p_p$ is the pilot power, and $\mathbf{n}_{k,n}^{\mathrm{P}} \sim \mathcal{CN}_{1\times\tau}\left(\mathbf{0}_{1\times\tau}, \mathbf{I}_\tau\right)$ denotes the AWGN noise at the $k$th relay. To estimate $g_{k',k,n}$, the received pilot signal at the users, $\boldsymbol{\psi}_{k'}$ is projected onto $\mathbf{y}_{k,n}^{\mathrm{P}}$ as $\tilde{y}_{k',k,n}^{\mathrm{P}} = \mathbf{y}_{k,n}^p\boldsymbol{\psi}_{k'} = \sqrt{\tau p_p}g_{k',k,n} + \mathbf{n}_{k,n}^{\mathrm{RP}}\boldsymbol{\psi}_{k'}$. The MMSE estimate of $g_{k',k,n}$ can then be derived as [7]

$$\hat{g}_{k',k,n} = \frac{\sqrt{\tau p_p}\beta_{k',k,n}}{1+\tau p_p\beta_{k',k,n}}\tilde{y}_{k',k,n}^p \triangleq \sqrt{\zeta_{k',k,n}}u_{k',k,n}, \quad (10)$$

where $u_{k',k,n}$ is the small scale fading coefficient with pdf $\mathcal{CN}(0,1)$ while $\zeta_{k',k,n} = \mathbb{E}[|\hat{g}_{k',k,n}|^2] = \tau p_p\beta_{k',k,n}^2/\left(1+\tau p_p\beta_{k',k,n}\right)$ is the large scale fading coefficient. The channel estimation error is $\epsilon_{k',k,n} = g_{k',k,n} - \hat{g}_{k',k,n}$. For MMSE estimation, $\epsilon_{k',k,n}$ and $\hat{g}_{k',k,n}$ are statistically independent and therefore, the variance of the channel estimation error is given as $\mathbb{E}[|\epsilon_{k',k,n}|^2] = \beta_{k',k,n} - \zeta_{k',k,n}$.

## III. Achievable Spectral Efficiency

We now derive closed form SE lower bound for the relay-aided massive MIMO-NOMA system with imperfect CSI/SIC. The SE for the $n$th user in the $k$th relay, denoted as $R_{k,n}$, can be expressed using (4) as follows

$$R_{k,n} = \frac{1}{2}\left(1-\frac{\tau}{\tau_c}\right)\mathbb{E}\left[\log_2(1+\gamma_{k,n})\right], \text{with } \gamma_{k,n} = \frac{P_{k,n,0}^o}{\sum_{i=1}^{5} I_{k,n,i}^o + 1},$$

where $P_{k,n,0}^o = \left|\mu_k f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k\right|^2 p_{k,n}$,

$$I_{k,n,1}^o = \left|f_{k,n}g_{k,k,n}\mu_k\mathbf{h}_k^T\mathbf{w}_k\right|^2\sum_{n'=1}^{n-1}p_{k,n'},$$

$$I_{k,n,2}^o = \sum_{n'=n+1}^{U_k}\left|f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k - f_{k,n}\hat{g}_{k,k,n}\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]\right|^2\mu_k^2 p_{k,n'},$$

$$I_{k,n,3}^o = \sum_{k'\neq k}^{K}\sum_{n'=1}^{U_{k'}}\left|f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k'}\right|^2 p_{k',n'},$$

$$I_{k,n,4}^o = \sum_{k'}^{K}\sum_{k''\neq k'}^{K}\sum_{n'}^{U_{k''}}\left|f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k''}\right|^2 p_{k'',n'}, \text{ and}$$

$$I_{k,n,5}^o = \sum_{k'=1}^{K}\left|f_{k,n}g_{k',k,n}\mu_{k'}\right|^2. \quad (11)$$

Using the method in [4], we express the signal received at the $n$th user associated with the $k$th relay in (4) as

