Finding the Shannon capacity of a wireless network is perhaps one of the most well-studied problem that has remained unsolved. For ad hoc wireless networks, two slightly relaxed capacity notions have been defined, transport [1] and transmission [2], for which theoretical results have been possible mainly because of two important simplifications: SINR model of communication, and assuming random locations for nodes [3].

Under the SINR model, communication between two nodes is deemed successful if the SINR between them is larger than a threshold that depends on the rate of transmission. The assumption of random location of nodes takes two forms, either nodes are assumed to be distributed uniformly on a unit-radius disk or nodes are assumed to be located on the whole of $\mathbb{R}^2$ as a homogenous Poisson point process (PPP) with a fixed density.

Even under these simplifications, as far as we know, there has been no fundamental characterization of the maximum throughput possible (under any reasonable capacity definition) in cellular wireless networks that are structured, rather than being ad hoc. Two important features of cellular wireless networks that may have been behind the lack of capacity results are: preclusion of large scale basestation (BS) coordination/scheduling (allowed in transport capacity) and universal implementation of automatic repeat request (ARQ) protocol, where packets are transmitted repeatedly until successful reception via acks/nacks. Retransmissions entail temporal correlations of SINRs, thereby making the analysis challenging.

In this paper, we consider the well-accepted model of a cellular wireless network, also called the tractable model [4], that has BS locations process distributed as a homogenous PPP, and the focus is on characterizing the ‘capacity’ in the downlink. The tractable model is a reasonable abstraction in the modern scenario, where multiple layers of BSs (macro, micro, femto) are overlaid over each other. Moreover, we consider the widely used BS-MU (mobile user) association rule, where each MU connects to its nearest BS, i.e., all MUs lying in a Voronoi cell connect to the representative Voronoi BS. The MU locations are also assumed to be distributed as a homogenous PPP, independent of the BS locations process.

We consider the SINR model of transmission for each BS-MU communication, where with ARQ, a packet is retransmitted from the BS until the SINR seen at the MU is above a certain threshold. Also, each BS serves all the MUs located in its Voronoi cell in a round-robin manner by dividing its slots/bandwidth equally among them to closely model the ‘fair’ practical implementation.

BSs in real-life cellular wireless networks are limited in their ability to coordinate their transmissions in order to control intercell interference. We begin by considering that each BS is only allowed to use local strategies, i.e., each BS’s transmission decisions are only based on local channel conditions (path-loss or fading gain) at the MU or some feedback (ack/nacks) from the MU it is serving. Note that with ARQ, the local strategies are allowed to be adaptive, for example, a BS can use power control, or choose whether to transmit at all, given the history of ack/nacks. We later extend our model to include small-scale BS coordination, where nearby basestations can schedule their transmissions together, similar to coordinated
With ARQ, let $D$ be the number of retransmissions needed for a packet from BS $x$ to be successful at its MU $y$, defined as

$$D = \min\{t : \text{SINR}_{xy}(t) \geq \beta\}.$$ 

$D$ has the interpretation of delay, the time (or retransmissions) needed to receive the packet successfully.

Let $\lambda$ be the BS density (per $m^2$) of the cellular network. With ARQ, to count for the actual rate of successful packet transmission, we consider a natural definition of capacity (in the downlink) to be the reciprocal of the expected delay. For analysis, we consider a typical MU, which without loss of generality is assumed to be located at the origin, that connects to its nearest BS. Let $n_0$ be the number of MUs in Voronoi cell that contains the origin where the typical-MU is located. For any local strategy $S$ used by BSs, the per-MU capacity is

$$C_m(S) = \frac{1}{\lambda_0 \mathbb{E}[D]} \text{ packets/sec},$$

the per-BS capacity with $S$ under round-robin policy is

$$C_b(S) = \frac{1}{\mathbb{E}[D]} \text{ packets/sec},$$

and the network wide capacity with $S$ is

$$C(S) = \frac{\lambda}{\mathbb{E}[D]} \text{ packets/sec/m}^2.$$ 

Hence, our capacity definition is

$$C = \sup_S C(S), \quad (1)$$

i.e., we are looking for the best possible local (adaptive) BS strategy $S$ that achieves the maximum network-wide throughput. Since we are assuming that each BS serves all the MUs connected to itself in a round-robin manner, $C$ is independent of the MU density. Note that even though we are assuming round robin scheduling, the performance analysis is for a typical user (e.g. randomly chosen), and not the worst case user, e.g., the cell-edge users. Thus, the round robin policy is not a limiting factor.

The SINR model of transmission for the BS-MU link can be thought of as multi-ary-non-symmetric erasure channel, that is an extension of a binary-erasure-channel (BEC) with feedback, where each BS has multiple choices of transmit powers, and where the probability of erasure depends on the choice of transmitted power. In comparison, in a BEC, there are only two possible transmission choices and the erasure probabilities (erasure probabilities) for both the choices are identical. As one may recall, the Shannon capacity for the BEC $C_{BEC}$ is achieved by a simple strategy of retransmitting the packet until it is correctly received, and $C_{BEC} = \frac{1}{\mathbb{E}[T]}$ [5], where $T$ is total number of retransmissions needed. Our capacity definition (1) is in similar spirit, where in addition, BS strategy can choose what power level to use after every erasure. One important distinction, however, compared to the usual BEC, is that the SINRs are temporally correlated, which in information theoretic language translates to the channel having memory.

### A. Prior work on Network Capacity

Two related metrics of capacity; transport and transmission, have been defined for ad hoc networks and for both, exact results have been obtained. The transport capacity framework [1] assumes that $\lambda$ nodes are distributed uniformly on a unit-radius disk, and $\lambda/2$ source-destination pairs among them are chosen randomly. The transport capacity is the measure of how many bit-meters can be simultaneously transported across the network, where one bit-meter is transported if one bit is successfully (SINR model) transmitted to a distance of one meter towards its destination. Under the path-loss only model, where fading gain is neglected, transport capacity has been shown to be $\Theta(\sqrt{\lambda})$ [1], [6]. Some extensions of transport capacity are also known, where with fast mobile node mobility it is known to scale as $\Theta(\lambda)$, and also in the information theoretic communication model, it scales as $\Theta(\lambda)$ [7].

Transmission capacity framework [2] assumes the other type of randomness, where (transmitter) node locations are distributed as a homogenous PPP with density $\lambda$. Each transmitter has a receiver at a fixed distance $d_f$ from it in a random direction. For each $\epsilon$ (reliability constraint), transmission capacity is defined as the largest density $\lambda$, such that the outage probability $P(\text{SINR} < \beta)$ for each transmission is less than $\epsilon$.

Let $\lambda$, be the transmission capacity with reliability constraint of $1 - \epsilon$. The distance scaled transmission capacity, where $\epsilon$ is multiplied with the fixed transmitter-receiver distance $d_f$, is comparable with transport capacity, and remarkably, also scales as $\Theta(\sqrt{\lambda})$ [2], [8], for both the path-loss only and the path-loss plus fading model.

Similar to Shannon capacity, transport capacity has in-built reliability (modulo the SINR model definition), i.e., all counted bits are actually successfully received. Transmission capacity on the other hand follows a fire and forget principle (one-shot transmission), where a packet is guaranteed to be successfully received with probability $1 - \epsilon$, but if unsuccessful, there is no penalty or a retransmission procedure. Transport capacity definition, however, allows for global arbitrary coordination (scheduling of transmission from all possible transmitters).

With retransmissions, a generalization of the transmission capacity, called the delay normalized transmission capacity $C_R(S)$ with BS strategy $S$ [9], is defined as the number of successfully delivered packets in the network under the SINR model subject to a maximum limit on the number of retransmissions $R$. Let $N_R(t)$ be the number of successfully received packets at the MU until time $t$ with at most $R$ retransmissions from the corresponding BS. Then,

$$C_R(S) = \lambda \lim_{t \rightarrow \infty} \frac{N_R(t)}{t}.$$ 

From the renewal reward Theorem [10], $C_R$ is also given by

$$C_R(S) = \frac{\lambda P_R^S}{\mathbb{E}[D(R)]}, \quad (2)$$

where $P_R^S$ is the probability that a packet is successfully received at the MU, and $D(R)$ is the random variable denoting...
the actual number of retransmissions needed, with at most R retransmissions. \( D(R) \) has the interpretation of delay, the time (or retransmissions) needed to receive the packet successfully.

Our capacity definition (1) is essentially the limit of \( C_R \) as \( R \to \infty \), where \( P_s^R \to 1 \), and yields

\[
C(S) = \lim_{R \to \infty} \frac{P_s^R}{E[D(R)]} = \frac{\lambda}{E[D]},
\]

where \( D \) is now the absolute delay between any BS \( x \) and MU \( y \) it is serving with no retransmission limit defined as

\[
D = \min\{t : \text{SINR}_{xy}(t) \geq \beta\}.
\]

Remark 1. Using the limit of (2) as \( R \to \infty \) for our capacity definition is actually not a simplification compared to (2), since each BS serves all its MUs in a round-robin fashion, thereby precluding selection of which MU to serve in each slot. Thus, for any particular MU, with or without constraining the maximum number of retransmissions, there is no change in the average number of successfully received packets, since the underlying random process that determines success/failure of packets does not depend on the packet index. Thus, dropping a packet after hitting the constraint \( R \) or continuing with it, yields the same average rate of successful packet transmissions (1/expected delay).

The delay normalized transmission capacity was first defined in [9], and later studied in [11]. The basic problem in studying the delay normalized transmission capacity is the complicated correlation of SINRs across time slots under the PPP assumption on BS locations [12]. For simplifying analysis, [9] made a limiting assumption that SINRs across time slots are independent. A more rigorous approach was taken in [11], but yielded limited analytical results because of the mentioned difficulty and the main results were derived only for very-low-BS density regime.

B. Our results

In this paper, we avoid making any simplifying/limiting assumptions for studying the joint distribution of SINRs across time slots for deriving bounds on the expected delay seen at any MU. The main result of our paper is that, when each BS is only allowed a local strategy or BSs can accomplish only small-scale coordination, upto constants, \(^1\)

\[
\min_{S} E\{D\} = \max \left\{ \frac{1}{\lambda^2}, \exp(\lambda) \right\},
\]

and consequently the capacity is

\[
C = \min \left\{ \lambda^{\frac{\alpha}{2} + 1}, \lambda \exp(-\lambda) \right\}.
\]

In Fig. 1, we illustrate the behaviour of the capacity (upto a constant) of the cellular wireless network as a function of the BS density as the blue curve, that tracks the minimum of the two functions \( \lambda^{\frac{\alpha}{2} + 1} \) and \( \lambda \exp(-\lambda) \).

\(^1\)The lower and upper bounds only differ by constants, and all \( c_i \)'s are constants.

Thus, the capacity of a cellular wireless network increases polynomially in the low-BS density regime, then increases as \( \lambda \exp(-\lambda) \) in the moderate BS density regime, and eventually starts to decreases exponentially with the density of basestations for local BS strategies or with small-scale BS coordination.

The main result is derived via the following sub-results.

- We begin by showing that for any local BS strategy \( S \)

\[
C(S) = \frac{\lambda}{E[D]} \leq \min \left\{ \frac{\lambda^{\frac{\alpha}{2} + 1}}{c_1}, c_3 \lambda \exp(-c_2 \lambda) \right\}.
\]

It is accomplished via deriving two separate bounds, first

\[
E\{D\} \geq \left( \frac{c_1}{\lambda^{\frac{\alpha}{2}}} \right),
\]

that is relevant for low-BS density regime, and then

\[
E\{D\} \geq c_4 \left( \exp(c_2 \lambda) \right),
\]

that is tighter in the high-BS density regime.
- The upper bound on the capacity is shown to hold even if small scale BS coordination is allowed, where a fixed number (independent of the BS density) of BSs can schedule their transmissions jointly.
- We show that a simple non-adaptive strategy \( S \) (first proposed by us in [13]), where each BS transmits power to completely nullify the path-loss based signal attenuation at the MU it is serving (using the knowledge of the distance to the MU), achieves

\[
E\{D\} = O \left( \sqrt{1 + \frac{\Gamma(\alpha + 1)}{(\pi \lambda)^{\alpha}}} \exp(c_5 \lambda) \right),
\]

\[
\Rightarrow C(S) \geq \left( \min \left( c_6 \lambda^{\frac{\alpha}{2} + 1}, \frac{\lambda \exp(-c_6 \lambda)}{c_7} \right) \right),
\]

matching the upper bound on the capacity for any strategy \( S \). Importantly, a single policy achieves the capacity upper bound for the both the low and the high-BS density regime.
regimes, for which the capacity scaling behaviour is quite different.

