Optimizing Data Aggregation for Uplink Machine-to-Machine Communication Networks

Derya Malak, Harpreet S. Dhillon, and Jeffrey G. Andrews

Abstract

Machine-to-machine (M2M) communication’s severe power limitations challenge the interconnectivity, access management, and reliable communication of data. In densely deployed M2M networks, controlling and aggregating the generated data is critical. We propose an energy efficient data aggregation scheme for a hierarchical M2M network. We develop a coverage probability-based optimal data aggregation scheme for M2M devices to minimize the average total energy expenditure per unit area per unit time or simply the energy density of an M2M communication network. Our analysis exposes the key tradeoffs between the energy density of the M2M network and the coverage characteristics for successive and parallel transmission schemes that can be either half-duplex or full-duplex. Comparing the rate and energy performances of the transmission models, we observe that successive mode and half-duplex parallel mode have better coverage characteristics compared to full-duplex parallel scheme. Simulation results show that the uplink coverage characteristics dominate the trend of the energy consumption for both successive and parallel schemes.

I. INTRODUCTION

Machine-to-machine (M2M) applications are rapidly growing, and will become an increasingly important source of traffic and revenue on 4G and 5G cellular networks. Unlike video applications, which are expected to consume around 70% of all wireless data by the end of the decade [1], M2M devices will use a comparatively small fraction. On the other hand, M2M communication has its own challenges. The air interface design for high-data-rate applications...
may not effectively support M2M’s vast number of devices, each usually having only a small amount of data to transmit. Therefore, M2M will require sophisticated access management and resource allocation techniques with QoS constraints to prevent debilitating random access channel (RACH) congestion \[2\], and to enable link adaptation with low overhead, reduced energy consumption, and efficient control channel design \[3\]. Bluetooth (IEEE 802.15.1), Zigbee (IEEE 802.15.4) and WiFi (IEEE 802.11b) are a few examples of current technologies that have been used for M2M communication \[3\]. Meanwhile, 3GPP has been studying M2M communication in its standardization process for Long Term Evolution-Advanced (LTE-A). In Release 13, a low-complexity user equipment (UE) category for M2M devices is specified, in which a UE has reduced bandwidth and lower maximum transmission power, and can operate in all the duplex modes supported by LTE \[4\]. However, there is still no consensus on the general network architecture for large scale M2M communication.

A. Motivation and Related Work

Unlike most human generated or consumed traffic, M2M as defined in this paper is characterized by a very large number of small transactions, often from battery powered devices. The power and energy optimal uplink design is studied for various multiple access strategies in \[5\], while an optimal uncoordinated strategy in terms of maximizing the average throughput for a time-slotted RACH is developed in \[6\]. It is shown that for the small payload sizes relevant for M2M, a strategy that transmits both identity and data over the RACH can support significantly more devices compared to the conventional approach, where data transmissions are scheduled after an initial random-access stage.

A comprehensive performance evaluation of the energy efficiency of the random access mechanism of LTE and its role for M2M is provided in \[7\]. A queueing model to analyze heterogeneous M2M communication where M2M devices coexist with human-to-human devices and perform relay transmissions for each other to reduce network congestion and improve energy efficiency is developed in \[8\]. An energy-efficient design for the uplink of LTE networks in M2M and human-to-human coexistence scenarios that also satisfies quality-of-service (QoS) requirements is developed in \[9\]. Similar to \[7\]–\[9\], we study an energy-efficient design for M2M uplink where devices perform multi-hop transmissions. We also incorporate the coverage characteristics for different transmission modes using tools from stochastic geometry.
Because low-power M2M devices may not be able to communicate with the BS directly, hierarchical architectures may be necessary. Hence, critical design issues also include optimizing hierarchical organization of the devices and energy efficient data aggregation. Although these issues have not been studied in the context of M2M communication, there is prior work on distributed and hierarchical networks in the context of wired communications. In [10], energy consumption is optimized by studying a distributed protocol for stationary ad hoc networks. In [11], a distribution problem which consists of subscribers, distribution points and concentration points for a wired network model is studied. The objective is to minimize a cost function by optimizing the density of distribution points. In [12], a hierarchical model for a sensor network including multiple sinks, compressors/aggregation nodes, and sensors is considered, and their approach is shown to yield significant energy savings.