$\bar{y}_{k,n} = \mu_k\mathbb{E}\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^H\mathbf{w}_k\right]\sqrt{p_{k,n}}s_{k,n} + \bar{z}_{k,n}$, where

$$\begin{aligned}\bar{z}_{k,n} &= \underbrace{\mu_k\left(f_{k,n}g_{k,k,n}\mathbf{h}_k^H\mathbf{w}_k - \mathbb{E}\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^H\mathbf{w}_k\right]\right)\sqrt{p_{k,n}}s_{k,n}}_{\text{signal leakage}} \\ &+ f_{k,n}g_{k,k,n}\mu_k\mathbf{h}_k^T\mathbf{w}_k\sum_{n'=1}^{n-1}\sqrt{p_{k,n'}}s_{k,n'} \\ &+ \sum_{n'=n+1}^{U_k}\mu_k\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k - f_{k,n}\hat{g}_{k,k,n}\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]\right]\sqrt{p_{k,n'}}s_{k,n'} \\ &+ \sum_{k'\neq k}^{K}\sum_{n'=1}^{U_{k'}}f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k'}\sqrt{p_{k',n'}}s_{k',n'} \\ &+ \sum_{k'=1}^{K}\sum_{k''\neq k'}^{K}\sum_{n'=1}^{U_{k''}}f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k''}\sqrt{p_{k'',n'}}s_{k'',n'} \\ &+ \sum_{k'=1}^{K}f_{k,n}g_{k',k,n}\mu_{k'}z_{R,k'} + z_{k,n}. \end{aligned} \quad (12)$$

The terms in $\bar{z}_{k,n}$ are random variables with realizations unknown to user. An achievable SE can hence be derived by treating these terms as additive worst case Gaussian noise [4].

*Theorem* 1. The SE of the $n$th user in the $k$th relay for a finite number of BS antennas relying on MMSE channel estimate based CSI/SIC is lower bounded as

$$\underline{R}_{k,n} = \frac{1}{2}\left(1-\frac{\tau}{\tau_c}\right)\log_2(1+\underline{\gamma}_{k,n}), \text{ with } \underline{\gamma}_{k,n} = \frac{P_{k,n,0}}{\sum_{i=0}^{5} I_{k,n,i}+1},$$

where $P_{k,n,0} = \mathbb{E}\left[\left|\mu_k\mathbb{E}\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k\right]\right|^2\right]p_{k,n}$,

$$I_{k,n,0} = \mathbb{E}\left[\left|\mu_k\left(f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k - \mathbb{E}\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k\right]\right)\right|^2\right]p_{k,n},$$

$$I_{k,n,1} = \mathbb{E}\left[\left|f_{k,n}g_{k,k,n}\mu_k\mathbf{h}_k^T\mathbf{w}_k\right|^2\right]\sum_{n'=1}^{n-1}p_{k,n'},$$

$$I_{k,n,2} = \sum_{n'=n+1}^{U_k}\mathbb{E}\left[\left|f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k - f_{k,n}\hat{g}_{k,k,n}\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]\right|^2\right]\mu_k^2 p_{k,n'},$$

$$I_{k,n,3} = \sum_{k'\neq k}^{K}\sum_{n'=1}^{U_{k'}}\mathbb{E}\left[\left|f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k'}\right|^2\right]p_{k',n'},$$

$$I_{k,n,4} = \sum_{k'}^{K}\sum_{k''\neq k'}^{K}\sum_{n'}^{U_{k''}}\mathbb{E}\left[\left|f_{k,n}g_{k',k,n}\mu_{k'}\mathbf{h}_{k'}^T\mathbf{w}_{k''}\right|^2\right]p_{k'',n'}, \text{ and}$$

$$I_{k,n,5} = \sum_{k'=1}^{K}\mathbb{E}\left[\left|f_{k,n}g_{k',k,n}\mu_{k'}\right|^2\right]. \quad (13)$$

