

Spectral Efficiency in Large Intelligent Surfaces: Asymptotic Analysis Under Pilot Contamination

Minchae Jung,¹ Walid Saad,² and Gyuyeol Kong¹

¹School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea,

Emails: {hosaly, gykong}@yonsei.ac.kr.

²Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, VA, USA,

Email: walids@vt.edu.

Abstract—Large intelligent surfaces (LISs) have emerged as a new and promising wireless communication paradigm that relies on equipping man-made structures such as walls with a massive number of antennas. However, despite their potential benefits, a fundamental analysis on the performance limits of LIS systems is lacking. In this paper, the system spectral efficiency (SSE) of an uplink LIS system is asymptotically analyzed under a practical frame structure and LIS environment. In order to quantify the impact on the SSE of pilot contamination, the SSE of a multi-LIS system is asymptotically studied and a theoretical bound on its performance is derived. Simulation results show that the derived analyses are in close agreement with the exact mutual information in presence of a large number of antennas. Moreover, the results show that the achievable SSE is limited by the effect of pilot contamination and intra/inter-LIS interference through the line-of-sight path, even if the LIS is equipped with an infinite number of antennas.

I. INTRODUCTION

THE notion of a large intelligent surface (LIS) that relies on equipping man-made structures, such as buildings, with massive antenna arrays is rapidly emerging as a key enabler of tomorrow's Internet of Things (IoT) and 6G applications [1]–[6]. An LIS system can potentially provide pervasive and reliable wireless connectivity by exploiting the fact that pervasive city structures, such as buildings, roads, and walls, will become electromagnetically active in the near future. If properly operated and deployed, the entire environment is expected to be active in wireless transmission providing near-field communications. In contrast, conventional massive multiple-input multiple-output (MIMO) systems is essentially regarded as far-field communications generating non-line-of-sight (NLOS) channels with a high probability. Indeed, the wireless channels of an LIS can become nearly line-of-sight (LOS) channels, resulting in several advantages compared to conventional massive MIMO system. First, noise and inter-user interference through a NLOS path become negligible as the number of antenna arrays on LIS increases [3]. Also, the inter-user interference through a LOS path is negligible providing an interference-free environment, when the distances between adjacent devices are larger than half the wavelength [4], [5]. Moreover, an LIS offers more reliable and space-intensive communications compared to conventional massive MIMO systems as clearly explained in [3] and [6].

However, recent works in [3]–[6] have not considered the effects on spectral efficiency (SE) resulting from the use of a practical uplink frame structure in which the pilot training and data transmission period are jointly considered. Given that statistical channel state information (CSI) is typically acquired by pilot signaling, and because the channel uses for data transmission are closely related to the length of the pilot sequence [7], an uplink frame structure that includes pilot training strongly impacts the achievable SE of LIS systems. Moreover, this pilot signal will be contaminated by inter-LIS interference, similar to inter-cell interference in multi-cell MIMO environment (e.g., see [8] and [9]). Therefore, accurate SE analysis constitutes an important challenge in multi-LIS systems where the pilot sequences are reused in adjacent LISs. In fact, prior studies on massive MIMO [7]–[9] do not directly apply to LIS, because the channel model of LIS is significantly different from the one used in these prior studies. For densely located LISs, all channels will be modeled by device-specific spatially correlated Rician fading depending on the distance between each LIS and device, however, the massive MIMO works in [7]–[9] rely on a Rayleigh fading channel considering far-field communications. Moreover, in LIS, each area of the large surface constitutes one of the key parameters that determine the performance of an LIS system [3]–[5], however, in existing massive MIMO works [7]–[9], this notion of an area is not applicable.

The main contribution of this paper is an asymptotic analysis of the uplink system SE (SSE) in a multi-LIS environment that considers a practical uplink frame structure based on the 3GPP model in [10]. We analyze the asymptotic SSE including its ergodic value, channel hardening effect, and performance bound, under pilot contamination considerations, relying on a scaling law for a large number of antennas. The devised approximation allows for accurate estimations of the SSE, deterministically, and it also allows verifying the reliability of an LIS system. Simulation results show that the impact on an LIS system of pilot contamination can be negligible when inter-LIS interference channels are generated from spatially correlated Rayleigh fading, which highlights a significant difference from conventional massive MIMO.

The rest of this paper is organized as follows. Section II presents the LIS-based system model. Section III describes the asymptotic analysis of the SSE and its performance bound. Simulation results are provided in Section IV to support and verify the analyses, and Section V concludes the paper.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1A6A3A11936259) and by the U.S. National Science Foundation under Grants CNS-1836802 and OAC-1638283.

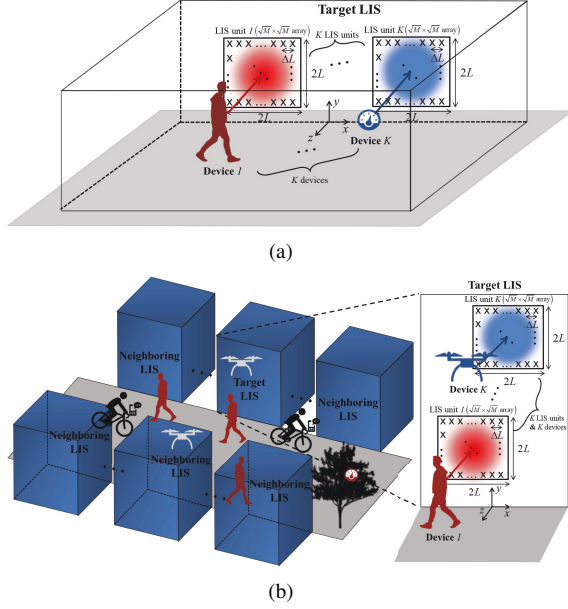


Fig. 1. Illustrative system model of the uplink LIS under consideration of (a) indoor case with single LIS, (b) outdoor case with multiple LISs.