- We show that for a pure ALOHA strategy $S$ that is commonly employed and most often studied, $C(S) = 0$. Moreover, any ALOHA type strategy, where the serving BS has the knowledge of the distance $d$ to its MU, can achieve non-zero capacity, only if its transmit power scales at least as $d^{-2}$, where $\alpha > 2$ is the path-loss exponent. Since $d$ is a random variable with infinite support, this result directly implies that if each BS has a strict peak power constraint, then its capacity is 0.

Our result shows that in the low-BS density regime the capacity increases polynomially with the BS density, since with low BS density, interference is weak and signal power can be boosted using the proposed power control strategy. In the high-BS density regime, our result shows that the ‘realistic’ capacity of a cellular wireless network is fundamentally different and significantly smaller than previously thought via the known results on the transport or the transmission capacity (both scale as $\Theta(\sqrt{\lambda})$) in the high-BS density regime.

The early indication for this distinction was visible in [14], that considered an ad hoc network where nodes can be either source/destination/relay in any time slot, rather than taking up a rigid/fixed BS or MU role as in our case. Assuming the location of all nodes to be a PPP, under the SINR model, [14] showed that the ALOHA strategy has infinite expected delay for a packet to leave its source (to be received successfully at any node in the network, in particular the nearest node), and consequently has zero effective rate of communication.

What our upper bound result on the capacity shows that even if each BS is allowed adaptive strategies using local information depending on distance, fading gain or the SINR seen at the MU, the delay grows at least exponentially in the density of the basestations in the high-BS density regime.

The proposed strategy to achieve the upper bound only depends on the distance between the BS and the MU, and does not require channel state information (CSI) that changes much faster than the distance. In particular, if the BS-MU distance is $d$, then the strategy transmits with power $\ell(d)^{-1}$ with probability $M/(d)^{-1}$ in any slot dedicated for the MU, where $\ell$ is the average power constraint. Note that the strategy is not adaptive, making it suitable for practical implementation with low complexity. The basic idea behind the strategy is that it is to transmit infrequently, but whenever an attempt is made, sufficient power is used to compensate the path-loss completely. Thus, this strategy limits the overall network interference, while keeping the signal power high enough whenever an attempt is made, and more importantly achieves the capacity upper bound in both the low as well as high BS density regimes, where the capacity scaling behaviour is very different.

One important design implication of our result is in terms of cell densification, where number of basestations are increased to decrease the individual cell-sizes in the hope of improving connectivity and communication rate. Single-shot performance metrics such as connection probability or average rate have been considered in past for analyzing the effects of cell densification [15]–[19], that typically conclude that there is a phase transition; initially performance improves as BS density increases, then it saturates, and eventually it starts to degrade, but explicit dependence was not identified. Also, it is believed that cell densification can improve the spectral efficiency via frequency reuse. Taking a more comprehensive view via including retransmissions, where the effects of temporal correlations of SINRs are fully incorporated, we characterize the exact effect of cell densification on the capacity/long-term throughput.

We show that in the low-BS density regime, cell densification increases the capacity polynomially, while in the high-BS density regime, BS densification leads to an exponential fall in the capacity. Thus, cell densification needs to be undertaken judiciously.

C. Comparison with Prior Work

To compare our capacity metric with transport capacity or transmission capacity (relevant for high BS density), one has to multiply the average distance between the MU and its nearest BS distance $d_0$ (that scales as $\Theta\left(\frac{1}{\lambda}\right)$) with the capacity, which again yields $\mathcal{C}_{i0} = \Theta\left(\exp\left(-\lambda\right)\right)$. The reasons for arriving at such fundamentally different scaling compared to the transport capacity and the transmission capacity, that both scale as $\Theta\left(\sqrt{\lambda}\right)$, are that transport capacity has stricter reliability requirement (SINR has to be $\geq \beta$ for each successful counted bit on realization basis) but allows for large scale BS/node scheduling, while transmission capacity has one-shot reliability constraint of $1 - \epsilon$, and does not allow for retransmissions. In cellular networks, large scale BS scheduling is not possible, while retransmissions are an integral part of any deployment. The inability of large scale BS scheduling ensures that interference seen at any MU increases with BS density, while with retransmissions, the temporal correlation of SINRs degrades the delay performance.

The positive time-correlation of SINRs with PPP distributed nodes that was first derived in [12], is the principal reason behind the exponential increase of the expected delay in the high BS density regime. The high positive time correlation in the high-BS density regime ensures that if a packet has not been received till a sufficient number of retransmissions, then increasingly it become less and less likely to succeed, and leads to expected delay being exponential in the BS density. We would like to note that the analysis in this paper is non-asymptotic and holds for all $\lambda$, in contrast to the achievability of transport capacity [6], that needed the number of nodes to go to infinity for the application of results from percolation theory.

It is worth noting that the analytical tools used for analysis of transport capacity [1], [6] are geometric, and percolation theory based, while in this paper, they are more probabilistic for estimating the joint distribution of the temporal SINRs.

An important result obtained in [4], under the exact same model considered in this paper (BS and MU location processes being independent PPPs, and each MU connects to its nearest BS), is that ignoring noise, the one-shot connection probability
Our results on the capacity of cellular wireless networks.

D. Tradeoffs

We discuss some of the tradeoffs involved in the context of our results on the capacity of cellular wireless networks.

- One integral feature used for improving the performance of a cellular networks has been power control, that is used extensively in CDMA networks, and also finds applicability in other deployments to control the out-of-cell interference. Under the SINR model considered in the paper, all the interference seen at any MU is coming from out of cell. Each BS tries to serve its MU as well as possible, essentially by transmitting with sufficient power which is in conflict with the rate achievable by MUs in other cells. Thus, the choice of transmit power is critical for maximizing the network-wide capacity, where the two conflicting objectives (maintaining sufficient SINR for own users, while minimizing interference for MUs in other cells) have to be balanced out by BSs. With increasing BS density, the number of BSs that are within a fixed distance (that really matter for SINR) from a MU increases linearly, thereby increasing interference rapidly. To counter this increased interference, the serving BS has to increase its power, but that ends up causing increased interference at out of cell MUs.

- Another dimension to the problem is the use of ARQ in cellular networks. With strong positive temporal correlations of SINR, the success probability progressively decreases with the retransmission index. Thus, if a packet has not succeeded in few retransmissions, it is likely to take a large number of retransmissions (heavy-tail behavior). The power control issue together with the temporal correlations of SINR, limits the performance of a cellular wireless network, and the expected delay increases at least exponentially with the BS density, leading to capacity falling exponentially with BS density.

- Path-loss exponent: The path-loss exponent $\alpha$ is an important feature governing the performance in a cellular network. Larger value of $\alpha$ leads to lower interference, but also leads to lower signal power. It is also implicitly connected with the BS density $\lambda$, since with large $\lambda$, the distance between the MU and the BS it connects to decreases, thereby increasing the signal power, but at the same time there are larger number of closeby interferers, even though each interferer contributes lower interference with larger $\alpha$. Thus, how does $\alpha$ impacts the capacity is an important question. We show that the capacity (both upper and lower bound) is indifferent to the value of $\alpha$ for $\lambda > 1$, while for low-BS density regime $\lambda < 1$, we show that the capacity decreases with $\alpha$ as $\lambda^{1+\alpha}/2$. Thus, the effect of $\alpha$ is only visible at low-BS densities, where with larger inter-BS distances, interference is weak, and larger $\alpha$ decreases the signal power.

- MU density: We have implicitly assumed that the MU density is much higher than the BS density for our capacity definition, which is typically the case in a well-designed system. In case, the MU density is comparable to the BS density, it will lead to some BSs not having any MUs in their Voronoi regions, and consequently becoming inactive. Since the BS and the MU processes are independent, this will lead to an independent thinning of the active BS process (a new PPP with less density). Thus, our results are again applicable in this case.

E. Limitations

By far, as one can argue, the biggest limitation of our work is the non-information-theoretic definition of capacity. However, as is well known, the Shannon capacity of networks with even 2/3/4 nodes, e.g., interference, broadcast channels, remains unknown. Thus, the hope of finding Shannon capacity of general wireless networks is rather slim. To better understand the fundamental tradeoffs and limitations in networks, starting with [1], many simpler but functional definitions of capacity [2], [20] have been proposed and analysed. Our definition is a step in the same direction, with primary focus on cellular networks that are structured, and operate under some well established protocols such as ARQ. Even under this simpler definition, the capacity analysis remains non-trivial, and we are able to derive the exact scaling of capacity with the BS density.

Another possible limitation of our work can be argued in terms of assuming independent point processes for the BS and the MU locations, which has become very popular starting with [4]. One might expect BSs to be deployed in areas where a large number of users tend to be present, thus coupling the BS and MU locations process as proposed in the cluster model of [21]. We expect our results to be applicable even in this model, since the nearest neighbour distances still have similar distributions, and order-wise results should be similar. The exact results are, however, out of scope of this present paper.

II. Notation

We follow the following order notation in this paper. $f(x) = O(g(x))$ if and only if there exists a positive real number $M$ and a real number $x_0$ such that $|f(x)| \leq M|g(x)|$ for all $x \geq x_0$. 

\[ P_c = P(SINR > \beta) \] for any one slot is independent of the BS density $\lambda$. This result pointed towards showing that with increase in $\lambda$, the increase in the signal power because of the decreasing nearest BS distance completely compensates for the increase in the interference, and cell densification can lead to linear increase in network capacity with BS density. This conclusion (if valid) will also point towards potentially showing that expected delay $E[D]$ is independent of $\lambda$, at least when noise is ignored, contradicting our results. The real reason for $P_c$ being independent of $\lambda$ under the no-noise condition [4], is actually the use of an amplifying path-loss model of $\ell(x) = x^{-\alpha}$ in [4], that produces unrealistic signal amplification for distances $x < 1$. If instead one uses a more realistic path-loss function, e.g., $\ell(x) = \min\{1, x^{-\alpha}\}$ or $\ell(x) = 1/(1 + x^\alpha)$ that do not have signal amplification, such a result will not be possible. Thus, in this paper, we consider path-loss function $\ell(x) = \min\{1, x^{-\alpha}\}$, but all our results will apply for any non-amplifying path-loss function, such as $\ell(x) = 1/(1 + x^\alpha)$.

\[ \lambda \] is independent of $x$.

We follow the following order notation in this paper.
with respect to basestation locations, all MUs in a Voronoi cell/region connect to its representative basestation. See Fig. 2. Cellular wireless network where basestations are denoted by circles, and whole area is divided in Voronoi regions, and each mobile (square) node connects to its nearest basestation.

Similarly, \( f(n) = o(g(n)) \) if and only if for each \( \epsilon > 0 \), there exists a constant \( N \) such that \( |f(n)| \leq \epsilon |g(n)| \) for all \( n \geq N \). Moreover, \( f(x) = \Theta(g(x)) \) if only if \( f(x) = O(g(x)) \) and \( g(x) = O(f(x)) \). For two random variables \( X \) and \( Y \), \( X \) is defined to be stochastically dominated by \( Y \) if \( \P(X \leq x) \geq \P(Y \leq x) \) for all \( x \).