*Proof:* The above expectation terms are evaluated in Appendix A, and are provided in Table I (Column-3) for reference. For the sake of brevity, we define $p_k = \sum_{n=1}^{U_k}p_{k,n}\ \forall k$, $N = N_T - (K-1)$, $\kappa_k = \alpha_k - \eta_k + \eta_k N\ \forall k$, and $N_c = \Gamma^2(N+0.5)/\Gamma^2(N)$. ■

**Asymptotic result derivation:** We now investigate the asymptotic behaviour of interference terms in (13) with the number of BS antennas $N_T\to\infty$. Each of the terms in $\underline{\gamma}_{k,n}$ is provided in Table I (Column-4). As $N_T\to\infty$, $N_c\approx N$, $\kappa_k\approx\eta_k N$ and therefore, $\underline{\gamma}_{k,n}$ in (13) is given as in (14) (below Table I). The $\underline{\gamma}_{k,n}$ in (14) asymptotically becomes independent of $N_T$ and the SE, therefore, saturates to a finite value.

**Table I:** Derived expectations of terms in $\underline{\gamma}_{k,n}$

| Term Desription | Term | Behaviour for finite $N_T$ | Asymptotic ($N_T \to \infty$) |
|---|---|---|---|
| Desired signal | $P_{k,n,0}$ | $\frac{\frac{\pi}{4}\eta_k\zeta_{k,k,n}N_c p_{k,n}q_k}{\eta_k N p_k + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$. | $\frac{\pi}{4}\frac{\zeta_{k,k,n}p_{k,n}q_k}{p_k}$ |
| Signal Leakage | $I_{k,n,0}$ | $\frac{\left(\beta_{k,k,n}\kappa_k-\frac{\pi}{4}\zeta_{k,k,n}\eta_k N_c\right)p_{k,n}q_k}{\eta_k N p_k + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | $\left(\beta_{k,k,n}-\frac{\pi}{4}\zeta_{k,k,n}\right)\frac{p_{k,n}q_k}{p_k}$ |
| Inherent intra-relay interference | $I_{k,n,1}$ | $\sum_{n'=1}^{n-1}\frac{\kappa_k\beta_{k,k,n}p_{k,n'}q_k}{\eta_k N p_k + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | $\frac{\beta_{k,k,n}q_k}{p_k}\sum_{n'=1}^{n-1}p_{k,n'}$ |
| Residual intra-relay interference | $I_{k,n,2}$ | $\sum_{n'=n+1}^{U_k}\frac{\left(\beta_{k,k,n}\kappa_k-\eta_k\zeta_{k,k,n}N_c\right)p_{k,n'}q_k}{\eta_k N p_k + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | $\frac{\left(\beta_{k,k,n}-\zeta_{k,k,n}\right)q_k}{p_k}\sum_{n'=n+1}^{U_k}p_{k,n'}$ |
| Inter-relay interference (2nd hop) | $I_{k,n,3}$ | $\sum_{k'\neq k}^{K}\sum_{n'=1}^{U_{k'}}\frac{\kappa_{k'}\beta_{k',k,n}p_{k',n'}q_{k'}}{\eta_{k'} N p_{k'} + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | $\sum_{k'\neq k}^{K}\beta_{k',k,n}q_{k'}$ |
| Inter-relay interference (1st hop) | $I_{k,n,4}$ | $\sum_{k'}^{K}\sum_{k''\neq k'}^{K}\sum_{n'}^{U_{k''}}\frac{\beta_{k',k,n}(\alpha_{k'}-\eta_{k'})p_{k'',n'}q_{k'}}{\eta_{k'} N p_{k'} + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | 0 |
| Forwarding Noise | $I_{k,n,5}$ | $\sum_{k'=1}^{K}\frac{\beta_{k',k,n}q_{k'}}{\eta_{k'} N p_{k'} + \sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$ | 0 |

$$\underline{\gamma}_{k,n} \xrightarrow{N_T\to\infty} \frac{\frac{\pi}{4}\zeta_{k,k,n}p_{k,n}q_k}{\left(\beta_{k,k,n}-\frac{\pi}{4}\zeta_{k,k,n}\right)q_kp_{k,n}+\sum_{n'=1}^{n-1}\beta_{k,k,n}q_kp_{k,n'}+\sum_{n'=n+1}^{U_k}(\beta_{k,k,n}-\zeta_{k,k,n})q_kp_{k,n'}+\sum_{k'\neq k}^{K}\beta_{k',k,n}q_{k'}p_k+p_k}. \tag{14}$$