II. SYSTEM MODEL

Consider an uplink LIS system with $N \geq 1$ LISs sharing the same frequency band. Each LIS is located in two-dimensional Cartesian space along the xy -plane, serving K devices, as shown in Fig. 1. Each LIS is composed of K LIS units, each of which serves a single-antenna device occupying a $2L \times 2L$ square shaped subarea of the entire LIS. A large number of antennas, M , are deployed on the surface of each LIS unit with ΔL spacing, arranged in a rectangular lattice centered on the (x, y) coordinates of the corresponding device. Considering the location of device k at LIS n as (x_{nk}, y_{nk}, z_{nk}) , antenna m of LIS unit k at LIS n will be located at $(x_{nkm}^{\text{LIS}}, y_{nkm}^{\text{LIS}}, 0)$ where $x_{nkm}^{\text{LIS}} \in [x_{nk} - L, x_{nk} + L]$ and $y_{nkm}^{\text{LIS}} \in [y_{nk} - L, y_{nk} + L]$. Fig. 1 illustrates our system model for an indoor case with a single LIS and an outdoor case with multiple LISs. In case of single LIS, as shown in Fig. 1(a), the desired signal is affected, exclusively, by *intra-LIS interference* which is defined as the interference generated by multiple devices located within the same LIS area. On the other hand, for the case of multiple LISs, as shown in Fig. 1(b), the desired signal can be affected by both *intra-LIS* and *inter-LIS interference* simultaneously. Here, *inter-LIS interference* corresponds to the interference generated by devices serviced by other LISs. We assume that each LIS consists of K non-overlapping LIS units that use an orthogonal multiple access scheme among devices with similar locations, and each device controls its transmission power toward the center of its LIS unit according to a target signal-to-noise-ratio (SNR), to avoid the near-far problem.

A. Wireless Channel Model

In LIS systems, entire man-made structures are electromagnetically active and can be used for wireless communication. We then consider the LIS channel $\mathbf{h}_{nnkk}^L \in \mathbb{C}^M$ between

device k at LIS n and LIS unit k part of LIS n as a LOS path defined by

$$\mathbf{h}_{nnkk}^L = [\beta_{nnkk1}^L h_{nnkk1}, \dots, \beta_{nnkkM}^L h_{nnkkM}]^T,$$

where $\beta_{nnkk}^L = \alpha_{nnkk}^L l_{nnkk}^L$ and $h_{nnkk} = \exp(-j2\pi d_{nnkk}/\lambda)$ denote a LOS channel gain and state, respectively, between device k at LIS n and antenna m of LIS unit k part of LIS n [11]. The terms $\alpha_{nnkk}^L = \sqrt{z_{nk}/d_{nnkk}}$ and $l_{nnkk}^L = 1/\sqrt{4\pi d_{nnkk}^2}$ represent, respectively, the antenna gain and free space path loss attenuation, where d_{nnkk} is the distance between device k at LIS n and antenna m of LIS unit k part of LIS n . λ is the wavelength of a signal. We model the interference channel $\mathbf{h}_{lnjk} \in \mathbb{C}^M$ between device j at LIS l and LIS unit k part of LIS n as a Rician fading channel with Rician factor κ_{lnjk} , given by $\mathbf{h}_{lnjk} = \tilde{\mathbf{h}}_{lnjk} + \tilde{\mathbf{h}}_{lnjk} = \sqrt{\frac{\kappa_{lnjk}}{\kappa_{lnjk}+1}} \mathbf{h}_{lnjk}^L + \sqrt{\frac{1}{\kappa_{lnjk}+1}} \mathbf{h}_{lnjk}^{\text{NL}}$, where $\mathbf{h}_{lnjk}^L \in \mathbb{C}^M = [\beta_{lnjk1}^L h_{lnjk1}, \dots, \beta_{lnjkM}^L h_{lnjkM}]^T$ and $\mathbf{h}_{lnjk}^{\text{NL}} \in \mathbb{C}^M = \mathbf{R}_{lnjk}^{1/2} \mathbf{g}_{lnjk}$ denote the deterministic LOS and the correlated NLOS component, respectively. Here, if $l = n$ and $j \neq k$, then $\mathbf{h}_{lnjk}$ indicates the intra-LIS interference channel, otherwise, if $l \neq n \forall j, k$, then $\mathbf{h}_{lnjk}$ indicates the inter-LIS interference channel. Considering P dominant paths among all NLOS paths, we define $\mathbf{R}_{lnjk} \in \mathbb{C}^{M \times P}$ and $\mathbf{g}_{lnjk} = [g_{lnjk1}, \dots, g_{lnjkP}]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_P)$ to be the deterministic correlation matrix and an independent fast-fading channel vector between device j at LIS l and LIS unit k part of LIS n , respectively. Since the LIS is deployed on the horizontal plane, as shown in Fig. 1, we can model it as a uniform planar array [12]. Then, the correlation matrix can be defined as $\mathbf{R}_{lnjk}^{\text{NL}} = \mathbf{l}_{lnjk}^{\text{NL}} \mathbf{D}_{lnjk}$, where $\mathbf{l}_{lnjk}^{\text{NL}} = \text{diag}(l_{lnjk1}^{\text{NL}}, \dots, l_{lnjkM}^{\text{NL}})$ is a diagonal matrix that includes the path loss attenuation factors $l_{lnjk}^{\text{NL}} = d_{lnjk}^{-\beta_{\text{PL}}/2}$ with a path loss exponent β_{PL} and $\mathbf{D}_{lnjk} = [\alpha_{lnjk1}^{\text{NL}} \mathbf{d}(\phi_{lnjk1}^v, \phi_{lnjk1}^h), \dots, \alpha_{lnjkP}^{\text{NL}} \mathbf{d}(\phi_{lnjkP}^v, \phi_{lnjkP}^h)]$. $\mathbf{d}(\phi_{lnjkp}^v, \phi_{lnjkp}^h) \in \mathbb{C}^M$ is the NLOS path p at given angles of $(\phi_{lnjkp}^v, \phi_{lnjkp}^h)$ defined as:

$$\begin{aligned} \mathbf{d}(\phi_{lnjkp}^v, \phi_{lnjkp}^h) &= \frac{1}{\sqrt{M}} \mathbf{d}_v(\phi_{lnjkp}^v) \otimes \mathbf{d}_h(\phi_{lnjkp}^h), \\ \mathbf{d}_v(\phi_{lnjkp}^v) &= [1, e^{j\frac{2\pi\Delta L}{\lambda}\phi_{lnjkp}^v}, \dots, e^{j\frac{2\pi\Delta L}{\lambda}(\sqrt{M}-1)\phi_{lnjkp}^v}]^T, \\ \mathbf{d}_h(\phi_{lnjkp}^h) &= [1, e^{j\frac{2\pi\Delta L}{\lambda}\phi_{lnjkp}^h}, \dots, e^{j\frac{2\pi\Delta L}{\lambda}(\sqrt{M}-1)\phi_{lnjkp}^h}]^T, \end{aligned}$$

where $\phi_{lnjkp}^v = \sin \theta_{lnjkp}^v$ and $\phi_{lnjkp}^h = \sin \theta_{lnjkp}^h \cos \theta_{lnjkp}^h$ when the elevation and azimuth angles of path p between device j at LIS l and LIS unit k part of LIS n are θ_{lnjkp}^v and θ_{lnjkp}^h , respectively. Further, $\alpha_{lnjkp}^{\text{NL}} = \sqrt{\cos \theta_{lnjkp}^v \cos \theta_{lnjkp}^h}$ denotes the antenna gain of path p with $\theta_{lnjkp} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\theta_{lnjkp} \in \{\theta_{lnjkp}^v, \theta_{lnjkp}^h\}$.

B. Uplink Pilot Training

We consider that a matched filter (MF) is used at the LIS to amplify the desired signals and suppress interfering signals. This MF receiver requires CSI which can be estimated by pilot signaling with known pilot signals being transmitted

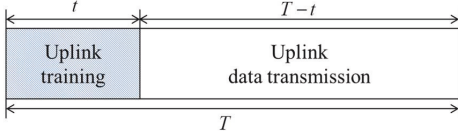


Fig. 2. Illustrative uplink frame structure with a pilot training period t and a data transmission period $T - t$.

from the device to the LIS. The device transmits its data signals immediately after sending the pilot signals within the channel coherence time T in which the uplink channel is approximately static. We consider the uplink frame structure shown in Fig. 2, in which the total duration of T channel uses is divided into a t period used for pilot training and a $T - t$ period used for data transmission. Every device simultaneously transmits $t \geq K$ orthogonal pilot sequence over the uplink channel to the LIS, so that the required CSI can be acquired. Given that those K pilot sequences are pairwise orthogonal to each other, we have $\Psi^H \Psi = \mathbf{I}_K$, where $\Psi = [\psi_1, \dots, \psi_K]$ and ψ_k is the $t \times 1$ pilot sequence for device k .

For the multi-LIS scenario in which the same frequency band is shared by all LISs and adjacent LISs reuse the pilot sequences, the pilot symbols between adjacent LISs are no longer orthogonal to each other and this non-orthogonality causes *pilot contamination*. In large antenna-array systems such as massive MIMO and LIS, the performance can be dominantly limited by residual interference from pilot contamination as explained in [8] and [9]. Since LISs will be located more densely than BSs, the LIS channels associated with pilot contamination will be significantly different than those of massive MIMO, and hence, prior studies on pilot contamination for massive MIMO [7]–[9] do not directly apply to LIS. In order to verify the effect of pilot contamination in an LIS system theoretically, we consider such multi-LIS scenario in which a total of N LISs share the same frequency band and each LIS reuses K pilot sequences. Moreover, all LISs are assumed to use the same uplink frame structure shown in Fig. 2, whereby a pilot sequence k is allocated to device k for all $1 \leq k \leq K$. On the basis of orthogonal characteristic of the pilot sequences, each LIS unit k multiplies the received pilot signal by ψ_k for channel estimation, as given by

$$\mathbf{Y}_{nk}^p \psi_k = \sqrt{t\rho_{p_{nk}}} \mathbf{h}_{nnkk}^L + \sum_{l \neq n} \sqrt{t\rho_{p_{lk}}} \mathbf{h}_{lnkk} + \mathbf{N}_{nk} \psi_k,$$