III. SYSTEM MODEL

We consider a cellular network model, that consists of basestations \( \{T_n\} \), whose locations are distributed according to a homogeneous PPP \( \Phi = \{T_n\} \) with density \( \lambda \), popularly called the tractable model starting from [4]. This model is reasonable since with a PPP, given the number of nodes lying in the given area, the node locations are uniformly distributed similar to what is seen practically with many different types of basestations (macro, micro, femto, etc.) overlaid on top of each other.

The MUs are also located according to an independent PPP \( \Phi_B \) with density \( \mu \), and each MU connects to its nearest basestation, which is what is typically the case in a practical deployment. Therefore, by the definition of Voronoi regions with respect to basestation locations, all MUs in a Voronoi cell/region connect to its representative basestation. See Fig. 2 for the considered network schematic.

We consider a typical MU \( m \), and to study the capacity \( C \), focus on the expected delay encountered by its packets transmitted from its nearest basestation \( n(m) \).

**Transmission model:** We consider that time is slotted, and each BS transmits to all the MUs that lie in its Voronoi region/cell in a round-robin manner, i.e., equally sharing time slots between them.

**Path-loss model:** We consider the distance based path-loss (signal attenuation) function to be

\[
\ell(d) = \min\{1, d^{-\alpha}\}, \quad \text{for } \alpha > 2.
\]

A simple function \( \ell(d) = d^{-\alpha} \) is used widely in literature, however, for \( d < 1 \), it produces signal amplification which is unrealizable. The results presented in this paper are valid for most other reasonable path-loss functions such as \( \ell(d) = \frac{1}{1+d^\alpha} \) that do not have signal amplification.

**Fading model:** We assume that each node in the network is equipped with a single antenna, and at time \( t \), the fading gain between BS \( x \) and MU \( y \) is denoted by \( h_{x,y}(t) \) that is assumed to be exponentially distributed with parameter 1. Moreover, \( h_{x,y}(t) \) is assumed to be independent and identically distributed for all time \( t \), and all location pairs \( x,y \).

**Power Constraint:** We assume that each BS has an average power constraint of \( M \).

Without loss of generality, we assume that the typical MU \( m \) is located at the origin, and the distance to the nearest basestation from origin is \( d_0 \), while any other BS \( z \) is located at distance \( z \) (abuse of notation) from the origin. Since we focus on only the MU \( m \) located at the origin, we abbreviate \( h_{z,0}(t) \) to \( h_z(t) \). For BS \( z \), let \( P_z(t) \) be the transmit power at time \( t \), and \( 1_z(t) \) be the indicator variable denoting whether BS \( z \) is transmitting at time \( t \) or not, with \( \P(1_z(t) = 1) = p_z(t) \).

The received signal at \( m \) located at the origin that is connected to its nearest basestation \( n(m) \) is given by

\[
y(t) = \sqrt{P_{n(m)}(t)\ell(d_0)h_{n(m)}(t)1_{n(m)}(t)s(t)} + \sum_{z \in \Phi \setminus \{n(m)\}} \sqrt{P_z(t)\ell(z)h_z(t)1_z(t)}s_z(t) + w, \quad (3)
\]

where \( s(t) \) is the signal transmitted from BS \( z \) with power \( P_z(t) \) at time \( t \), \( w \) is the AWGN with variance \( N \), and \( 0 < \gamma \leq 1 \) is the processing gain of the system or the interference suppression parameter which depends on the transmission/detection strategy, for example, on the orthogonality between codes used by different legitimate nodes during simultaneous transmissions.

Thus, the SINR at \( m \) from its nearest BS \( n(m) \) in time slot \( t \) is given by

\[
\text{SINR}_{n(m),m}(t) = \frac{P_{n(m)}(t)h_{n(m)}(t)\ell(d_0)1_{n(m)}(t)}{\gamma \sum_{z \in \Phi \setminus \{n(m)\}} P_z(t)1_z(t)h_z(t)\ell(z) + N}. \quad (4)
\]

The transmission from BS \( x \) to MU \( y \) is deemed successful at time \( t \), if \( \text{SINR}_{xy}(t) > \beta \), where \( \beta > 0 \) is a fixed threshold depending on the rate of information transfer. Let

\[
e_{xy}(t) = \begin{cases} 1 & \text{if } \text{SINR}_{xy}(t) > \beta, \\ 0 & \text{otherwise}. \end{cases} \quad (5)
\]

Since \( h_{ij}(t) \) is a random variable, multiple transmissions may be required for a packet to be successfully received at any node.
Thus, a measure of delay, i.e., the number of retransmissions needed to successfully receive packets, is required, which is defined as follows.

**Definition 1.** Let the minimum time (delay) taken by any packet to be successfully received at any BS be

\[ D = \min \{ t > 0 : e_n(t) = 1 \}. \]

Thus, conditioned on \( D \), the number of retransmissions needed to successfully receive packets, is required, which is defined as follows.

**Definition 2.** As stated earlier, we consider the capacity to be

\[ C = \sup_S \frac{\lambda}{E[D]}, \]

where \( S \) is any BS strategy followed by all BSs.

Thus, we are fixing BS density \( \lambda \) and SINR threshold \( \beta \), and looking for the best possible strategy that achieves the capacity. Next, we derive upper bounds on this capacity definition.

**IV. UPPER BOUND ON CAPACITY**

**A. ALOHA type strategies**

We begin by considering the most common ALOHA type strategies, where each BS transmits with probability \( p \) i.i.d. in each slot with power \( P \), such that \( pP = M \) to satisfy the average power constraint of \( M \). For finding a lower bound on the expected delay, we consider the scenario where other than BS \( n(m) \), there is no other active BS in the network. Thus, the interference seen at \( m \) is zero, and we only consider additive noise, in which case SINR = SNR. Clearly, this will result in lower expected delay.

Without any interference, there are only two sources of randomness, i) distance \( d_0 \) between BS \( n(m) \) and \( m \) that is fixed for all time slots given a realization of \( \Phi \) and ii) fading variables \( h_{(m)}(t) \) that are i.i.d. for each time slot \( t \). Note that \( p \) and \( P \) can be arbitrary, however, with ALOHA strategy, \( p \) has to be constant across time slots, so it cannot depend on \( h_{(m)}(t) \), but can depend on \( d_0 \).

We first show that if the transmit power \( P \) does not scale as fast as \( d_0^{-2} \), then the expected delay is infinity.

**Theorem 1.** For any ALOHA type strategy, if \( P = o(d_0^{\alpha-2}) \),

\[ E[D] = \infty, \]

and consequently

\[ C = 0. \]

**Proof.** To prove this result, consider that other than BS \( n(m) \) there is no other active BS in the network. Thus, the interference seen at \( m \) is zero. Let the BS \( n(m) \) use an ALOHA type policy where it transmits in each slot with probability \( p \) i.i.d. with power \( P \), such that \( pP = M \) to satisfy the average power constraint.

Given a realization of the BS PPP \( \Phi \), the distance \( d_0 \) is fixed. Thus, conditioned on \( d_0 \), the event of transmission \( 1_{n(m)} = 1 \), and the success event \( e_{n(m),m}(t) = 1 \) (\( h_{(m)}(t) > \beta^{\frac{1}{1-\epsilon}}(d_0) \)) are independent. Moreover, \( e_{n(m),m}(t) \) is also independent across \( t \) because of i.i.d. assumption on \( h_{(m)}(t) \). Recall that with independent trials and per-trial success probability of \( q \), the expected time till the earliest success is \( 1/q \). Hence

\[ E[D|\Phi] = \frac{1}{pP\left(\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right)}, \]

\[ = \frac{1}{p \exp\left(-\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right)}, \]

since \( h_{(m)}(t) \sim EXP(1) \). Since \( pP \leq M \), we have

\[ E[D|\Phi] \geq \frac{P \exp\left(-\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right)}{M}. \]

Note that \( \ell(d_0) \leq d_0^{-\alpha} \). Thus,

\[ E[D|\Phi] \geq \frac{P \exp\left(-\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right)}{M}. \]

Hence

\[ E[D] = E[E[D|\Phi]] \geq \frac{P \exp\left(-\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right)}{M}. \]

From Proposition 5, we know that \( f_{d_0}(y) = 2\pi y \exp(-\lambda\pi y^2) \). Thus,

\[ E[D] \geq \frac{1}{M} \int_0^{\infty} 2\pi y \exp(-\lambda\pi y^2) \exp\left(-\frac{\beta^{\frac{1}{1-\epsilon}}(d_0)}{P}\right) dy, \]

\[ = \infty, \]

if \( P = o(d_0^{\alpha-2}) \) when \( \alpha > 2 \).

Following important corollaries are immediate.

**Corollary 3.** With the pure ALOHA strategy, where the BS transmits independently of the distance \( d_0 \) to the MU it is transmitting to,

\[ E[D] = \infty. \]

**Corollary 4.** If there is a peak power constraint at each BS, then again with the ALOHA strategy,

\[ E[D] = \infty. \]

**Proof.** Note \( d_0 \) can be arbitrarily large, but a peak power constraint will always have \( P = o(d_0^{\alpha-2}) \), since \( \alpha > 2 \). Hence

\[ E[D] = \infty. \]

Corollary 3 shows that the pure ALOHA strategy that assumes knowledge of distance \( d_0 \), has infinite delay and leads to zero capacity. This is similar to what was shown in [14] for a single ad hoc network, that the expected delay for any packet leaving its source successfully is infinite using an ALOHA strategy. Even when an ALOHA strategy can modulate its power depending on distance \( d_0 \), this result shows that the power has to scale at least as much as \( d_0^{-2} \) to get non-zero capacity. Corollary 4 implies that any BS can only serve MUs that lie within a bounded area in its Voronoi cell under the peak-power constraint using the ALOHA strategy.
Proposition 5. The cumulative distribution function and probability distribution function of nearest BS distance $d_0$ is
\[
\mathbb{P}(d_0 > y) = \exp(-\lambda \pi y^2)
\]
and
\[
f_{d_0}(y) = 2\lambda \pi y \exp(-\lambda \pi y^2).
\]

We next consider more general (potentially adaptive) strategies, and derive lower bounds on their expected delays.

B. General Strategies

Definition 6. Let $\mathcal{S} = \{(p_i, P_i), i \in \mathcal{I} : p_i P_i \in [M/\tau, M]\}$, where $\tau \geq 1$ is a constant, be a collection of probability of transmission $p_i$ and power transmission $P_i$ pairs. A strategy for BS is to choose any element of $\mathcal{S}$ at any given time slot, where the current choice can be adaptive, i.e., it can depend on entire history of earlier choices. 2 Note that with our definition, the average power constraint of $M$ is satisfied automatically.

In this paper, for being as close to practical wireless networks, we begin by restricting ourselves to local strategies that are defined as follows. In the sequel, we consider small-scale BS coordination strategies as well.

Definition 7. A transmission strategy adopted by a BS is called local, if it only depends on either the fading gain or the distance between itself and the connected MU $m$, or the history of success/failure event $e_{n(m), m}$ at $m$ i.e., ack/nack signal sent back by the $m$. In particular, BS $n(m)$, can choose $p_{n(m)}(t)$, $P_{n(m)}(t)$ pair for time slot $t$, depending on $h_{n(m)}(t)$, $d_0$, and history of $e_{n(m), m}(s)$, and $p_{n(m)}(s)$, $P_{n(m)}(s)$ for $1 \leq s < t$.

ALOHA is one simple example of a local strategy. Essentially local strategies preclude global coordination or scheduling across BSs that is very costly and seldom employed in practice in cellular networks. Importantly, local strategies allow adaptation, e.g., power control, that is commonly implemented in practice.

Next, we present the first main result of the paper that upper bounds the capacity of any cellular wireless network, when each BS follows a local strategy.