Hierarchical networks can provide efficient data aggregation in M2M or other power-limited communication systems to enable successful end-to-end data transmission. Despite previous research efforts, to the best of our knowledge, there has been no study focusing on data aggregation schemes for wireless networks together with the rate coverage characteristics in M2M communication, especially from an energy optimal design perspective. Providing such a study is the main contribution of this paper.

B. Contributions and Organization

A large-scale hierarchical wireless network model for M2M communication. We propose a new communication model for the M2M uplink. In Sect. II, we consider a hierarchical scenario to model the wireless transmissions where the hierarchical levels describe the multi-hop stages of the model. Each level is composed of the transmitter and aggregator processes. Once the transmissions of the current hierarchical level are completed, the transmitters are either turned off or kept on depending on the transmission scheme used, and in the subsequent hierarchical level, the aggregator process forms the new transmitter process. Since each device can act as a transmitter or an aggregator in a particular time slot, we model the devices as wireless transceivers. In Sect. IV, the aggregation process is repeated over multiple levels to generate the hierarchical transmission model. We show there is an optimal fraction of aggregators that minimizes the overall energy consumption.

SIR and rate coverage probability for a heterogeneous M2M network. In Sect. V, we
analyze the SIR coverage and rate coverage characteristics of the multi-stage transmission process to determine the optimal number of stages. We consider two possible transmission techniques: i) *successive scheme*, where the hierarchical levels are not active simultaneously, and ii) *parallel scheme*, where either all the levels are active simultaneously as in full-duplex transmission or where the active levels are interleaved as in half-duplex mode. We describe the proposed transmission models in detail in Sect. VI. We compare their energy efficiencies in Sects. IV and VII by averaging the total energy consumption over multiple transmission cycles.

**Optimizing the number of multi-hop stages.** We propose a general aggregator model for power limited M2M communication devices and BSs. Since the M2M devices are power limited and thus range limited, multi-hop routing is a feasible transmission strategy rather than direct transmissions. However, in designing multi-hop communication protocols, the number of hops cannot be increased arbitrarily due to the additional energy consumption incurred by relays; long-hop routing is a competitive strategy for many networks [13]. In Sect. V, we find an upper and lower bound on the optimal number of hops.

**Optimizing the network energy density.** We use a generic transceiver scheme to model the energy consumption of the devices. For multi-hop transmission models, we find the optimal fraction of aggregators independently chosen from the set of devices that minimizes the average total energy consumption per unit area, i.e., the mean total energy density, for a fixed payload per device of an M2M network. We incorporate the energy consumption of aggregators and transceiver circuit components into the energy model. In Sects. III and IV, we optimize the total energy density of the transmissions based on a given SIR coverage requirement for the devices.

We evaluate the performance of the proposed transmission techniques in Sect. VII and provide a comparison in terms of their energy densities, communication rates and transmission delays through numerical investigations.

**II. System Model**

We consider a cellular-based uplink model for M2M communication where the BS and device locations are distributed as independent Poisson Point Processes (PPPs) with respective densities of $\lambda_{BS}$ and $\lambda$ with $\lambda \gg \lambda_{BS}$. Each BS has an average coverage area of $\lambda_{BS}^{-1}$, and each device has a fixed payload of $M$ bits to be transmitted to the BS. We also assume open loop power control under which the average received power at the BS from the devices in its coverage area is
constant and equal to $P_R$. With this assumption, the transmit power of a device located at distance $d$ from the BS is $P_T = P_Rd^\alpha$, where $\alpha$ is the path loss exponent. Assuming $N$ simultaneous transmissions, the uplink SINR for any device is $\text{SINR} = \frac{P_R((N-1)P_R + I_{oc} + N_0W)^{-1}}{1}$, where $I_{oc}$ is the out of cell interference, $N_0$ is the power spectral density and $W$ is the bandwidth.