where $\rho_{p_{nk}}$ and $\rho_{p_{lk}}$ are the transmit SNRs for the pilot symbols of device k at LIS n and device k at LIS l , respectively, and $\mathbf{N}_{nk} \in \mathbb{C}^{M \times t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is a noise matrix at LIS unit k part of LIS n . In most prior research on pilot contamination in large antenna-array systems such as in [8] and [13], the minimum mean square error (MMSE) channel estimator is assumed to estimate a desired channel given that the BS has knowledge of every correlation matrix between itself and interfering users located in adjacent cells. However, this assumption is impractical for LIS systems because a massive number of devices will be connected to an LIS, and thus, processing complexity will increase tremendously when estimating and sharing device information. Therefore,

we consider a simple least square (LS) estimator which does not require such information as a practical alternative [14]. The LS estimate of the deterministic desired channel $\mathbf{h}_{nnkk}^L$ is then obtained as follows:

$$\hat{\mathbf{h}}_{nnkk} = \mathbf{h}_{nnkk}^L + \mathbf{e}_{nk}, \quad (1)$$

where $\mathbf{e}_{nk}$ indicates the estimation error vector given by

$$\mathbf{e}_{nk} = \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}}}{\rho_{p_{nk}}}} (\tilde{\mathbf{h}}_{lnkk} + \tilde{\mathbf{h}}_{lnkk}) + \frac{1}{\sqrt{t\rho_{p_{nk}}}} \mathbf{w}_{nk},$$

where $\mathbf{w}_{nk} = [w_{nk1}, \dots, w_{nkM}]^T \in \mathbb{C}^M \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$.

C. Uplink SSE

The uplink signal received from all devices at LIS unit k part of LIS n is given by:

$$\mathbf{y}_{nk} = \sqrt{\rho_{nk}} \mathbf{h}_{nnkk}^L x_{nk} + \sum_{j \neq k} \sqrt{\rho_{nj}} \mathbf{h}_{nnjk} x_{nj} + \sum_{l \neq n} \sum_{j=1}^K \sqrt{\rho_{lj}} \mathbf{h}_{lnjk} x_{lj} + \mathbf{n}_{nk},$$

where x_{nk} , x_{nj} , and x_{lj} are uplink transmit signals of device k at LIS n , device j at LIS n , and device j at LIS l , respectively, and ρ_{nk} , ρ_{nj} , and ρ_{lj} are their transmit SNRs. Also, $\mathbf{n}_{nk} \in \mathbb{C}^M \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the noise vector at LIS unit k part of LIS n . We consider an MF receiver such that $\mathbf{f}_{nk} = \hat{\mathbf{h}}_{nnkk}$. Under the imperfect CSI results from an LS estimator, $\mathbf{f}_{nk}$ can be obtained from (1) as $\mathbf{f}_{nk} = \mathbf{h}_{nnkk}^L + \mathbf{e}_{nk}$ where $\mathbf{e}_{nk}$ is the estimation error vector uncorrelated with $\mathbf{n}_{nk}$. Therefore, the received signal-to-interference-plus-noise ratio (SINR) at LIS unit k part of LIS n will be $\gamma_{nk} = \rho_{nk} S_{nk} / I_{nk}$, where $S_{nk} = |\mathbf{h}_{nnkk}^L|^4$, and

$$I_{nk} = \rho_{nk} X_{nk} + \sum_{j \neq k} \rho_{nj} Y_{nj} + \sum_{l \neq n} \sum_{j=1}^K \rho_{lj} Y_{lnjk} + Z_{nk}, \quad (2)$$

where $X_{nk} = |\mathbf{e}_{nk}^H \mathbf{h}_{nnkk}^L|^2$, $Y_{nj} = |\hat{\mathbf{h}}_{nnkk}^H \mathbf{h}_{nnjk}|^2$, $Y_{lnjk} = |\hat{\mathbf{h}}_{nnkk}^H \mathbf{h}_{lnjk}|^2$, and $Z_{nk} = |(\mathbf{h}_{nnkk}^L)^H + \mathbf{e}_{nk}^H|^2$. Considering t and $T - t$ periods used for pilot training and data transmission, respectively, the SSE can be obtained as follows:

$$R_n^{\text{SSE}} = \left(1 - \frac{t}{T}\right) \sum_{k=1}^K \log(1 + \gamma_{nk}).$$

Given this SSE, we will be able to analyze the asymptotic value of the SSE and obtain its performance bound as M increases without bound. Note that in the following sections, we use a generalized value of $N \geq 1$ in order to analyze both single- and multi-LIS cases, simultaneously (i.e., $N = 1$ and $N \geq 2$ indicate a single-LIS and N -LIS cases, respectively).

III. ASYMPTOTIC SSE ANALYSIS

We analyze the asymptotic value of the SSE under consideration of the pilot contamination as M increases to infinity. In an uplink LIS system with MF receiver, the desired signal power, S_{nk} , converges to a deterministic value as M increases to infinity as proved in [3] and [4]:

$$S_{nk} - \bar{p}_{nk} \xrightarrow{M \rightarrow \infty} 0,$$

$$\sigma_{y_{njk}}^2 = \frac{1}{\kappa_{nnjk} + 1} \left(\sum_{p=1}^P |\bar{\mathbf{q}}_{nk}^H \mathbf{c}_{nnjkp}|^2 + \sum_{l \neq n} \frac{\rho_{p_{lk}} \left\| (\mathbf{R}_{lnkk}^{1/2})^H \mathbf{R}_{nnjk}^{1/2} \right\|_F^2}{\rho_{p_{nk}} (\kappa_{lnkk} + 1)} + \sum_{m,p}^{M,P} \frac{(\alpha_{nnjkp}^{\text{NL}} l_{nnjkm}^{\text{NL}})^2}{Mt \rho_{p_{nk}}} \right), \quad (4)$$