Theorem 2. For any local strategy $\mathcal{S}$ employed by a BS, the expected delay at the typical MU $m$ satisfies
\[
\mathbb{E}\{D\} \geq \left(\frac{c_1}{\lambda^2}\right),
\]
as well as
\[
\mathbb{E}\{D\} \geq c_4 (\exp(c_2 \lambda)).
\]
Consequently, the network wide capacity is
\[
C = \frac{\lambda}{\mathbb{E}\{D\}} \leq \min \left\{\frac{\lambda^{\frac{\alpha+1}{c_1}}}{c_3 \lambda \exp(-c_2 \lambda)}\right\}.
\]

2This strategy can closely emulate any transmission policy used by BS under an average power constraint $M$. The lower bound on $p_i, P_i \geq M/\tau$ reflects the considered setting of BS density being smaller than the MU density, where each BS has at least one MU in its Voronoi cell.

Theorem 2 shows that for low-BS densities, the expected delay decreases polynomially with the BS density, where (10) dominates. In the low-BS density regime, interference is weak, and increasing BS density results in decreased BS-MU distances, boosting the signal power. For moderate and high BS densities, (11) dominates, and our result shows that the expected delay grows at least exponentially with the BS density.

For large BS densities, Theorem 2 is essentially a negative result that shows that even when each BS has all the local information, that can be used adaptively, the expected delay increases exponentially with the density of BSs. Consequently, the network wide capacity decreases exponentially with the increase in the density of BSs in the high-BS density regime.

Theorem 2 also suggests that both transport and transmission capacity definitions (both scale as $\Theta(\sqrt{\lambda})$) overestimate the practically realizable capacity of a cellular wireless network in the high-BS density regime. The reason for this is that even though with transport capacity, the reliability condition is strict, but large scale BS coordination or scheduling is allowed, while with transmission capacity, the reliability constraint is loose and communication is one-shot.

Proof. First, we derive the upper bound that is tighter for lower values of BS densities (10). Consider that other than BS $n(m)$ there is no other active BS in the network. Thus, the interference seen at $m$ is zero. Then the SINR seen at $m$ in any time slot $t$
\[
\text{SINR}_{n(m), m}(t) = \frac{P_{n(m)} h_{n(m)}(t) \mathbb{1}_{n(m)}(t) \xi(d_0)}{N},
\]
The lower bound (10) is derived in Appendix A using the following proposition.

Proposition 8. [22] Under average power constraint of $M$, to minimize the outage probability $\mathbb{P}(h_i P_i \leq \gamma \beta)$, where $h_i$ is i.i.d. ~ $\text{EXP}(1)$, the optimal power allocation
\[
P_i = \begin{cases} 
\frac{(\gamma \beta)}{h_i} & h_i \geq \delta, \\
0 & h_i < \delta,
\end{cases}
\]
where $\delta$ is chosen to satisfy the average power constraint, and where lower average power constraint implies larger $\delta$, and vice versa.

The basic idea used to prove (10) is to first condition on $\Phi$, and then bound the expected delay when the BS can use the optimal power transmission strategy (Proposition 8) depending on the fading gains $h_{n(m)}$.

Now we consider a more general upper bound while incorporating the interference that will be tight for intermediate and higher values of BS densities. As defined before, the typical BS $n(m)$ follows a local strategy, i.e., it transmits with power $P_{n(m)}(t)$ with probability $p_{n(m)}(t)$ in any time slot $t$, such that $p_{n(m)}(t) \mathbb{1}_{n(m)}(t) \in [M/\tau, M]$, where $\tau \geq 1$ is a constant, where $p_{n(m)}(t), P_{n(m)}(t)$ can depend on $d_0$, $h_{n(m)}(t)$, and history $e_{n(m), m}(s), p_{n(m)}(s), P_{n(m)}(s)$, $1 \leq s < t$. 
To obtain a lower bound on the expected delay, we set $\ell(d_{\ell}) = 1$ for the typical user $m$, to maximize the signal power in terms of path-loss, since $\ell(.) \leq 1$. This will allow us to remove the dependence of $(p_{n(m)}(t), P_{n(m)}(t))$ on $d_0$. Note that we are not putting any restriction on $d_0$, since that would impact the interference term in the SINR expression.

Let $U_z$ be the set of MUs connected to the BS $z$. As before, we consider the practical setting where the MU density is much larger than the BS density, and $|U_z| \geq 1$ for all BSs $z$. For BSs other than $n(m)$, we let all the MUs $u \in U_z$ connected to BS $z$ be at most a unit distance away from $z$, i.e., $d_{zu} \leq 1$, $\forall z, u \in U_z$, which implies $\ell(d_{zu}) = 1$. This minimizes the path-loss between $z$ and $u \in U_z$, and gives the largest gain $\ell(d_{zu}) = 1$ for each BS-MU pair $(z, u)$ for $u \in U_z$. Moreover, we also let that each $u \in U_z$ receives no interference from any basestation. Let this be called restriction 1 that applies to all non-typical BSs $z \neq n(m)$ and its connected users in $U_z$. Thus, the SINR seen at any MU $u \in U_z$ under restriction 1 is

$$\text{SINR}_{z,u}(t) = \frac{P_z(t)h_{zu}(t)\mathbb{1}_z(t)}{N}. \quad (13)$$

As a function of time, let $\hat{p}_z(t)$ and $\hat{P}_z(t)$ (where $\hat{p}_z(t)\hat{P}_z(t)$ $\in [M/\tau, M]$) from strategy Definition 6), be the power profile used by BS $z$ under restriction 1 (SINR given by (13)) to get the same delay/capacity without restriction 1 while using the optimal (unknown) power profile $p_z(t)$ and $P_z(t)$. Clearly, since there is no path-loss and no interference seen at any $u \in U_z$ with restriction 1 (13), power profile $\hat{p}_z(t)$ and $\hat{P}_z(t)$ is stochastically dominated by $p_z(t)$ and $P_z(t)$. Equivalently, the interference seen at the typical MU $m$ from BS $z \neq n(m)$ is stochastically dominated with restriction 1.

Recall that the SINR (4) seen at the typical user $m$ critically depends on the power profile $(p_z(t)$ and $P_z(t))$ of BS $z \in \Phi\setminus(n(m))$. Thus, using the stochastically dominated power profile $\hat{p}_z(t)$ and $\hat{P}_z(t)$ with restriction 1 for each BS $z \neq n(m)$ cannot increase the delay at $m$. Thus, we work under restriction 1 to lower bound the expected delay seen at $m$.

It is important to note that under restriction 1, power profile $\mathbb{1}_z(t)$ and $\mathbb{P}_z(t)$ only depend on $h_{zu}(t)$. Since $h_{zu}(t)$’s are independent for each BS-MU pair $(z, u)$ $u \in U_z$ and across time $t$, restriction 1 helps simplify the ensuing analysis significantly.

We also notice that additive noise in the SINR expression (4) for deriving the lower bound on the expected delay.

Under this setup ($\ell(d_{\ell}) = 1$, no noise, and restriction 1 for BSs $z \neq n(m)$), from (4) for the $n(m) - m$ link

$$\text{SINR}_{n(m),m}(t) = \frac{P_{n(m)}(t)h_{n(m)}(t)\mathbb{1}_{n(m)}(t)}{\gamma \sum_{z \in \Phi \setminus(n(m))} \hat{P}_z(t)\mathbb{1}_z(t)h_z(t)\ell(z)}. \quad (14)$$

3. Otherwise, since MU and BS processes are independent, we get a thinned BS process, for which the derived results apply directly.

4. Essentially, what we are doing is that let $S^* = (p(t), P(t))$ be the (unknown) optimal strategy (power profile) that achieves minimum delay $\mathbb{E}(D^*)$ for each BS $z$. Then, under restriction 1, the power profile needed $\hat{S}_z = (\hat{p}_z(t), \hat{P}_z(t))$ for all BSs $z \neq n(m)$ to achieve $\mathbb{E}(D^*)$ at their respective MUs, is stochastically dominated by $S^*$. Thus, with each BS using $\hat{S}_z$, the expected delay at the typical MU $m$ is at most $\mathbb{E}(D^*)$.

Note that we have kept the path-loss $\ell(z)$ from BS $z$ to the typical MU $m$ as it is with restriction 1.

Thus, we have the tail probability $P(D > n|\Phi, \Phi_R)$

$$= \sum_{t=1}^{n} \mathbb{P}\left(\text{SINR}_{m}(t) < \beta|\Phi, \Phi_R\right), \quad (15)$$

where the second equality follows since given $\Phi, \Phi_R$, SINRs are independent; under restriction 1, $\mathbb{1}_z(t)$ and $P_z(t)$ only depend on $h_{zu}(t)$ that are independent across time, and $1_{n(m)}(t)$ and $P_{n(m)}(t)$ only depend on $h_{n(m)}(t)$ that are independent across time, and $h_z(t)$ is also independent.

In the Appendix B, we present the simpler proof for the case when the fading gain $h_{n(m)}(t)$ is not known at the typical BS, popularly called as the channel state information only at the receiver (CSIR). We proceed with the more challenging channel state information at both the transmitter and the receiver (CSIT) case as follows.

From (14), $P\left(\text{SINR}_{m}(t) < \beta|\Phi, \Phi_R\right)$

$$= \mathbb{P}\left(\frac{P_{n(m)}(t)h_{n(m)}(t)\mathbb{1}_{n(m)}(t)}{\gamma \sum_{z \in \Phi \setminus(n(m))} \hat{P}_z(t)\mathbb{1}_z(t)h_z(t)\ell(z)} < \beta|\Phi, \Phi_R\right). \quad (16)$$

Note that with CSIT, $1_{n(m)}(t)$ and $P_{n(m)}(t)$ can depend on $h_{n(m)}(t)$, and thus we need the following result from [22].

Under restriction 1, both $\mathbb{1}_z(t)$ and $1_{n(m)}(t)$ are independent across time, and are unknown at $m$. Moreover, there distributions are also unknown. Thus, to derive a lower bound, we let BS $n(m)$ choose the optimal transmission policy to minimize $\mathbb{P}(P_{n(m)}(t)h_{n(m)}(t)\mathbb{1}_{n(m)}(t) < \gamma\beta)$ from Proposition 8 that only depends on $h_{n(m)}(t)$ and $\gamma, \beta$. Thus, the outage probability $P\left(\text{SINR}_{m}(t) < \beta|\Phi, \Phi_R\right)$

$$= \mathbb{P}\left(h_{n(m)}(t) < \delta\right) \mathbb{P}\left(\frac{0}{I(t)} < \gamma\beta|\Phi, \Phi_R\right)$$

$$+ \mathbb{P}\left(h_{n(m)}(t) \geq \delta\right) \mathbb{P}\left(\frac{\beta}{I(t)} < \gamma\beta|\Phi, \Phi_R\right).$$

$$= \mathbb{P}\left(h_{n(m)}(t) < \delta\right) + \mathbb{P}\left(h_{n(m)}(t) \geq \delta\right) \mathbb{P}\left(\hat{I}(t) > 1|\Phi, \Phi_R\right),$$

$$\geq \mathbb{P}\left(h_{n(m)}(t) \geq \delta\right) \mathbb{P}\left(\hat{I}(t) > 1|\Phi, \Phi_R\right),$$

$$= \exp(-\delta) \mathbb{P}\left(\hat{I}(t) > 1|\Phi, \Phi_R\right), \quad (17)$$

where the last inequality follows since $h_{n(m)}(t) \sim EXP(1)$.

Now we derive bounds on $\mathbb{P}\left(\hat{I}(t) > 1|\Phi, \Phi_R\right)$ to derive a lower bound on the expected delay.

3 Under $\Phi, \Phi_R$ interference terms are independent across time and hence previous failures/successes have no effect on $1_{n(m)}(t)$ and $P_{n(m)}(t)$.