A. Data Aggregation and Transmission Model

Devices transmit data to the BS by aggregating data. The initial device process $\Psi$ is independently thinned by probability $\gamma < 0.5$ to generate two independent processes\footnote{Later in Sect. V in evaluating the SIR-based coverage probability, we also incorporate the small-scale fading into the analysis that is assumed to be independent and identically distributed (iid) with unit mean. Therefore, incorporating fading yields the same average energy analysis. To keep the notation simple, we do not incorporate fading in Sects. II, III and IV.} the aggregator process $\Psi_a$ with density $\lambda_a$ and the device (transmitter) process $\Psi_u$ with density $\lambda_u$, where $\lambda = \lambda_u + \lambda_a$. Each transmitting device picks the closest receiving device (aggregator) to transmit its payload, i.e., the devices within the Voronoi cell of the typical aggregator device will transmit to that aggregator. In Fig. 1 we illustrate a single-stage system model, where $\Psi_a$ is obtained by independently thinning $\Psi$ with probability $0.3$, i.e., $\Psi_a$ has a density of $0.3\lambda$. In this particular example, aggregators directly relay all data to the BS.

The aggregation process can be extended to multiple stages. Each hierarchical level is composed of the transmitter and the aggregator processes, where the aggregator processes of all stages are initially determined such that they are disjoint from each other. At each stage, after the set of transmitters transmit their payloads to their nearest aggregators, and once the transmissions of a hierarchical level are completed, the transmitters are turned off and excluded from the process. In the subsequent hierarchical level, the aggregators of the previous stage become the transmitters, and they transmit their data to the aggregator process of the new stage. The data aggregation process is repeated over multiple stages to generate the hierarchical transmission model. The process ends when all the payload is transmitted to the BS in the last stage of this multi-hop process, which we call a transmission cycle. The data aggregation model will be detailed in Sects. III and IV.

2In Sect. III we will motivate the choice of $\gamma < 0.5$ in our setup.
B. Wireless Transceiver Model

We model the devices as wireless transceivers, since each device can act as a transmitter or an aggregator in a particular time slot. In [14], a generic architecture for energy-limited wireless transceivers is provided. The transceiver has four major building blocks. The transmit block (TX) is responsible for modulation and up-conversion, the receive block (RX) for down-conversion and demodulation, the local oscillator (LO) block for the generation of the required carrier frequency, and the power amplifier (PA) block for amplification of the signal to produce the required RF transmit power $P_T$. The power consumption in the receive and transmit paths are

$$P_{rxr} = P_{LO} + P_{RX} + P_O, \quad P_{txr} = P_{LO} + P_{TX} + P_{PA},$$

where the power consumption of LO, RX and TX blocks are denoted by $P_{LO}$, $P_{RX}$ and $P_{TX}$, respectively, which are non-negative constants. $P_O$ is the receiver overhead power that is assumed to be constant. The PA power consumption is given by $P_{PA} = \eta^{-1} P_T$, where $\eta$ is the PA efficiency, which is constant in the linear regime. The average energy cost of the transceiver is given by $\beta_R \cdot P_{rxr} + \beta_T \cdot P_{txr}$ where $\beta_T = \bar{t}_{tx}/(\bar{t}_{tx} + \bar{t}_{rx})$ and $\beta_R = \bar{t}_{rx}/(\bar{t}_{tx} + \bar{t}_{rx})$, and $\bar{t}_{rx}$ and $\bar{t}_{tx}$ stand for the average receive and transmit times. The symbol definitions are given in Table I.
| Parameter                                                                 | Symbol     | Parameter                          | Symbol     |
|--------------------------------------------------------------------------|------------|------------------------------------|------------|
| Power consumption in the receive (transmit) path                         | $P_{txr}$  | PA efficiency                      | $\eta$    |
| Power consumption of the receive (transmit) block                        | $P_{RX}$   | Transmit distance                  | $d$        |
| Power consumption of the local oscillator (LO)                           | $P_{LO}$   | Path loss exponent                 | $\alpha$  |
| Receiver overhead power                                                  | $P_O$     | Average receive (transmit) time in one cycle | $\bar{t}_{rx} \ (\bar{t}_{tx})$ |
| Power consumption of the power amplifier PA                               | $P_{PA}$   | Transmission/reception time slot duration | $\Delta t$ |
| RF transmit (received) power                                             | $P_T \ (P_R)$ | Receiver (transmitter) activity factors | $\beta_T \ (\beta_R)$ |