**Intuitive insights for derived lower bound using finite and asymptotic $N_T$ values:** We infer the following from Table I:

- The residual intra-relay interference term ($I_{k,n,2}$), which strongly depends on the quality of CSI of the relay-to-user channel, is finite, and *cannot be ignored even as $N_T \to \infty$.*
- The inter-relay interference from the 1st hop ($I_{k,n,4}$) varies inversely with $N_T$, and becomes negligible even with imperfect CSI. The relay forwarding noise ($I_{k,n,5}$) also varies inversely with $N_T$, and becomes negligible asymptotically.
- The asymptotic SE in (14) becomes independent of the first-hop path-loss $\eta_k$ and $\alpha_k$.

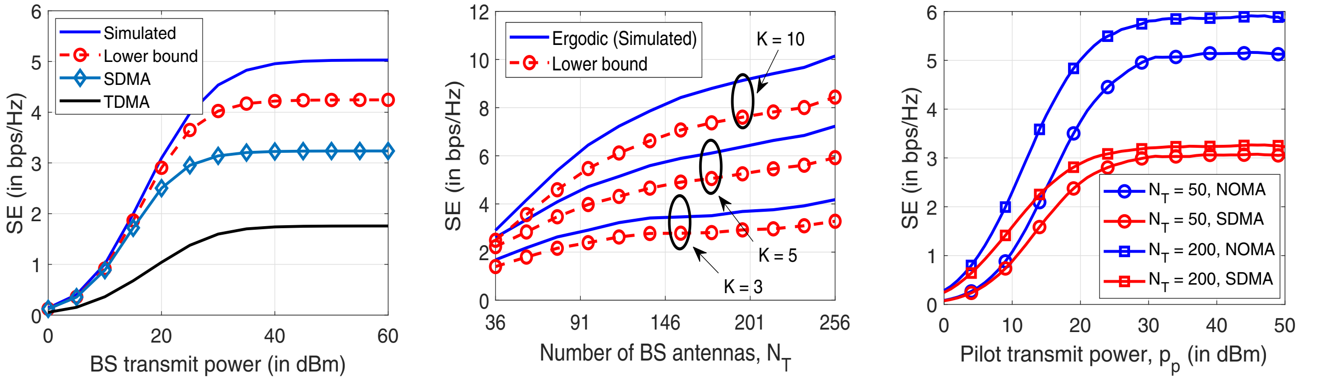
## IV. Simulation Results

We now numerically validate the derived lower-bound for a multiple-relay-aided massive MIMO NOMA system with imperfect CSI (resp. SIC) at the BS (resp. users). We consider a 20 MHz system with $N_T = 100$ BS antennas, with $K = 5$ relays and total 20 users. We assume that the i) relays and the users are distributed uniformly on a circle of radius 500 m and radius 800 m respectively, with the BS at the center; and ii) each relay is allocated four users, i.e. $U_k = 4$. We set the length of the coherence interval $\tau_c = 196$ symbols, training length $\tau = K$, pilot power $p_p = 30$ dBm, and circuit power consumption $p_c = 30$ dBm. We model the large scale fading coefficient as $\beta_{k',k,n}[\text{dB}] = \Upsilon - 10\alpha\log_{10}\left(\frac{d_{k',k,n}}{1\text{km}}\right) + F_{k',k,n}$, where $\Upsilon$ denotes the median channel gain at a reference distance of 1 km, $\alpha$ is the path-loss exponent, $d_{k',k,n}$ represents the separation distance between the $k'$th relay and the $n$th user associated with the $k$th relay, and $F_{k',k,n} \sim \mathcal{N}\left(0, \sigma_{\text{sf}}^2\right)$ is the shadow fading term. We set $\Upsilon$ to be $-130$ dB, path loss coefficient $\alpha$ to be 3.76 and $\sigma_{\text{sf}}^2$ to be 10 dB.