4 $\hat{p}_z(t)$, $\hat{P}_z(t)$, and $h_z(t)$’s are independent and are not available at the BS.
Recall that \( \mathbb{P}\left( \hat{I}(t) > 1 \mid \Phi, \Phi_R \right) \)
\[
= \mathbb{P}\left( \sum_{z \in \Phi \setminus \{n(m)\}} \hat{P}_z(t) \mathbb{I}_z(t) h_z(t) \ell(z) > 1 \mid \Phi, \Phi_R \right),
\]
\[
\geq 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \mathbb{P}\left( \hat{P}_z(t) \mathbb{I}_z(t) h_z(t) \ell(z) < 1 \mid \Phi, \Phi_R \right),
\]
\[
= 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \mathbb{E} \left\{ \left( 1 - \exp \left( \frac{-1}{\hat{P}_z(t) \mathbb{I}_z(t) \ell(z)} \right) \right) \mid \Phi, \Phi_R \right\},
\]
where the second equality follows since \( h_z(t) \)'s are independent for \( z, t \), and \( \hat{P}_z(t), \mathbb{I}_z(t) \) are independent for \( z \) under restriction 1, and the last inequality follows by taking the expectation with respect to \( h_z(t) \sim EXP(1) \).

Taking the expectation with respect to \( \mathbb{I}_z(t) \), we get
\[
\mathbb{P}\left( \hat{I}(t) > 1 \mid \Phi, \Phi_R \right) \geq 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \mathbb{E} \left\{ \left( 1 - \exp \left( \frac{-1}{\hat{P}_z(t) \ell(z)} \right) \right) \mid \Phi, \Phi_R \right\},
\]
\[
\geq 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \mathbb{E} \{ \hat{P}_z(t) \mid \Phi, \Phi_R \} \exp \left( \frac{-1}{\ell(z)} \right) \right),
\]
where the second inequality follows since \( \hat{P}_z(t) \hat{P}_z(t) \geq M/\tau \) for each BS \( z \), and the final inequality follows since \( \hat{P}_z(t) \leq 1 \), and the fact that \( \hat{P}_z(t) \) are independent for different BSs \( z \).

**Remark 2.** Recall that the BS density is much smaller than MU density, hence \(|U_z| > 1\), i.e., each BS \( z \) is transmitting to multiple MUs in its Voronoi cell. Thus, for \( \hat{P}_z(t) \) (that is independent across \( z \) and \( t \)) that only depends on fading gains from \( z \) to \( U_z \), we let \( \mathbb{E} \{ \hat{P}_z(t) \mid \Phi, \Phi_R \} \geq \eta > 0 \) (where \( \eta \) is a constant) where the expectation is with respect to the fading gains. Essentially, we have that the probability of transmission of any BS is bounded away from zero for any BS deployment \( \Phi \), which is reasonable to assume in practice.

Thus, it follows that \( \mathbb{P}\left( \hat{I}(t) > 1 \mid \Phi, \Phi_R \right) \)
\[
= 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right). \tag{18}
\]

Substituting (19) into (17), we get
\[
\mathbb{P}\left( \text{SINR}_m(t) < \beta \mid \Phi, \Phi_R \right) \geq \exp(-\delta) \left( 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right) ) \right),
\]
which from (15), gives \( \mathbb{P}(D > n \mid \Phi, \Phi_R) \)
\[
\geq \prod_{t=1}^{n} \exp(-\delta) \left[ 1 - \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right) \right]. \tag{20}
\]

**Remark.** Let \( g = \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right) \). Then
\[
\mathbb{E}\{D\mid \Phi, \Phi_R\} \geq \exp(\delta) \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right).
\]

Thus, using the definition of \( g \),
\[
\mathbb{E}\{D\mid \Phi, \Phi_R\} \geq \exp(\delta) \prod_{z \in \Phi \setminus \{n(m)\}} \left( 1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right) \right).
\]

To find \( \mathbb{E}\{D\} \) from \( \mathbb{E}\{D\mid \Phi, \Phi_R\} \), we first use the probability generating function \( \mathbb{P}(\Phi, \Phi_R) \) for the PPP \( \Phi \setminus \{n(m)\} \) to get \( \mathbb{E}\{D\mid d_0\} \)
\[
\geq \exp(\delta) \exp \left( \lambda \int_{\mathbb{R}^2} \mathbb{I}_{d_0} \left( \frac{1}{1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right)} - 1 \right) dz \right).
\]

Splitting the integral into two parts for \( d_0 \leq 1 \) and \( d_0 > 1 \), and keeping only the part with \( d_0 \leq 1 \), we get \( \mathbb{E}\{D\mid d_0\} \)
\[
\geq \exp(\delta) \exp \left( 2\pi \lambda \int_{\mathbb{R}, z > d_0, d_0 \leq 1} \frac{\eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right)}{1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right)} dz \right).
\]

For \( z < 1, \ell(z) = 1 \), thus, \( \mathbb{E}\{D\mid d_0\} \)
\[
\geq \exp(\delta) \exp \left( 2\pi \lambda \int_{d_0, d_0 \leq 1} \frac{\eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right)}{1 - \eta \exp \left( \frac{-1}{(M/\tau)\ell(z)} \right)} dz \right),
\]
\[
= \exp(\delta) \exp \left( 2\pi \lambda c_1 \int_{d_0} dz \right),
\]
\[
= \exp(\delta) \exp \left( 2\pi \lambda c_1 \left( 1 - \frac{d_0^2}{2} \right) \right),
\]

\footnote{Note that even for \( M = 1, \delta \approx .267. \) This approximation can be avoided, and exact expression can be derived but results in more tedious computations as will be evident next.}
where \( c_1 = \frac{\pi \exp \left( \frac{d_0^2}{2} \right)}{(1 - \pi \exp \left( \frac{d_0^2}{2} \right))} \). Then
\[
\mathbb{E}\{D|d_0\} \geq \exp(\delta) \exp \left( 2\pi \lambda c_1 \left( \frac{1 - d_0^2}{2} \right) \right),
\]
and taking the expectation with respect to the nearest BS distance \( d_0 \), whose PDF is given by Proposition 5, we get
\[
\mathbb{E}\{D\}
\]
\[
\geq \exp(\delta) \int_{x \leq 1} \exp \left( -\pi \lambda c_1 x^2 \right) f_{d_0}(x) dx,
\]
\[
= \exp(\delta) \exp \left( \pi \lambda c_1 \right)
\int_{x \leq 1} \exp \left( -\pi \lambda c_1 x^2 \right) 2\pi \lambda x \exp(-\lambda \pi x^2) dx,
\]
\[
= \exp(\delta) \exp \left( \pi \lambda c_1 \right)
\frac{1}{1 + c_1} (1 - \exp(-\pi \lambda (c_1 + 1))).
\]
Thus, (51) together with (22) give the required lower bound on the expected delay stated in Theorem 2.

The derived lower bound shows that the expected delay is u-shaped as a function of the BS density. For low-BS densities, ignoring the interference, where anyway it is weak, the delay is directly controlled by the “large” BS-MU distances. Thus, in the low-BS density regime, as expected, the delay decreases as the BS density increases, and our lower bound shows that the expected delay decreases polynomially with the increasing BS density.

For the high-BS density regime, the basic idea in showing that the expected delay increases exponentially with the BS density is to first make the path-loss for the typical user \( \ell(d_0) = 1 \) to completely eliminate the need for the BS strategy to depend on distance to its MU, since the path-loss function is at most 1. The possibility of arbitrarily large \( d_0 \) was the limitation for the ALOHA strategy that was exploited to show its unbounded expected delay in Corollary 3, which is avoided by using \( \ell(d_0) = 1 \) for deriving a lower bound over all strategies.

More importantly, we also make \( \ell(d) = 1 \) for all the non-typical BS-MU pairs, and completely remove the interference seen at any non-typical MU. This in turn makes the power transmission strategy for all non-typical BSs only depend on the fading gains that are independent across space and time. This enhancement provides a lower bound to the actual system performance. Using this independence of power transmission strategies from all the BSs that contribute interference at the typical MU, we then bound the expected delay via the tail probability (joint distribution of consecutive failure events) by considering only the interferers (BSs other than \( n(m) \)) that lie inside a disc of unit radius around the typical MU \( m \). Essentially, as \( \lambda \) increases, there are enough interferers in the unit disk, and the temporal correlation of SINRs is strong enough even under this relaxed setting.

\textbf{Proposition 9.} Let \( \mathcal{G} \) be the family of all non-negative, bounded measurable functions \( g : \mathbb{R}^d \to \mathbb{R} \) on \( \mathbb{R}^d \) whose support \( \{x \in \mathbb{R}^d : g(x) > 0\} \) is bounded. Let \( \mathcal{F} \) be the family of all functions \( f = 1 - g \), for \( g \in \mathcal{G} \), \( 0 \leq g \leq 1 \). Then the probability generating functional for a point process \( \Phi = \{x_n\} \) is defined as
\[
PGF(f) = \mathbb{E}\left\{ \prod_{x_n \in \Phi} f(x_n) \right\}.
\]

For a homogenous PPP \( \Phi \) with density \( \lambda \), the probability generating functional is given by
\[
PGF(f) = \exp[-f(1-f(x))\lambda dx].
\]

One question that is of immediate interest is: can the expected delay be decreased if small-scale BS coordination is allowed. We answer this in the following Corollary.

\textbf{Corollary 10.} When small scale BS coordination is allowed, i.e., if a fixed number (independent of BS density) of nearest BSs can schedule their transmissions jointly, then for any BS strategy
\[
\mathbb{E}\{D\} \geq \left( \frac{c_1}{\lambda^2} \right),
\]

as well as
\[
\mathbb{E}\{D\} \geq c_4 \left( \exp(c_2\lambda) \right).
\]

Consequently, the network wide capacity is
\[
C = \frac{\lambda}{\mathbb{E}\{D\}} \leq \min \left\{ \frac{\lambda^{\frac{d+1}{c_1}}}{c_3\lambda \exp(-c_2\lambda)} \right\}.
\]

\textbf{Proof.} The proof corresponding to the low-BS density regime (23) that is obtained by ignoring the interference remains the same as the bound (10).

Even when the interference is considered to get (24), most of the proof is identical to Theorem 2, where once again we make path-loss function \( \ell(d_0) = 1 \) for the typical BS-MU link, and enforce restriction 1. Supposing \( k \) BSs are scheduling their transmissions together, then at best the interference from the nearest \( k - 1 \) BSs (other than \( n(m) \)) can be eliminated from SINR_{n(m),k}(t) (4). Hence, following identical analysis as in Theorem 2, starting from (16), where replacing the sum in the interference term from \( z \in \Phi\setminus\{n(m)\} \) by \( z \in \Phi\setminus\{k \) nearest BSs from the origin\), till (21), where replacing the nearest BS distance \( d_0 \) with the \( k \)th nearest BS distance \( d_k \). we get
\[
\mathbb{E}\{D|d_k\} \geq \frac{1}{M} \exp\left( 2\pi \lambda c_1 \left( \frac{1 - d_k^2}{2} \right) \right).
\]

Taking the expectation with respect to the \( k \)th nearest BS distance \( d_k \), whose PDF is given by Proposition 11, we get the same order expression as in (22). Essentially, for deriving (22), we have shown that enough interferers lie in a unit disk around \( m \). Similar conclusion can be made even when considering interfering BSs other than the \( k - 1 \) nearest ones as \( \lambda \) increases.

\textbf{Proposition 11.} [23] The cumulative distribution function and probability distribution function of the \( k \)th nearest BS distance \( d_k \) is
\[
P(d_k > y) = \sum_{i=1}^{k-1} \frac{(\lambda \pi y)^i}{i!} \exp(-\lambda \pi y^2),
\]
\[
P(d_k = y) = \frac{(\lambda \pi y)^{k-1}}{(k-1)!} \exp(-\lambda \pi y^2).
\]
and
\[ f_{d_k}(y) = \frac{2(\lambda \pi)^k y^{2k-1}}{(k-1)!} \exp(-\lambda \pi y^2). \]

Proof. For \( d_k > y \), there can be at most \( k-1 \) BSs in the disc of radius \( y \) with node \( m \) as center located at the origin. Hence
\[ P(d_k > y) = \sum_{i=1}^{k-1} \frac{(\lambda \pi y)^i}{i!} \exp(-\lambda \pi y^2), \]
and differentiating \( P(d_k > y) \), we get the mentioned PDF. \( \square \)

V. ACHIEVABILITY

In this section, we consider a simple BS strategy (proposed by us in [13]) to achieve the capacity upper bound (Theorem 2) up to the same order.