**TABLE I: Notation.**

In the proposed model, multiple devices (transmitters) send data to an aggregator (receiver), where each device is allocated a different time slot on the same frequency, i.e., TDMA. Therefore, $\beta_T = 1$ and $\beta_R = 0$ for transmitting devices, and their energy cost is proportional to $P_{txr}$. For the aggregator, $\beta_T = 0$ and $\beta_R = 1$, and its energy cost is proportional to $P_{txr}$. Let $E_R$ denote the average energy required by an aggregator. $E_T$ denotes the average energy required for all transmissions within the coverage of an aggregator node, i.e., if $N_a$ devices transmit to the aggregator, $E_T$ is the sum of their average transmission energies. Hence, for a transmission/reception time slot of duration $\Delta t$, aggregator and transmitter energy consumptions are

$$E_R = \Delta t (P_{LO} + P_O + P_{RX}), \quad E_T = \Delta t \ (P_{LO} + P_{PA}) N_a + \Delta t \cdot P_{TX},$$

where we assume $N_a$ devices sequentially transmit to the aggregator and $\bar{t}_{tx} \leq \bar{t}_{rx} = N_a \bar{t}_{tx} = \Delta t$, and we assume that the communication delay due to processing of the data is negligible. Hence, transmitted data can be received within the time slot allocated and the aggregator can decode the received data. We also assume that the transmissions and receptions are synchronized. $E_R$ is mainly determined by the reception duration, while $E_T$ depends on the energy consumption of the PA, $\bar{t}_{tx}$ and $N_a$. Furthermore, whenever a device is in sleep mode, its transmitter and receiver modules are not active but other components are still consuming energy.

We will use this transceiver model and the data aggregation strategy of the devices described in this section for formulating our energy optimization problems in Sects. III and IV and analyzing the SIR coverage probability in Sect. V.
III. SINGLE-STAGE ENERGY DENSITY OPTIMIZATION

The main focus of this section is to model the average total energy density for a single-stage data aggregation scheme, which is a single hop energy model incorporating the data aggregation and transceiver model described in Sect. II. This model paves the way for understanding the multi-stage energy model to be discussed in Sect. IV.

For the single-stage model, recall that initially a density $\lambda_a < \lambda/2$ of the devices will be independently selected as aggregators, and a density of $\lambda_u = \lambda - \lambda_a$ will be the transmitters. Then, the set of aggregators, $\Psi_a$, collects the data from the remaining devices, $\Psi_u$, based on nearest aggregator association, and each aggregator might have multiple devices assigned to it. However, multiple transmissions to an aggregator at a particular time slot are not allowed, hence the devices are scheduled based on TDMA. Once all devices complete their transmissions, the aggregators also incorporate their payloads, and then transmit the whole data to the BS.

We assume perfect channel inversion power control, under which the received power at the BS from any device is unity. Then, the average total uplink power is given by Theorem [I]

**Theorem 1.** The mean total uplink power of the devices in the Voronoi cell of the typical aggregator is given by

$$P(\lambda_a) = \frac{\lambda_u \alpha/2}{\lambda_a^{1+\alpha/2} \pi^{\alpha/2}} \Gamma\left(\frac{\alpha}{2}\right).$$  \hspace{1cm} (3)

*Proof:* Proof is given in Appendix A.

Note that (3) assumes the receiver power is unity, and it needs to be scaled to capture the actual uplink power of the transmission model. The mean total uplink power so that the received power is $P_R$ is found by incorporating the PA efficiency $\eta$ and scaling the result in (3) as follows

$$P(\lambda_a) = \frac{P_R}{\eta E[N_a]} \frac{\lambda_u \alpha/2}{\lambda_a^{1+\alpha/2} \pi^{\alpha/2}} \Gamma\left(\frac{\alpha}{2}\right),$$  \hspace{1cm} (4)

where $N_a$ is a random variable that denotes the number of devices within the Voronoi cell of a typical aggregator and its average can be obtained as

$$E[N_a] = E[\Psi_a] \sum_n 1_{\tilde{X}_n \in V_0} = 2\pi \lambda_u \int_{r>0} e^{-\lambda_a \pi r^2} r \, dr = \frac{\lambda_u}{\lambda_a},$$

where $E[\Psi_a]$ is the expectation with respect to the Palm probability conditioned on $0 \in \Psi_a$ [15].
Definition 1. **Average total energy in a Voronoi cell of an aggregator.** This is the average total energy consumption due to the transmissions within the coverage of the typical aggregator for a fixed payload per device. The average total energy in a typical cell is given by
\[
\mathcal{E}_V(\lambda_a) = E_R(\lambda_a) + E_T(\lambda_a). \tag{5}
\]