$$\sigma_{y_{lnjk}}^2 = \frac{1}{\kappa_{lnjk} + 1} \left(\sum_{p=1}^P |\bar{\mathbf{q}}_{nk}^H \mathbf{c}_{lnjkp}|^2 + \sum_{l \neq n} \frac{\rho_{p_{lk}} \left\| (\mathbf{R}_{lnkk}^{1/2})^H \mathbf{R}_{lnjk}^{1/2} \right\|_F^2}{\rho_{p_{nk}} (\kappa_{lnkk} + 1)} + \sum_{m,p}^{M,P} \frac{(\alpha_{lnjkp}^{\text{NL}} l_{lnjkm}^{\text{NL}})^2}{Mt \rho_{p_{nk}}} \right). \quad (5)$$

where $\bar{p}_{nk} = \frac{M^2 p_{nk}^2}{16\pi^2 L^4}$ and $p_{nk} = \tan^{-1} \left(L^2 / (z_{nk} \sqrt{2L^2 + z_{nk}^2}) \right)$. Given the definition of $\gamma_{nk} = \rho_{nk} S_{nk} / I_{nk}$, we have $\gamma_{nk} - \bar{\gamma}_{nk} \xrightarrow{M \rightarrow \infty} 0$, where

$$\bar{\gamma}_{nk} = \frac{\rho_{nk} p_{nk}^2}{16\pi^2 L^4 I_{nk} / M^2}, \quad (3)$$

We can observe from (3) that the distribution of $\bar{\gamma}_{nk}$ depends exclusively on the distribution of I_{nk} . In order to analyze the distribution of I_{nk} theoretically, we derive the following lemmas. Here, we define $\mathbf{R}_{lnjk}^{1/2} = [\mathbf{c}_{lnjk1}, \dots, \mathbf{c}_{lnjkP}] = [\mathbf{r}_{lnjk1}^H, \dots, \mathbf{r}_{lnjkM}^H]^H$, where $\mathbf{c}_{lnjkp} \in \mathbb{C}^{M \times 1}$ and $\mathbf{r}_{lnjkm} \in \mathbb{C}^{1 \times P}$.

Lemma 1. The mean of X_{nk} is obtained by $\mu_{X_{nk}} = \sigma_{x_{nk}}^2 + |\mu_{x_{nk}}|^2$, where

$$\mu_{x_{nk}} = \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}}}{\rho_{p_{nk}}}} \bar{\mathbf{h}}_{lnkk}^H \mathbf{h}_{nnkk}^L, \quad \sigma_{x_{nk}}^2 = \sum_{l \neq n} \frac{\rho_{p_{lk}}}{\rho_{p_{nk}} (\kappa_{lnkk} + 1)} \sum_{p=1}^P |\mathbf{c}_{lnkkp}^H \mathbf{h}_{nnkk}^L|^2 + \sum_{m=1}^M \frac{\beta_{nnkk}^2}{t \rho_{p_{nk}}}.$$

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

Lemma 2. The mean values of Y_{njk} and Y_{lnjk} follow $\mu_{Y_{njk}} - \bar{\mu}_{Y_{njk}} \xrightarrow{M \rightarrow \infty} 0$ and $\mu_{Y_{lnjk}} - \bar{\mu}_{Y_{lnjk}} \xrightarrow{M \rightarrow \infty} 0$, respectively, where

$$\bar{\mu}_{Y_{njk}} = \sigma_{y_{njk}}^2 + |\mu_{y_{njk}}|^2, \quad \bar{\mu}_{Y_{lnjk}} = \sigma_{y_{lnjk}}^2 + |\mu_{y_{lnjk}}|^2,$$

where

$$\mu_{y_{njk}} = (\mathbf{h}_{nnkk}^L)^H \bar{\mathbf{h}}_{nnjk} + \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}}}{\rho_{p_{nk}}}} \bar{\mathbf{h}}_{lnkk}^H \bar{\mathbf{h}}_{nnjk}, \quad \sigma_{y_{njk}}^2 = \sigma_{y_{njk}^{\text{EL}}}^2 + \sigma_{y_{njk}^{\text{EN}}}^2, \quad \sigma_{y_{njk}^{\text{EL}}}^2 = \sum_{l \neq n} \frac{\rho_{p_{lk}}}{\rho_{p_{nk}}} \sum_{p=1}^P \frac{|\mathbf{c}_{lnkkp}^H \bar{\mathbf{h}}_{nnjk}|^2}{(\kappa_{lnkk} + 1)} + \sum_{m=1}^M \frac{\beta_{nnjkm}^2}{t \rho_{p_{nk}}},$$

and

$$\mu_{y_{lnjk}} = (\mathbf{h}_{nnkk}^L)^H \bar{\mathbf{h}}_{lnjk} + \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}}}{\rho_{p_{nk}}}} \bar{\mathbf{h}}_{lnkk}^H \bar{\mathbf{h}}_{lnjk}, \quad \sigma_{y_{lnjk}}^2 = \sigma_{y_{lnjk}^{\text{EL}}}^2 + \sigma_{y_{lnjk}^{\text{EN}}}^2, \quad \sigma_{y_{lnjk}^{\text{EL}}}^2 = \sum_{l \neq n} \frac{\rho_{p_{lk}}}{\rho_{p_{nk}}} \sum_{p=1}^P \frac{|\mathbf{c}_{lnkkp}^H \bar{\mathbf{h}}_{lnjk}|^2}{(\kappa_{lnkk} + 1)} + \sum_{m=1}^M \frac{\beta_{lnjkm}^2}{t \rho_{p_{nk}}}.$$