Strategy: Let for MU \( u \), the distance to its nearest BS \( n(u) \) be \( d_u \). Let each BS know \( d_u \) for all the users connected to it. As usual, each BS serves all its connected users in a round-robin fashion by equally splitting the total time or bandwidth. For a slot \( t \) designated for a particular MU \( u \), BS \( n(u) \) transmits with probability \( P_n(u)(t) \) with power \( P_{n(u)}(t) \) given by
\[ P_{n(u)}(t) = c \ell^{-1}(d(u)), \quad P_n(u)(t) = M(P_n(u)(t))^{-1}, \tag{25} \]
where \( c = M(1-\epsilon)^{-1}, 0 < \epsilon < 1, \beta \gamma (1-\epsilon) < 1 \) is a constant, and \( M = P_{n(u)}(t)\rho_{n(u)}(t) \) is the average power constraint. Condition \( \beta \gamma (1-\epsilon) < 1 \) is technical and is required for ensuring certain function is \( < 1 \) in (41). Note that \( P_{n(u)}(t) \leq 1-\epsilon \), since \( \ell(d_u) \leq 1 \). Thus, in each time slot, with the proposed strategy, each BS makes transmission attempts with transmission power proportional to the distance to the MU it is serving, to completely nullify the path-loss. The transmission probability is chosen so as to satisfy the average power constraint of \( M \). It is worthwhile noting that the strategy does not use the knowledge of fading gain \( h_{n(m)}(t) \), and is not an adaptive strategy.

Remark 3. Note that for the proposed strategy (25), the expected power \( \mathbb{E}\{P_n(u)(t)\} < \infty \) and expected probability of transmission \( \mathbb{E}\{P_n(u)(t)\} > 0 \) for any BS \( n(u) \) from Proposition 5.

From hereon, for analysis purposes, we focus on a typical MU \( m \) and its nearest BS \( n(m) \), and derive an upper bound on the expected delay seen at MU.

Theorem 3. The power control strategy (25) achieves the following performance
\[ \mathbb{E}\{D\} \leq \sqrt{c_3 \left(1+\frac{\Gamma(\alpha+1)}{(\pi \lambda)^\alpha}\right) \exp(c_3 \lambda)}, \]
where \( c_3 \) and \( c_4 \) are constants. Thus,
\[ \mathbb{E}\{D\} = O(\exp(\lambda)), \]
and
\[ C = \Omega(\lambda \exp(-\lambda)). \]

Theorem 3 shows that a simple non-adaptive strategy that does not need to learn the local fading gain is capable of achieving (order-wise) the upper bound on the capacity \( C \). The only local information it needs is the distance \( d_0 \). Recall that Theorem 1 suggests that any strategy that can achieve non-zero capacity \( C \) needs to know \( d_0 \). Thus, the considered power control strategy is minimal in terms of its learning requirements. Distance \( d_0 \) can be learned easily via ranging or RSSI measurements and the considered power control strategy can be implemented easily in practice.

The considered power control strategy completely nullifies the path-loss seen at MU \( m \), and makes the received signal power independent of distance \( d_0 \). Essentially, Theorem 3 shows that as long as signal attenuation because of the path-loss function is compensated, order-optimal expected delay and capacity \( C \) can be achieved. Theorem 3 shows that a single strategy is sufficient to achieve the capacity upper bound in both the low and high BS density regimes, where the capacity behaviour is quite different. We would like to note that the following proof is similar to the one derived in [13], but is described in detail because of some key differences.

Proof. Without loss of generality, we will derive the upper bound on the expected delay for the typical user \( m \) that is served by its nearest BS \( n(m) \) at a distance of \( d_0 \) from it. We consider \( k \) successive slots (not necessarily consecutive) that are dedicated for transmission to \( m \) by BS \( n(m) \), and are interested in probability \( \mathbb{P}\{D > k\} \) to upper bound the expected delay, where delay \( D \) is as defined in Definition 1.

Typically, the number of MUs connected to different BSs are different. Thus, during the \( k \) considered slots for \( m \), any BS other than \( n(m) \) transmits potentially to different MUs. Let \( \mathcal{G}_k \) be the sigma field generated by the BS point process \( \Phi \) and MU point process \( \Phi_R \) and the choice (index) of MUs being served by BSs of \( \Phi \) at the above described \( k \) slots \( t = 1, 2, \ldots, k \).

With the above mentioned local power control strategy (25), the SINR seen at \( m \) in time slot \( t \) is
\[ \text{SINR}_{n(m),m}(t) = \frac{c h_{n(m)}(t) 1_{n(m)}(t)}{\gamma I(t)} + \mathcal{N}, \tag{26} \]
where
\[ I(t) = \sum_{z \in \Phi \setminus \{n(m)\}} 1_z(t) P_z(t) h_z(t) \ell(z). \tag{27} \]

With \( c_{n(m),m}(t) = 1 \) if \( \text{SINR}_{n(m),m}(t) > \beta \), and 0 otherwise, we have
\[ \mathbb{P}\{D > k | \mathcal{G}_k\} = \mathbb{E}\{ \mathbb{P}\{c_{n(m),m}(t) = 0, \forall t = 1, \ldots, k | \mathcal{G}_k\}\}. \tag{28} \]

Given \( \mathcal{G}_k \), with the described strategy (25), the transmission events \( 1_z(t) \), and the transmit powers \( P_z(t) \) are independent across time slots \( t \) for all BSs \( z \). Moreover, the fading gains \( h_z(t) \) are all independent. Hence, we get
\[ \mathbb{P}\{D > k | \mathcal{G}_k\} = \mathbb{E}\left\{ \prod_{t=1}^{k} \mathbb{P}\{c_{n(m),m}(t) = 0 | \mathcal{G}_k\}\right\}. \tag{29} \]

Let \( A(t) \) be the event that \( 1_{n(m)}(t) = 0 \), i.e., the BS \( n(m) \) does not transmit in the slot designated for user \( m \), while \( B(t) \) be
the event that \( \{1_{n(m)}(t) = 1\} \cap \{e_{n(m),m}(t) = 0\} \), i.e., BS \( n(m) \) transmits in slot \( t \) but the transmission fails. Then
\[
P[ D > k | G_k ] = \mathbb{E} \left\{ \prod_{t=1}^{k} P[A(t) \cup B(t) | G_k] \right\} ,
\]  
(30)

Because of our power transmission strategy, given \( \Phi \), transmission event \( 1_{n(m)}(t) \) and the success event \( e_{n(m),m}(t) \) are independent, hence \( P(A(t)|G_k) = 1 - p_{n(m)}(t) \), while \( P(B(t)|G_k) = p_{n(m)}(t) \)
\[
= p_{n(m)}(t) \left( 1 - \mathbb{E} \left\{ \exp \left( -\frac{\beta}{c} (N + \gamma I(t)) \right) | G_k \right\} \right) ,
\]
(31)
that follows by taking expectation with respect to \( h_{n(m)}(t) \sim EXP(1) \). Using the union bound, from (30), we get
\[
P[ A(t) \cup B(t) | G_k ] \leq 1 - p_{n(m)}(t) + p_{n(m)}(t) \left( 1 - \mathbb{E} \left\{ \exp \left( -\frac{\beta}{c} (N + \gamma I(t)) \right) | G_k \right\} \right) \]
\[
\leq 1 - p_{n(m)}(t) \exp \left( -\frac{\beta N}{c} \right) \mathbb{E} \left\{ \exp \left( -\frac{\beta}{c} I(t) \right) | G_k \right\} ,
\]
(32)
Let \( a = \frac{\beta N}{c} \), and focus on finding a lower bound on
\[ \mathbb{E} \{ \exp ( -aI(t) ) | G_k \} , \]
that is independent of the choice of the MU being served by BS \( z \). To this end, we first expand
\[
\mathbb{E} \{ \exp ( -aI(t) ) | G_k \} = \prod_{z \in \Phi \setminus\{n(m)\}} \mathbb{E} \{ \exp ( -a_1z(t)p_z^{(u(z))}(t)h_z(t)\ell(z) ) | G_k \} ,
\]
(33)
where BS \( z \in \Phi \setminus\{n(m)\} \) transmits to the MU \( u(z) \) in time slot \( t \). This fixes the transmission probability \( p_z^{(u(z))}(t) \) and power \( P_z^{(u(z))}(t) \) (where we have included the index \( u \) to make the dependence on the MU explicit). Then, taking the expectation with respect to \( 1_z(t) \), we have
\[
\mathbb{E} \{ e^{-a_1z(t)p_z^{(u(z))}(t)h_z(t)\ell(z)} | G_k \} = (1 - p_z^{(u(z))}(t)) + p_z^{(u(z))}(t) \mathbb{E} \{ \exp ( -aP_z^{(u(z))}(t)h_z(t)\ell(z) ) | G_k \} \]
\[
= (1 - p_z^{(u(z))}(t)) + p_z^{(u(z))}(t) \frac{1}{1 + a^{\gamma}(\ell(z)P_z^{(u(z))}(t))} ,
\]
(34)
where the second equality follows by taking expectation with respect to the independent fading gains \( h_z(t) \sim EXP(1) \).

Let \( u^*(z) \) be the MU for which the right hand expression in (34) is minimized, i.e., BS \( z \) causes maximum interference at \( m \) when it is serving MU \( u^*(z) \). Let \( p_z^*, P_z^* \) denote the corresponding transmission probability and power, respectively, for BS \( z \). Denote by \( 1_z^* \) an independent Bernoulli random variable with \( \mathbb{P}[1_z^* = 1] = p_z^* \). Define
\[
I^*(t) = \sum_{z \in \Phi \setminus\{n(m)\}} 1_z^* P_z^* h_z(t)\ell(z) .
\]
(35)
Essentially \( I^*(t) \) dominates the actual interference \( I(t) \) seen at \( m \). Substituting \( I^*(t) \) for \( I \) in (32) along with the observation that given \( \Phi, \Phi_R \), \( I^*(t) \geq I^*(1) \), we get
\[
P[ A(t) \cup B(t) | G_k ] \leq 1 - p_{n(m)}(t) \exp \left( -\frac{\beta N}{c} \right) \mathbb{E} \{ \exp ( -aI^*(1) ) | \Phi, \Phi_R \} .
\]
(36)
Let
\[
\theta = \exp \left( -\frac{\beta N}{c} \right) p_{n(m)}(t) \mathbb{E} \{ \exp ( -aI^*(1) ) | \Phi, \Phi_R \} .
\]
(37)
Substituting from (36) in (29), we get
\[
P[ D > k | \Phi, \Phi_R ] \leq (1 - \theta)^k .
\]
(38)
Then the expected delay can then be written as
\[
\mathbb{E}[D] = \sum_{k \geq 0} P[D > k] = \mathbb{E} \left\{ \sum_{k \geq 0} P[D > k | \Phi, \Phi_R] \right\} \leq \mathbb{E}[\theta^{-1}] ,
\]
where the last inequality follows from (37). Using the Cauchy-Schwartz inequality, on the two random variables in \( \theta^{-1} \), we get
\[
\mathbb{E}[D] \leq \exp (\beta N/c) \left( \frac{1}{\mathbb{E}\{\exp (-aI^*(1)) | \Phi, \Phi_R\}^2} \right) \mathbb{E}\{p_{n(m)}(t)^{-2}\}^{1/2} .
\]
(39)
From the definition of the transmission probability \( p_{n(m)}(t) = (M/c) \ell(d_0) = (M/c) \min\{1, d_0^{-\alpha}\} \), we get \( \mathbb{E}[p_{n(m)}(t)^{-2}] = \left( \frac{c}{M} \right)^2 \left( \int_0^1 f_{d_0}(x)dx + \int_1^{\infty} d_0^{2\alpha} f_{d_0}(x)dx \right) \),
\[
\leq \left( \frac{c}{M} \right)^2 \left( 1 + \int_0^{\infty} d_0^{2\alpha} f_{d_0}(x)dx \right) ,
\]
\[
= \left( \frac{c}{M} \right)^2 \left( 1 + \frac{\Gamma(\alpha + 1)}{(\pi\lambda)^\alpha} \right) .
\]
(40)
using the PDF of \( d_0 \) from Proposition 5.