Definition 2. **Average total energy density.** This is the average total energy consumption per unit area per unit time for a fixed payload per device, and can be found by scaling the average energy density per Voronoi cell by the number of Voronoi cells per area. The average number of Voronoi cells per unit area is given by
\[
\mathbb{E}[N_v] = \frac{\lambda}{\lambda_u/\lambda_a + 1} = \lambda_a.
\]

The dissipated energy density, i.e., the energy consumption per unit area, is defined as
\[
\mathcal{E}(\lambda_a) = \lambda_a \mathcal{E}_V(\lambda_a) = \lambda_a (E_R(\lambda_a) + E_T(\lambda_a)). \tag{6}
\]

Our objective is to find the best \(\lambda_a < \lambda_u\) that minimizes the total energy consumption of both the transmitters and the aggregators per unit area. As \(\lambda_a\) increases, the total number of aggregators and their total energy consumption to receive data will increase, and as \(\lambda_a\) decreases, the distance between the aggregators and the energy requirement of the typical transmitter will increase. Hence, the optimal value of \(\lambda_a\) that minimizes the total energy per unit area is of interest. We can modify (2) as
\[
E_R(\lambda_a) = \Delta t (P_{LO} + P_O + P_{RX}) \, , \quad E_T(\lambda_a) = \bar{t}_{tx} \mathbb{E}[N_a] P_{LO} + \mathbb{E}[N_a] P_{TX} + P(\lambda_a), \tag{7}
\]
where we recall \(\bar{t}_{tx} = \bar{t}_{tx} \mathbb{E}[N_a] = \Delta t\).

Any aggregator device aggregates data from multiple devices. Hence, the average number of devices per aggregator, i.e., \(\mathbb{E}[N_a] = \frac{1+\gamma}{\gamma}\), is always greater than 1. Thus, the optimal fraction of aggregators that minimizes the overall energy density should satisfy \(\gamma < 0.5\).

IV. **Multi-Stage Energy Density Optimization**

We extend the model in Sect. III to multiple stages where the aggregation process is repeated \(K > 1\) times before the data is eventually transmitted to the BS. Let \(E_{R,k}(\lambda_a(k))\) and \(E_{T,k}(\lambda_a(k))\) be the average energies required for reception and transmission within the coverage of an aggregator node and \(\lambda_a(k)\) denotes the density of aggregators at stage \(k\). The average total energy consumption density at stage \(k\) is
\[
\mathcal{E}_k(\lambda_a(k)) = \lambda_a(k)(E_{R,k}(\lambda_a(k)) + E_{T,k}(\lambda_a(k))). \tag{8}
\]
The initial total density of the devices is given by \( \lambda \). The initial process with density \( \lambda \) is independently thinned to obtain the set of aggregator processes with density \( \lambda_a(k) \) for stage \( k \) to form a disjoint set of aggregators for each stage. The density of aggregators at stage \( k \) is denoted by \( \lambda_a(k) = \lambda \gamma^k \), for \( k \in \{1, \ldots, K - 1\} \). The total density of active devices that can be either transmitter and receiver at stage \( k \) is given by

\[
\lambda(k) = \lambda_a(k) + \lambda_a(k).
\]  

(9)

The transmitter device density of the first stage is found by subtracting the total density of aggregators from the initial density of the device process as

\[
\lambda_u(1) = \lambda - \sum_{k=1}^{K-1} \lambda(a(k)).
\]

The aggregation process can be extended to \( k > 2 \) stages in a similar manner by letting \( \lambda_u(k) = \lambda_a(k-1) \) for \( k \geq 2 \).

Generalizing (7), the average aggregator and transmitter energy consumptions per unit area at stage \( k \) are given by

\[
E_{R,k}(\lambda_a(k)) = \Delta t(k) P_{cov}(k-1) (P_{LO} + P_O + P_{RX})
\]

\[
E_{T,k}(\lambda_a(k)) = \bar{t}_{tx}(k) P_{cov}(k-1) \left[ \mathbb{E} [N_a(k)]^2 P_{LO} + \mathbb{E} [N_a(k)] P_{TX} + P(\lambda_a(k)) \right],
\]