Also, $\sigma_{y_{njk}^{\text{EN}}}^2$ and $\sigma_{y_{lnjk}^{\text{EN}}}^2$ are given by (4) and (5), respectively, where

$$\bar{\mathbf{q}}_{nk} = \mathbf{h}_{nnkk}^L + \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}}}{\rho_{p_{nk}}}} \bar{\mathbf{h}}_{lnkk}.$$

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

Lemma 3. The mean of Z_{nk} is obtained by $\mu_{Z_{nk}} = \sum_{m=1}^M \left(\sigma_{z_{nkm}}^2 + |\mu_{z_{nkm}}|^2 \right)$, where

$$\mu_{z_{nkm}} = \beta_{nnkk}^L h_{nnkk}^* + \sum_{l \neq n} \sqrt{\frac{\rho_{p_{lk}} \kappa_{lnkk}}{\rho_{p_{nk}} (\kappa_{lnkk} + 1)}} \beta_{lnkk}^L h_{lnkk}^*, \quad \sigma_{z_{nkm}}^2 = \sum_{l \neq n} \frac{\rho_{p_{lk}} |\mathbf{r}_{lnkkm}|^2}{\rho_{p_{nk}} (\kappa_{lnkk} + 1)} + \frac{1}{t \rho_{p_{nk}}}.$$

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

In Lemmas 1–3, the variables, $\mu_{X_{nk}}$, $\bar{\mu}_{Y_{njk}}$, $\bar{\mu}_{Y_{lnjk}}$, and $\mu_{Z_{nk}}$, are obtained by the deterministic information such as the locations of the devices and covariance matrices. On the basis of Lemmas 1–3, we can asymptotically derive the mean of I_{nk} from (2) as follows: $\mu_{I_{nk}} - \bar{\mu}_{I_{nk}} \xrightarrow{M \rightarrow \infty} 0$, where

$$\bar{\mu}_{I_{nk}} = \rho_{nk} \mu_{X_{nk}} + \sum_{j \neq k} \rho_{nj} \bar{\mu}_{Y_{njk}} + \sum_{l \neq n} \sum_{j=1}^K \rho_{lj} \bar{\mu}_{Y_{lnjk}} + \mu_{Z_{nk}}.$$

Since the variance of $\bar{\gamma}_{nk}$ exclusively depends on the variance of I_{nk}/M^2 from (3), the following Lemma 4 is used to obtain $\sigma_{I_{nk}}^2/M^4$ based on the scaling law for M .

Lemma 4. According to the scaling law for M , the variance of I_{nk}/M^2 asymptotically follows $\sigma_{I_{nk}}^2/M^4 \xrightarrow{M \rightarrow \infty} 0$.

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

Lemma 4 shows that I_{nk}/M^2 converges to the deterministic value $\bar{\mu}_{I_{nk}}/M^2$ without any variance, as M increases. Then, $\bar{\gamma}_{nk}$ converges to a deterministic value as M increases, and finally, we have the following Theorem 1 related to the asymptotic convergence of R_n^{SSE} .

Theorem 1. As M increases to infinity, we have the following asymptotic convergence of SSE: $R_n^{\text{SSE}} - \bar{\mu}_n^{\text{SSE}} \xrightarrow{M \rightarrow \infty} 0$, where

$$\bar{\mu}_n^{\text{SSE}} = \left(1 - \frac{t}{T} \right) \sum_{k=1}^K \log \left(1 + \frac{\rho_{nk} p_{nk}^2}{16\pi^2 L^4 \bar{\mu}_{I_{nk}} / M^2} \right). \quad (6)$$

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

Theorem 1 shows that the multi-LIS system will experience a channel hardening effect resulting in the deterministic SSE. This deterministic SSE provides the improved system reliability and a low latency. Moreover, we can observe from (6) that the asymptotic SSE can be obtained from the deterministic information such as the locations of the devices and correlation matrices. Therefore, this asymptotic approximation enables accurate estimation of the SSE without the need for extensive simulations.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	3 GHz
Target SNR for pilot / data	0 dB / 3 dB
Coherence block length (T)	500 symbols
Length of LIS unit ($2L$)	0.5 m
Rician factor (κ [dB]) [16]	$13 - 0.03d[\text{m}]$
LOS path loss model [11]	$11 + 20\log_{10}d[\text{m}]$
NLOS path loss model [8]	$37\log_{10}d[\text{m}]$ ($\beta_{\text{PL}} = 3.7$)

Next, we utilize Theorem 1 to derive a performance bound on the SSE, asymptotically, by analyzing $\bar{\mu}_n^{\text{SSE}}$ via a scaling law for an infinite M . As M increases, the SSE converges to $\bar{\mu}_n^{\text{SSE}}$ which depends on the value of $\bar{\mu}_{I_{nk}}/M^2$, as seen from (6). Hence, in an LIS system equipped with an infinite number of antennas, it is important to obtain its limiting value, $\lim_{M \rightarrow \infty} \bar{\mu}_{I_{nk}}/M^2$, and this can provide the performance bound of the SSE, asymptotically. In order to derive the performance bound of LIS system, we first determine the scaling law of $\bar{\mu}_{I_{nk}}/M^2$ according to M . Then, we have the following result related to the performance bound of SSE.