Now we work towards bounding
\[
\mathbb{E}\{\exp(-aI^*(1)) | \Phi, \Phi_R\} .
\]
Recall,
\[
\mathbb{E}\{\exp(-aI^*(1)) | \Phi, \Phi_R\} = \prod_{z \in \Phi \setminus\{n(m)\}} \mathbb{E}\{\exp(-aP_z^*h_z(1)\ell(z)) | \Phi, \Phi_R\} .
\]
(41)
Taking expectations, first with respect to \( I_z^x \) and then with respect to fading gain \( h_z(1) \), we get
\[
\mathbb{E}\left\{ (a - a I_z^x h_z(1)) \ell(z) | \Phi, \Phi_R \right\} = (1 - p_z) + p_z \mathbb{E}\left\{ e^{a h_z(1) \ell(z)} | \Phi, \Phi_R \right\}
\]
\[
= 1 - p_z \left( 1 - \frac{1}{1 + a P_z^x} \ell(z) \right)
\]
\[
\overset{(a)}{=} 1 - \frac{\beta^2 P_z^x \ell(z)}{e + \beta^2 P_z^x \ell(z)}
\]
\[
\overset{(b)}{\geq} 1 - \frac{\beta^2 M(z)}{e} = 1 - \beta \gamma (1 - \epsilon) \ell(z),
\]
where (a) follows by resubstituting \( a = \frac{\beta^2}{2} \), and (b) by invoking the average power constraint \( p_z P_z^x \leq M \) for all \( z \) and in particular \( p_z P_z^x \leq M \) and \( e = M(1 - \epsilon)^{-1} \). Let \( c_2 = \beta \gamma (1 - \epsilon) \), where note that because of assumption on \( \epsilon \) (\( \epsilon \) satisfies \( (1 - \epsilon) \beta \gamma < 1 \)) in the power control strategy (25), \( c_2 \) < 1. Substituting the above bound in (40), we get
\[
\mathbb{E}\left\{ e^{a I_z^x(1)} | \Phi, \Phi_R \right\} \geq \prod_{z \in \Phi \setminus \{n(m)\}} (1 - c_2 \ell(z)). \quad (41)
\]
Hence
\[
\mathbb{E}\left\{ \frac{1}{(\mathbb{E}\{ e^{a I_z^x(1)} | \Phi, \Phi_R \})^2} \right\} \leq \mathbb{E}\left\{ \prod_{z \in \Phi \setminus \{n(m)\}} (1 - c_2 \ell(z)) \right\}. \quad (42)
\]
Once again using the probability generating functional (Proposition 9) for the PPP \( \Phi \setminus \{n(m)\} \), we get
\[
\mathbb{E}\left\{ \frac{1}{(\mathbb{E}\{ e^{a I_z^x(1)} | \Phi \})^2} \right\} \leq \mathbb{E}_0 \left\{ \exp \left( \lambda \int_{\mathbb{B}(0, d_0)} (\exp(-2 \log (1 - c_2 \ell(z))) - 1) \, dz \right) \right\},
\]
\[
\leq \mathbb{E}_0 \left\{ \exp \left( \frac{2 \lambda c_2}{(1 - c_2)^2} \int_{\mathbb{R}^2} \ell(z) \, dz \right) \right\}, \quad (43)
\]
where \( \mathbb{B}(0, d_0) \) is the disc with radius \( d_0 \) centered at the origin, and the last inequality follows by noting that \( \ell(z) \leq 1 \). Similar to the proof of Theorem 2, we can separate the integral into two parts \( z > d_0 < 1 \) (where \( \ell(z) = 1 \)) and \( z > d_0 > 1 \) (where \( \ell(z) = z^{-\alpha} \)), we get
\[
\mathbb{E}\left\{ \frac{1}{(\mathbb{E}\{ e^{a I_z^x(1)} | \Phi \})^2} \right\} \leq \int_0^1 \exp \left( \frac{2 \lambda c_2}{(1 - c_2)^2} \left( 1 - \frac{x^2}{2} \right) \right) f_{d_0}(x) \, dx + \int_1^\infty \exp \left( \frac{2 \lambda c_2}{(1 - c_2)^2} \frac{x^{2-\alpha}}{\alpha - 2} \right) f_{d_0}(x) \, dx,
\]
where \( c_3 \) is a constant. Combining this with (39), from (38) we get
\[
\mathbb{E}\{D\} \leq \sqrt{c_4 \left( 1 + \frac{\Gamma(\alpha + 1)}{(\pi \lambda)^{\alpha}} \right)} \exp(c_3 \lambda),
\]
where \( c_4 \) is a constant. This completes the proof of Theorem 3. \( \square \)

VI. Extensions

We consider two potential extensions of our system model that can possible provide with better capacity than the results presented in this paper.

A. Multiple Antennas

One possible remedy to improve the capacity scaling with increasing BS density is to use multiple antennas [24]–[26] at the BS and the MUs. We next show via a lower bound on the expected delay that the order-wise scaling of the expected delay and the capacity remains unchanged even with multiple antennas. We consider the case when each BS is equipped with \( N \) antennas. Case of MUs also having multiple antennas follows similarly. For sake of brevity, we just indicate the change that multiple antennas can make to the lower bound of Theorem 2.

With \( N \) antennas at each BS, that are used for beamforming, compared to the SINR expression with single antennas (4), the only change in the SINR is in terms of the signal power, that is now \( |h_t(n(m))| \), a chi-square distributed random variable with \( 2N \) degrees of freedom.

Similar to the technique used to find the lower bound corresponding to low-BS densities in Theorem 2, we remove all the interference seen at \( m \), to get
\[
\text{SINR} = \frac{P_{n(m)}(t)|h_{n(m)}(t)| \ell(d_o) 1_{n(m)}(t)}{N}. \quad (45)
\]

Thus, similar to the idea used to derive lower bound (10), in Theorem 2, for \( \ell(d_o) \geq \frac{\beta}{\lambda} \), we let \( |h_t(n(m))| = N \), and when \( \ell(d_o) < \frac{\beta}{\lambda} \), we consider a stronger channel \( |h_t(n(m))| \) with PDF \( f_{h_t}(x) = \frac{c}{(N-1)!} x^{\alpha - 1} \exp(-x) \) compared to \( |h_t(n(m))| \sim \chi_N^2 \) that has PDF \( f_{h_t}(x) = \frac{1}{\sqrt{2\pi N}} x^{N-1} \exp(-x) \). The rest of the proof follows identically to the proof of (10), where the resulting lower bound with multiple antennas differs only in constants compared to lower bound (10).

Next, to derive bound similar to (11), we consider interference, where the SINR at the typical MU \( m \) is
\[
\text{SINR} = \frac{P_{n(m)}(t)|h_{n(m)}(t)| \ell(d_o) 1_{n(m)}(t)}{\sum_{z \in \Phi \setminus \{n(m)\}} P_z(t) 1_z(t) h_z(1) z + N}. \quad (46)
\]
From (15), recall that for the \( (n(m) - m) \) link, the tail probability is
\[
\mathbb{P}(D > |t| \Phi) = \prod_{l=1}^n \mathbb{P}(\text{SINR}_m(t) < \beta |\Phi, \Phi_R),
\]
where \( \mathbb{P}(D > |t| \Phi) \) is a normalizing constant.
The CDF of a chi-square distributed random variable $X$ with $2N$ degrees of freedom is given by $P(X \leq x) = \int_0^x \frac{x^{N-1}}{(N-1)!} \exp(-s) ds$. For large $x$, the CDF is well approximated by $P(X \leq x) \approx \frac{x^{N-1}}{(N-1)!} \exp(-x)$, which is more importantly bounded by
\[
P(X \leq x) \leq \frac{1}{(N-1)!} \exp\left(-\frac{x}{N-1}\right),
\]
for large $x$.

With $\ell(d_0) = 1$ for the typical user $m$, and dropping additive noise contribution in (46), as before, we get $P(\text{SINR}_m(t) < \beta | \Phi, \Phi_R) = P\left(\frac{P_{n(m)}(t) | 1_{n(m)}(t) | h_z(t) \ell(z)}{\gamma \sum_{z \in \Phi \setminus n(m)} P_z(t) 1_z(t) h_z(t) \ell(z)} < \beta | \Phi, \Phi_R\right)$.

Using the CDF approximation for $h_{n(m)}(t)$ from (47), since interference is large, we get $P(\text{SINR}(t) < \beta | \Phi, \Phi_R) = 1 - \frac{1}{(N-1)!}$.

\[
E\left\{\exp\left(-\beta \gamma \sum_{z \in \Phi \setminus n(m)} P_z(t) 1_z(t) h_z(t) \ell(z)\right) \mid \Phi, \Phi_R\right\}.
\]

From hereon, we replace $\beta$ by $\beta/(N-1)$, which allows identical analysis as in the proof of Theorem 2, starting from (15) under restriction 1. Thus, even with multiple antennas, the scaling of the expected delay with respect to the BS density remains the same as in Theorem 2.

\section{Coordinated Multi-Point}

Another modern technique to improve the performance of cellular wireless networks, is the use of coordinated multi-point (CoMP) \cite{27}, where a BS shares its signal with multiple BSs, which then transmit it coherently. Essentially, if $k$ BSs are cooperating using CoMP, then the received signal model at the MU is equivalent to having $k$ transmit antennas at the BS it is connected to without power normalization, and in addition removing the interference from the $k-1$ coordinating other BSs. Thus, CoMP is a combination of having multiple antennas at BS, and small-scale BS coordination. Thus, using results from Corollary 10, and Subsection VI-A, it easily follows that CoMP also cannot change the order-wise scaling of the expected delay and the capacity.

\section{Simulation Results}

In this section, we provide some numerical results to better understand the expected delay performance of cellular wireless networks. We consider a BS location process, where the number of BSs is Poisson distributed with mean 200, and each BS lies on a two-dimensional disk of radius $r$. To simulate a particular BS density $\lambda$, the disk radius $r$ is adjusted accordingly, and the typical MU is placed at the center of the disk. In all simulations, we use the path-loss exponent $\alpha = 3$, and SINR threshold $\beta = 1$ corresponding to $R = 1$ bits/sec/Hz, and average power constraint $M = 1$.

In Fig. 3, we begin by plotting the expected delay of the power control strategy (25) in the low-BS density regime. As derived, the expected delay closely follows the polynomial lower bound of $1/\lambda^{\alpha/2}$ in the low-density regime. In Fig. 4, we plot the expected delay with the power control strategy for higher BS densities, where as expected, the expected delay grows exponentially closely following the derived lower and upper bounds.

In Fig. 5, we illustrate the expected delay for an ALOHA strategy with BS coordination, where to emulate BS coordination, interference from the nearest $K$ BSs is neglected at the typical MU located at the origin. We see that the delay performance improves, but not order-wise. Finally, in Fig. 6, we simulate the performance of CoMP, where for the ALOHA strategy, BS coordination (similar to Fig. 5) together with multiple antennas at BS is employed. Each BS has $N$ antennas that it uses for beamforming towards its MU. Once again, as discussed earlier, we see that there is no substantial improvement in expected delays with employing multiple antennas at each BS.