**Validation of the derived lower-bound in** (13)**:** We do this by numerically comparing lower-bound in (13) in Fig. 2 with its exact expression in (11). We first plot in Fig. 2a the SE versus BS transmit power by fixing maximum relay transmit power as $Q_T = 25$ dBm. We observe that the gap between the lower-bound and the exact expression is marginal. We also see that the SE initially increases with increase in BS transmit power, and then saturates around 40 dBm due to the intra- and inter-relay interference. We also compare the relay-aided NOMA scheme with spatial division multiple access (SDMA) and OMA time division for multiple access (TDMA) schemes. In TDMA scheme, one user is served in one time slot via the closest relay, while in the SDMA scheme, in one time slot, users equal to the number of relays are served via $K$ relays simultaneously. We observe that NOMA significantly outperforms both SDMA and TDMA counterparts.

We next plot in Fig. 2b the SE versus number of BS antennas $N_T$ for different values of $K$ by fixing maximum BS and relay transmit power to be 30 dBm and 25 dBm, respectively. We see that the SE lower bound closely matches the ergodic expression for all $N_T$ values. We observe that the SE increases with increase in $N_T$ because large number of antennas helps in mitigating the signal leakage and the intra/inter relay interferences. We also observe this from the theoretical evaluation of asymptotic SE terms in Table I.

**Channel estimation performance:** We next investigate in Fig. 2c the impact of varying pilot power in the first hop (BS-to-relay) and the second hop (relay-to-users), on the SE for both NOMA and SDMA schemes. For this study, we set the BS transmit power $P_T = 30$ dBm and the relay transmit power $Q_T = 25$ dBm. We observe from Fig. 2b that for lower pilot power $p_p$ both NOMA and SDMA schemes have similar SE. As we increase $p_p$, NOMA, however, performs better than SDMA, for example, with $p_p = 30$ dBm and $N_T = 200$, NOMA achieves GEE which is 50% higher than that for SDMA. This is because users utilize the first hop CSI to design the precoder and second-hop CSI to perform SIC for canceling intra-cluster interference. NOMA performance, therefore, very strongly depends on $p_p$. We conclude that with the perfect CSI, NOMA considerably outperforms SDMA. The




**(a)** **(b)** **(c)**

**Fig. 2:** (a) SE vs BS transmit power, (b) SE vs number of BS antennas, and (c) SE vs pilot transmit power ($p_p$).

NOMA performance, will be limited by CSI, since the residual interference due to imperfect SIC will degrade NOMA spectral gains. This study, therefore, captures the effect of channel estimation error on the NOMA performance, which [6] did not.

## V. Conclusion

We considered a multi-relay-aided massive MIMO NOMA system, and derived a lower-bound for its ergodic SE with imperfect CSI and SIC, and characterize its degradation with respect to orthogonal access. We also showed that a multi-relay massive MIMO NOMA system outperforms its orthogonal counterpart only with accurate channel information.

## Appendix A

We first simplify few expressions, needed for derivations.

*Derivations of expectations involving $\mathbf{h}_k^T\mathbf{w}_k$:* We note from Section II that $\mathbf{w}_k$ is the projection of $\hat{\mathbf{h}}_k = \sqrt{\eta_k}\mathbf{v}_k$ in the null-space of $\mathbf{H}_k$ of size $(N_T \times (K-1))$, and therefore, $|\hat{\mathbf{h}}_k^T\mathbf{w}_k|^2$ has scaled $\chi^2\left(2\left(N_T-(K-1)\right)\right)$ distribution by a factor $\eta_k/2$ . Therefore, $\mathbb{E}\left[|\hat{\mathbf{h}}_k^T\mathbf{w}_k|^2\right] = \eta_k(N_T-(K-1))$ and using (7) we compute $\mathbb{E}\left[\left|\mathbf{h}_k^T\mathbf{w}_{k'}\right|^2\right]$, $\forall k'$ as