Theorem 2. As M increases, $\bar{\mu}_n^{\text{SSE}}$ asymptotically converges to its performance bound, $\hat{\mu}_n^{\text{SSE}}$, as given by: $\bar{\mu}_n^{\text{SSE}} - \hat{\mu}_n^{\text{SSE}} \xrightarrow{M \rightarrow \infty} 0$, where

$$\hat{\mu}_n^{\text{SSE}} = \left(1 - \frac{t}{T}\right) \sum_{k=1}^K \log \left(1 + \frac{M^2 \rho_{nk} p_{nk}^2}{16\pi^2 L^4 \bar{\mu}_{I_{nk}}}\right),$$

$$\hat{\mu}_{I_{nk}} = \rho_{nk} |\mu_{x_{nk}}|^2 + \sum_{j \neq k}^K \rho_{nj} |\mu_{y_{nj}}|^2 + \sum_{l \neq n}^N \sum_{j=1}^K \rho_{lj} |\mu_{y_{ljk}}|^2.$$

Proof: Due to the page limitation, the detailed proof can be found in our technical report in [15]. ■

Based on Lemmas 1 and 2, the terms $\mu_{x_{nk}}$, $\mu_{y_{nj}}$, and $\mu_{y_{ljk}}$ in $\hat{\mu}_{I_{nk}}$ are determined by the LOS channels depending on the locations of the devices. Therefore, the asymptotic SSE performance bound can be obtained, deterministically, and that deterministic bound leads to *several important implications when evaluating an LIS system, that significantly differ from conventional massive MIMO*. First, an LIS system has a particular operating characteristic whereby the pilot contamination and intra/inter-LIS interference through the NLOS path and noise become negligible as M increases. If all of the inter-LIS interference is generated from the NLOS path, the pilot contamination and inter-LIS interference will vanish lead to a performance convergence between the SSE of single- and multi-LIS system. Moreover, unlike conventional massive MIMO in which the performance is dominantly limited by pilot contamination, the channel estimation error including pilot contamination gradually loses its effect on the SSE, and eventually, the SSE of a multi-LIS system will reach that of a single-LIS system with perfect CSI. More practically, even if all of the intra/inter-LIS interference channels are generated by device-specific spatially correlated Rician fading, an LIS system also has a particular characteristic whereby its SSE is bounded by three factors that include pilot contamination, intra-, and inter-LIS interference through the LOS path.

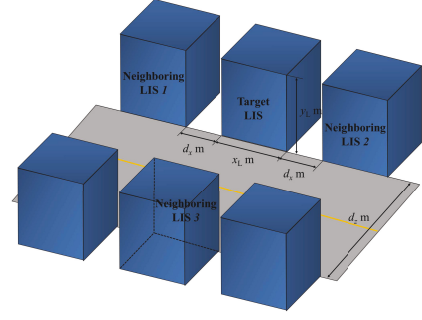


Fig. 3. Illustrative system model of the multiple LISs when $N = 4$.

IV. SIMULATION RESULTS AND ANALYSIS

We present Monte Carlo simulation results for the uplink SSE in an LIS system, and compare them with the results of the asymptotic analyses. All simulations are statistically averaged over a large number of independent runs. The simulation parameters are based on the LTE specifications, presented in Table I, and the probability of LOS path and Rician factor are applied based on the 3GPP model [3]. In the simulations, we consider both single- and multi-LIS cases. In both cases, we consider a scenario in which the devices are randomly and uniformly distributed within a three-dimensional space of $4 \text{ m} \times 4 \text{ m} \times 2 \text{ m}$ in front of each LIS. For the single-LIS case, $N = 1$ and only a single target LIS is located in two-dimensional Cartesian space along the xy -plane. For the multi-LIS case, to be able to consider the effect of pilot contamination, we assume a total of $N = 4$ LISs consist of one target LIS and three neighboring LISs, located on both sides and in front of the target LIS, as shown in Fig. 3. The parameters presented in Fig. 3 are such that $x_L = y_L = 4$, $d_x = 4$, and $d_z = 6$. All LISs are assumed to share the same frequency band, each of which serves K devices and reuses K pilot sequences.

In Figs. 4–6, Theorems 1 and 2 are verified in the following scenario. All intra-LIS interference channels are generated by device-specific spatially correlated Rician fading. In Fig. 4, all inter-LIS interference channels are also generated by that Rician fading, however, in Figs. 5 and 6, those channels are generated entirely from the NLOS path such as spatially correlated Rayleigh fading. Fig. 4 shows the channel hardening effect of an LIS system whereby the variances of individual SE in both single- and multi-LIS cases converge to zero as M increases, despite the pilot contamination effect. In both Figs. 4 and 5, the asymptotic results from Theorem 1 become close to the results of our simulations and these results gradually approach to their performance bounds obtained from Theorem 2, as M increases. Moreover, those performance bounds also converge to the limiting values resulting from the intra/inter-LIS interference through the LOS path. In Fig. 4, the performance gap between the results of the single- and multi-LIS is roughly 33 bps/Hz at $M = 900$, and it is expected to converge to 36 bps/Hz from the bound gap between the two systems, as M increases. This performance gap between the two systems stems from pilot contamination and inter-LIS interference generated from the LOS path,

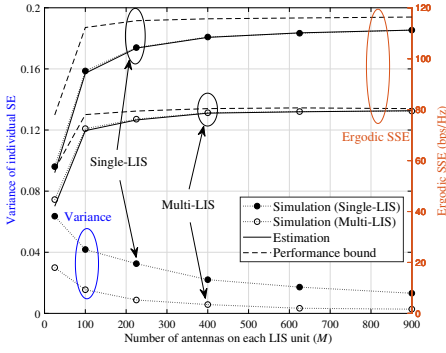


Fig. 4. Variance and ergodic SSE with Rician fading interference when $K = 20$.