\section{Conclusion}

In this paper, we derived tight bounds for a natural and practical definition of capacity for cellular wireless networks. Even though our capacity metric is weaker than the Shannon capacity, however, it closely matches the throughput measure observed in a real-life implementation of cellular wireless networks, where ARQ is implemented universally. Most importantly, we were able to derive the exact dependence of the BS density on the capacity of cellular networks, which has generally escaped analytical tractability. Conventional wisdom suggests that there is a advantage in increasing BS density; via increasing the SINR for the cell-edge users or improving the frequency reuse. Our work, exactly characterizes the capacity behaviour as a function of the BS density. For low-BS densities, the capacity increases polynomially with the BS density, while as BS density is increased further the capacity starts to decrease exponentially. The prior conclusions were drawn depending
wireless networks, in this paper, we also upper bound the (capacity or connection probability) analysis of a fixed BS is achievable by a simple BS strategy, that only depends on the distance between the BS and the MU, that we proposed in our earlier work.

Fig. 4. Expected delay of the power control strategy for higher BS densities referenced with the lower and the upper bound.

Fig. 5. Comparison of the expected delay of the ALOHA strategy with BS cooperation, where 7 closest BSs are not transmitting.

Fig. 6. Comparison of the expected delay of the ALOHA strategy with CoMP; BS cooperation where 7 closest BSs are not transmitting, and there are N antennas at the BS.

REFERENCES

Given a realization of the BS PPP $\Phi$, the distance $d_0$ is fixed. Given $\Phi$, we will find a lower bound on $E\{D|\Phi\}$. With fixed $M, \beta$ for two different cases depending on $d_0$, we will consider stronger fading gains to lower bound the outage probabilities, and consequently lower bound the expected delay, as follows.

Case 1: $\ell(d_0) \geq \frac{1}{M\beta}$. For this case, we first show that the outage probability $p_o = P(\text{SINR} \leq \beta|\Phi)$ (49) with $h \sim EXP(1)$ is larger than when $h \equiv 1$ (no fading/line of sight). With $h \sim EXP(1)$, from Proposition 8, for any $M$, the outage probability $p_o = P(\text{SINR} \leq \beta|\Phi)$ is at least $1 - \exp(\delta)$, where $\delta$ is chosen to satisfy the average power constraint $\int_0^\infty \frac{\beta}{M(\beta)} x \exp(-x) dx = \frac{1}{M\beta}$.

Case 2: $\ell(d_0) < \frac{1}{M\beta}$. To derive a lower bound in this case, we replace $h_{n(m)} \sim EXP(1)$ with ‘stronger’ $h'_{n(m)}$ that has PDF $f_{h'_{n(m)}}(x) = x \exp(-x)$ and CDF $P(h'_{n(m)} < x) = 1 - \exp(-x)$. With $h_{n(m)} \sim EXP(1)$, the CDF is $P(h_{n(m)} < x) = 1 - \exp(-x)$.

With the stronger fading gain $h'_{n(m)}$, conditioned on $d_0$, from Proposition 8, with an average power constraint of $M$, the outage probability $p_o = \mathbb{P}\left(\frac{P_{n(m)/h'_{n(m)}}}{N}I_{n(m)}(t)\ell(d_0) \leq \beta | \Phi\right)$ is minimized when $P_{n(m)} = \frac{\beta}{\ell(d_0)h'_{n(m)}}$ and consequently lower bound the expected delay, as follows.

Thus, combining case 1 and 2, given $\Phi$, the expected delay is at least

$$E\{D|\Phi\} = 1 + \frac{M(\beta)}{\ell(d_0)} < 1 + \frac{1}{\beta\ell(d_0)} < 1 + \beta \left(1 - \ln \frac{M(\beta)}{\beta}\right) \frac{1}{\beta\ell(d_0)} \leq \frac{\beta}{M\beta} \leq 1,$$

where the first equality follows since given $\Phi$, the success events ($h_{n(m)} > \beta$) are independent across time slots. From Proposition 5, we know that $f_{d_0}(y) = 2\pi \lambda y \exp(-\lambda y^2)$. Hence, the resulting outage probability, $p_o = \mathbb{P}(h'_{n(m)} < \delta) = \int \frac{\beta}{M\beta} \leq 1 \left(1 - \frac{1}{\beta\ell(d_0)} \frac{M(\beta)}{\beta}\right)$, and the success probability $p_s = 1 - p_o = \left(1 - \frac{1}{\beta\ell(d_0)} \frac{M(\beta)}{\beta}\right)$. For this case, we first show that

$$\mathbb{P}(D_0 < d) = 2\pi \lambda y \exp(-\lambda y^2) dy,$$

where $c$ is a constant.

APPENDIX B

UPPER BOUND ON CAPACITY UNDER CSIR

As before, consider the restriction 1 for all the non-typical BSs and the MUs served by them. Moreover, $\ell(d_0) = 1$ for the typical MU. Let $\Phi, \Phi_R$ be the sigma field generated by the BS point process $\Phi$, and the MU point process $\Phi_R$. Thus, we have the tail probability $P(D > n|\Phi, \Phi_R)$

$$= \mathbb{P}\left(\text{SINR}_m < \beta, \ldots, \text{SINR}_m < \beta | \Phi, \Phi_R\right),$$

$$= \prod_{i=1}^n \mathbb{P}\left(\text{SINR}_m(t) < \beta | \Phi, \Phi_R\right),$$

where the second equality follows since given $\Phi, \Phi_R$, the only randomness left in the SINRs (SINR) is because of the fading gains $h$'s that are independent across time slots and all BS-MU pairs. Thus, $P(D > n|\Phi, \Phi_R)$

$$= \mathbb{P}\left(\sum_{z \in \Phi, \Phi_R} P_{n(m)}(t)h_{n(m)}(t)I_{n(m)}(t) \leq \beta | \Phi, \Phi_R\right).$$

Taking the expectation with respect to $h_{n(m)}(t) \sim EXP(1)$ (note that $I_{n(m)}(t)$ and $P_{n(m)}(t)$ cannot depend on $h_{n(m)}(t)$)
under CSIR, which makes the analysis easy as follows, we get
\[ \mathbb{P}\left(\text{SINR}(t) < \beta|\Phi, \Phi_R\right) = 1 - \mathbb{E}\left\{ \exp\left(\frac{-\beta\gamma}{P_{n}(t)} \sum_{z \in \Phi(\{n\})} \tilde{P}_z(t) \tilde{1}_z(t) h_z(t) \ell(z) \right) | \Phi, \Phi_R\right\}. \]

Taking the expectation with respect to $1_{n(m)}(t)$,
\[ \mathbb{P}\left(\text{SINR}(t) < \beta|\Phi, \Phi_R\right) = 1 - \mathbb{E}\left\{ p_{n(m)}(t) \exp\left(\frac{-\beta\gamma}{P_{n}(t)} \sum_{z \in \Phi(\{n\})} \tilde{P}_z(t) \tilde{1}_z(t) h_z(t) \ell(z) \right) | \Phi, \Phi_R\right\} = 1 - M \mathbb{E}\left\{ \exp\left(\frac{-\beta\gamma}{P_{n}(t)} \sum_{z \in \Phi(\{n\})} \tilde{P}_z(t) \tilde{1}_z(t) h_z(t) \ell(z) \right) | \Phi, \Phi_R\right\}, \]
where the second equality follows since $p_{n(m)}(t)P_{n}(t) \leq M$.

Rearranging terms, \[ \mathbb{P}\left(\text{SINR}(t) < \beta|\Phi, \Phi_R\right) \geq 1 - M \prod_{z \in \Phi(\{n\})} \left( \frac{1 - \tilde{P}_z(t)}{1 + \tilde{P}_z(t)} \left( M/\tau + M\gamma\beta\ell(z) \right) \right), \]
which is (55) into (52), \[ \mathbb{P}(D > n|\Phi, \Phi_R) \geq \prod_{t=1}^{n} \left( 1 - M \prod_{z \in \Phi(\{n\})} \left( \frac{1 - \tilde{P}_z(t)(M/\tau + M\gamma\beta\ell(z))}{1 + M\gamma\beta\ell(z)} \right) \right). \]

As mentioned before, under restriction 1, the power profile $\tilde{P}_z(t)$ and $\tilde{P}_z(t)$ of BS $z$ only depend on $h_{zu}(t)$’s that are independent for each BS-MU pair $(z,u)$, $u \in U(t)$ across time $t$. Thus, taking the expectation of (20) with respect to $\tilde{P}_z(t)$ that are independent (across both $z$ and $t$), and denoting $\eta = \mathbb{E}\{\tilde{P}_z(t)\}$, we get \[ \mathbb{P}(D > n|\Phi) \geq \prod_{t=1}^{n} \left( 1 - \eta(M/\tau + M\gamma\beta\ell(z)) \frac{1 - \eta(M/\tau + M\gamma\beta\ell(z))}{1 + M\gamma\beta\ell(z)} \right). \]

Let $g = M \sum_{z \in \Phi(\{n\})} \left( 1 - \eta(M/\tau + M\gamma\beta\ell(z)) \frac{1 - \eta(M/\tau + M\gamma\beta\ell(z))}{1 + M\gamma\beta\ell(z)} \right)$, where $g \leq 1$. Then \[ \mathbb{E}\{D|\Phi\} \geq \sum_{n=0}^{\infty} \sum_{n=0}^{\infty} \left( 1 - g \right)^n, \]
and \[ \mathbb{E}\{D|\Phi\} \geq M \int_{\mathcal{P}} \lambda \int_{\mathbb{R}} \int_{\mathbb{R}} \left( 1 + M\gamma\beta\ell(z) - \eta M\gamma\beta\ell(z) \right) \right. \]

Splitting the integral into two parts for $d_0 \leq 1$ and $d_0 > 1$, and keeping only the part with $d_0 \leq 1$, we get \[ \mathbb{E}\{D|d_0\} \geq \frac{1}{M} \exp \left( 2\pi\lambda \int_{\mathbb{R}^2} \left( 1 + M\gamma\beta\ell(z) - \eta M\gamma\beta\ell(z) \right) \right. \]

where the second and the third inequalities follow since $\tilde{P}_z(t) \tilde{P}_z(t) \geq M/\tau$, and $\tilde{P}_z(t) \tilde{P}_z(t) \leq M$, respectively. For the final inequality, we bound the $\tilde{P}_z(t) \leq 1$ in the denominator by $\tilde{P}_z(t) = 1$.

Thus, substituting (55) into (52), \[ \mathbb{P}(D > n|\Phi, \Phi_R) \geq \prod_{t=1}^{n} \left( 1 - M \prod_{z \in \Phi(\{n\})} \left( \frac{1 - \tilde{P}_z(t)(M/\tau + M\gamma\beta\ell(z))}{1 + M\gamma\beta\ell(z)} \right) \right). \]

Let $c_1 = \left( \frac{1 + M\gamma\beta}{1 + M\gamma\beta - \eta(M/\tau + M\gamma\beta)} \right)$. Then \[ \mathbb{E}\{D|d_0\} \geq \frac{1}{M} \exp \left( 2\pi\lambda c_1 \left( 1 - \frac{d_0^2}{2} \right) \right). \]
and taking the expectation with respect to the nearest BS distance $d_0$, whose PDF is given by Proposition 5, we get

$$
\mathbb{E}\{D\} \geq \frac{1}{M} \exp(\pi \lambda c_1) \int_{x \leq 1} \exp(-\pi \lambda c_1 x^2) f_{d_0}(x) dx,
$$

$$
= \frac{1}{M} \exp(\pi \lambda c_1) \int_{x \leq 1} \exp(-\pi \lambda c_1 x^2) 2\pi \lambda x \exp(-\lambda \pi x^2) dx,
$$

$$
= \frac{1}{M} \exp(\pi \lambda c_1) \frac{1}{1 + c_1} (1 - \exp(-\pi \lambda(c_1 + 1))). \tag{59}
$$

Thus,

$$
\mathbb{E}\{D\} = \Omega(\exp(\lambda)),
$$

and consequently

$$
C = O(\lambda \exp(-\lambda)).
$$