$$\mathbb{E}\left[\left|\mathbf{h}_k^T\mathbf{w}_{k'}\right|^2\right] = \mathbb{E}\left[\left|\left(\hat{\mathbf{h}}_k^T+\boldsymbol{\varepsilon}_k^T\right)\mathbf{w}_{k'}\right|^2\right] = \mathbb{E}\left[\left|\boldsymbol{\varepsilon}_k^T\mathbf{w}_{k'}\right|^2\right]$$

$$= \mathbb{E}\left[\left|\boldsymbol{\varepsilon}_k\right|^2\right]\mathbb{E}\left[\left|\mathbf{w}_{k'}\right|^2\right]/N_T = (\alpha_k-\eta_k),\ \forall k' \neq k, \quad (17)$$

$$\mathbb{E}\left[\left|\mathbf{h}_k^T\mathbf{w}_k\right|^2\right] = \mathbb{E}\left[\left|\left(\hat{\mathbf{h}}_k^T+\boldsymbol{\varepsilon}_k^T\right)\mathbf{w}_k\right|^2\right] = \mathbb{E}\left[\left|\hat{\mathbf{h}}_k^T\mathbf{w}_k\right|^2\right] + \mathbb{E}\left[\left|\boldsymbol{\varepsilon}_k^T\mathbf{w}_k\right|^2\right]$$

$$= \eta_k(N_T-(K-1))+(\alpha_k-\eta_k) = \alpha_k+\eta_k\left(N_T-K\right). \quad (18)$$

To compute $\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]$, we substitute (7) into $\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]$ as

$$\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right] = \mathbb{E}\left[\left(\sqrt{\eta_k}\mathbf{v}_k^T+\boldsymbol{\varepsilon}_k^T\right)\mathbf{w}_k\right] = \sqrt{\eta_k}\mathbb{E}\left[\mathbf{v}_k^T\mathbf{w}_k\right]$$

$$= \sqrt{\eta_k}\frac{\Gamma\left(N_T-(K-1)+\frac{1}{2}\right)}{\Gamma\left(N_T-(K-1)\right)}, \quad (19)$$

where $\Gamma(.)$ denotes the gamma function and $|\hat{\mathbf{h}}_k^T\mathbf{w}_k|$ has scaled chi-distribution with a factor $\sqrt{\eta_k/2}$.

*Derivation of expectations involving $f_{k,n}g_{k,n}$:* By using the independence of channels $g_{k',k,n}$ and $g_{k,k,n}$, we deduce that $\mathbb{E}\left[f_{k,n}g_{k',k,n}\right] = \mathbb{E}\left[\frac{\hat{g}_{k,k,n}^*}{|\hat{g}_{k,k,n}|}\right]\mathbb{E}\left[g_{k',k,n}\right] = 0,\ \forall\ k' \neq k$. Next, by using the fact that $\epsilon_{k',k,n} = g_{k',k,n} - \hat{g}_{k',k,n}$ and substituting (10) in $\mathbb{E}\left[f_{k,n}g_{k,k,n}\right]$ we obtain

$$\mathbb{E}\left[f_{k,n}g_{k,k,n}\right] = \mathbb{E}\left[\frac{\hat{g}_{k,k,n}^*}{|\hat{g}_{k,k,n}|}g_{k,k,n}\right] = \mathbb{E}\left[|\hat{g}_{k,k,n}|\right]. \quad (20)$$