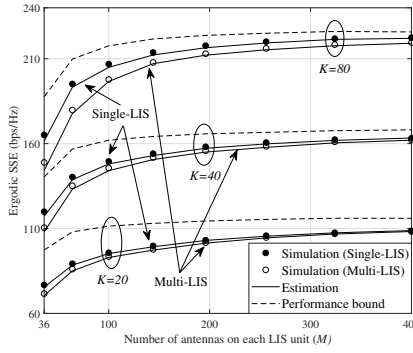


Fig. 5. Ergodic SSE with NLOS inter-LIS interference.

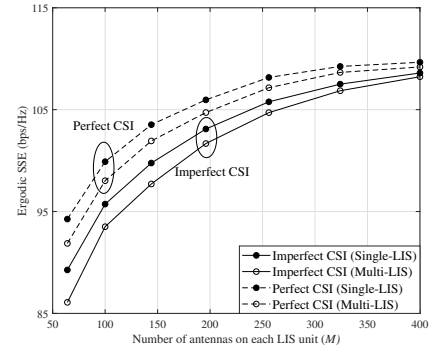


Fig. 6. Comparison of ergodic SSE when $K = 20$ with NLOS inter-LIS interference.

as proved in Theorem 2. In Fig. 5, the performance gap between the results of the single- and multi-LIS converges to zero even at $K = 80$, as M increases, and their bounds achieve an equal performance over the entire range of M . Since the pilot contamination and inter-LIS interference generated from the NLOS path become negligible compared to the intra-LIS interference through the LOS path, this results in the performance convergence between the two systems and eventually the multi-LIS system becomes an inter-LIS interference-free environment.

Fig. 6 compares the ergodic SSE resulting from cases with perfect CSI and imperfect CSI, when $K = 20$ and all inter-LIS interference channels are generated by spatially correlated Rayleigh fading. We can observe that all ergodic SSE converge to same value of roughly 110 bps/Hz. Hence, despite the pilot contamination in the multi-LIS case, the ergodic SSE of the multi-LIS system with the imperfect CSI converges to that with the perfect CSI, and it eventually reaches the single-LIS performance with perfect CSI, as M increases. This clearly shows a particular characteristic of LIS systems whereby pilot contamination and inter-LIS interference become negligible, representing a significant difference from massive MIMO.

V. CONCLUSION

In this paper, we have asymptotically analyzed the performance of an LIS system under practical LIS environments with a well-defined uplink frame structure and the pilot contamination. In particular, we have derived the asymptotic SSE by considering a practical LIS environment in which the interference channels are generated by device-specific spatially correlated Rician fading and channel estimation errors can be caused by pilot contamination based on a practical uplink frame structure. We have shown that the asymptotic results can accurately and analytically determine the performance of an LIS without the need for extensive simulations. Moreover, we have proved that the SSE of a multi-LIS system is bounded by three factors: pilot contamination, intra-LIS interference, and inter-LIS interference generated from the LOS path. On the other hand, the pilot contamination and intra/inter-LIS interference generated from the NLOS path and noise become negligible as M increases. Simulation results show that our analytical results are in close agreement with the results arising from extensive simulations. Our results also show that,

unlike conventional massive MIMO system, the effect of pilot contamination has been shown to become negligible when the inter-LIS interference is generated from NLOS path.

REFERENCES

- [1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *available online: arxiv.org/abs/1902.10265*, Feb. 2019.
- [2] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3949–3963, Jun. 2016.
- [3] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Performance analysis of large intelligent surfaces (LISs): Asymptotic data rate and channel hardening effects," *available online: arxiv.org/abs/1810.05667*, Oct. 2018.
- [4] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746–2758, May 2018.
- [5] S. Hu, K. Chitti, F. Rusek, and O. Edfors, "User assignment with distributed large intelligent surface (LIS) systems," in *IEEE PIMRC*, Sep. 2018.
- [6] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Reliability analysis of large intelligent surfaces (LISs): Rate distribution and outage probability," *available online: arxiv.org/abs/1903.11456*, Mar. 2019.
- [7] T. Kim, K. Min, M. Jung, and S. Choi, "Scaling laws of optimal training lengths for TDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7128–7142, Aug. 2018.
- [8] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [10] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Physical channels and modulation," TR 36.211, Dec. 2017, v15.0.0.
- [11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge Univ. Press, 2005.
- [12] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, Apr. 2017.
- [13] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [14] A. Khansefid and H. Minn, "On channel estimation for massive MIMO with pilot contamination," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1660–1663, Sep. 2015.
- [15] M. Jung, W. Saad, and G. Kong, "Performance analysis of large intelligent surfaces (LISs): Uplink spectral efficiency and pilot training," *available online: arxiv.org/abs/1904.00453*, Mar. 2019.
- [16] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Spatial channel model for Multiple Input Multiple Output (MIMO) simulations," TR 25.996, Mar. 2017, v14.0.0.