Here we have exploited the fact that $\hat{g}_{k,k,n}$ and $\epsilon_{k,k,n}$ are independent in the last step. Next, we know that $\hat{g}_{k,k,n} = \sqrt{\zeta_{k,k,n}}u_{k,k,n}$ where $u_{k,k,n} \sim \mathcal{CN}(0,1)$, and hence, $|u_{k,k,n}|^2 \sim \exp(1)$. Therefore, $\mathbb{E}\left[f_{k,n}g_{k,k,n}\right] = \mathbb{E}\left[|\hat{g}_{k,k,n}|\right] = \sqrt{\pi\zeta_{k,k,n}}/2$. From $f_{k,n} = \hat{g}_{k,k,n}^*/|\hat{g}_{k,k,n}|$ and using the fact $|u_{k,k,n}|^2 \sim \exp(1)$, we can derive $\mathbb{E}\left[\left|f_{k,n}g_{k',k,n}\right|^2\right]$ as

$$\mathbb{E}\left[\left|f_{k,n}g_{k',k,n}\right|^2\right] = \mathbb{E}\left[\left|f_{k,n}\right|^2\right]\mathbb{E}\left[\left|g_{k',k,n}\right|^2\right] = \beta_{k',k,n}\ \forall k' \neq k,$$

$$\mathbb{E}\left[\left|f_{k,n}g_{k,k,n}\right|^2\right] = \mathbb{E}\left[\left|\frac{\hat{g}_{k,k,n}^*}{|\hat{g}_{k,k,n}|}g_{k,k,n}\right|^2\right] = \beta_{k,k,n}. \quad (21)$$

*Computing the fixed-relaying gain, $\mu_k$:* By using independence of $z_{R,k}$ and $\hat{h}_k$ and by using (1)

$$\mu_k = \sqrt{\frac{q_k}{\mathbb{E}\left[\left|\sum_{k'}\mathbf{h}_k^T\mathbf{w}_{k'}x_k+z_{R,k}\right|^2\right]}} = \sqrt{\frac{q_k}{\eta_k N p_k + \sum_{k'=1}^{K}(\alpha_{k'}-\eta_{k'})p_{k'}+1}}.$$

We now evaluate the expectation terms in $\underline{\gamma}_{k,n}$ in detail. Using (18), (21) and $\mu_k$, we first compute the desired signal term, $P_{k,n,0}$ as $P_{k,n,0} = \mathbb{E}\left[\mu_k\mathbb{E}\left[f_{k,n}g_{k,k,n}\mathbf{h}_k^T\mathbf{w}_k\right]^2\right]p_{k,n} = \mu_k^2 p_{k,n}\mathbb{E}\left[|f_{k,n}g_{k,k,n}|\right]^2\left|\mathbb{E}\left[\mathbf{h}_k^T\mathbf{w}_k\right]\right|^2 = \frac{0.25\pi\eta_k\zeta_{k,k,n}p_{k,n}q_k N_c}{\eta_k N p_k+\sum_{j=1}^{K}(\alpha_j-\eta_j)p_j+1}$, where for ease of notation, we define $p_k = \sum_{n=1}^{U_k}p_{k,n}\ \forall k$, $N = N_T-(K-1)$, $\kappa_k = \alpha_k-\eta_k+\eta_k N\ \forall k$, and $N_c = \Gamma^2\left(N+\frac{1}{2}\right)/\Gamma^2(N)$. The interference terms $(I_{k,n,0-5})$ can be computed similarly and have been provided in Table I.

## 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] H. V. Cheng, E. Björnson, and E. G. Larsson, "Performance analysis of NOMA in training-based multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 372–385, 2018.

[3] X. Yue, Z. Qin, Y. Liu, S. Kang, and Y. Chen, "A unified framework for non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5346–5359, 2018.

[4] Y. Li and G. Amarasuriya, "Relay-aided massive MIMO NOMA downlink," in *IEEE Global Communications Conference, GLOBECOM 2018, Abu Dhabi, United Arab Emirates, December 9-13, 2018*, 2018, pp. 1–7.

[5] 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.

[6] 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.

[7] S. M. Kay, "Fundamentals of statistical signal processing, volume I: estimation theory," 1993.