(11)

where \( \Delta t(k) = \bar{t}_{tx}(k) \mathbb{E} [N_a(k)] \), and \( P_{cov}(k) \) is the joint SIR coverage probability up to and including stage \( k \) at a given threshold. SIR coverage probability gives the fraction of successful transmissions with respect to an SIR threshold, and the actual average energy at stage \( k \) is determined by the fraction of successful transmissions up to stage \( k \), i.e., the joint SIR coverage \( P_{cov}(k-1) \). Hence, the current stage energy consumption scales with the fraction of the successful transmissions. For the devices in outage, no additional energy expenditure is incurred in the current and consecutive stages. In this paper, we interchangeably use \( \lambda_a \) (or \( \lambda_a(k) \)), \( \lambda_u \) (or \( \lambda_u(k) \)), \( N_a \) (or \( N_a(k) \)), and \( \Delta t(k) \) (or \( \Delta t \)) when the stage index is insignificant or clear from context.

A. An Upper Bound on the Energy Density

Let \( \lambda_a(K) = [\lambda_a(1) \ \lambda_a(2) \ \cdots \ \lambda_a(K)]^T \). Following the aggregation procedure described, an upper bound for the average total energy consumption density over \( K \) stages can be found
by ignoring the SIR coverage probability as follows

\[
E_U(\lambda_a(K)) = \sum_{k=1}^{K} E_k(\lambda_a(k)) \bigg|_{P_{\text{cov}}(k)=1, \forall k}
\]

\[
= \sum_{k=1}^{K} \bar{t}_{tx}(k) \left[ \lambda_a(k) P_C + \lambda_a(k) P(\lambda_a(k)) + \frac{\lambda_u(k)^2}{\lambda_a(k)} P_{LO} \right],
\]

(12)

where \( P_C = P_{TX} + P_{RX} + P_{LO} + P_O \). We simplify (12) by letting \( \bar{t}_{tx}(1) = 1 \) as this scaling does not affect the result of the optimization formulation.

The slot duration at stage \( k \) is \( \Delta t(k) = \bar{t}_{tx}(k) \) and the duration that the transmit block is on is \( \bar{t}_{tx}(k) \). The relation between the average transmit times are as \( \bar{t}_{tx}(k+1) = \bar{t}_{tx}(k) \mathbb{E} [N_a(k)] \). The transmit duration at slot \( k \) is proportional to the mean total number of bits to be transmitted as

\[
\bar{t}_{tx}(k) \propto \mathbb{E} \left[ \prod_{i=1}^{k-1} N_a(i) \right].
\]

(13)

Note that \( N_a(i) \)'s for \( i \in \{1, \ldots, k-1\} \) are dependent random variables because of the correlation between the Voronoi cell areas of the aggregators at subsequent stages [16]. However, for tractability, we assume independence among \( N_a(i) \)'s, and modify (13) as

\[
\bar{t}_{tx}(k) \propto \prod_{i=1}^{k-1} \mathbb{E} [N_a(i)] = \prod_{i=1}^{k-1} \lambda_u(i)/\lambda_a(i),
\]

(14)

where the independence assumption is validated through simulations in Sect. VII.

**Remark 1.** If there are \( K \) stages in total, at the last stage, all the aggregators of the previous stage (stage \( K-1 \)) with density \( \lambda_u(K-1) = \lambda_a(K-1) \) are transmitting to the BSs with density \( \lambda_{BS} \). Thus, at stage \( K \), \( \lambda_a(K-1) = \lambda_{BS} \). Therefore, the fraction of aggregators at the last stage is

\[
\gamma(K) = \lambda_{BS} \left( \frac{\lambda_u}{\lambda_a} \right)^{K-1} + \lambda_{BS} \right)^{-1},
\]

(15)

which is incorporated into the numerical analysis to find the minimizing aggregator fraction.

**B. Average Energy Density with an SIR Coverage Constraint**

This section mainly concentrates on the average energy consumption in the case of SIR outage. The average energy density of the proposed aggregator model is found by incorporating the SIR coverage probability of the typical transmitting device. If the received SIR at any stage \( k \) cannot exceed the threshold \( T \), then the transmission is unsuccessful. Therefore, devices with a density of \( \mathcal{P}_k(T) \lambda_u(k) \) will be successful at stage \( k \), i.e., out of \( N_a(k) \) transmitting devices served per
aggregator, on average, only $P_k(T)E[N_a(k)]$ of them will successfully transmit their data, and the rest of the devices will not meet the minimum SIR requirement despite consuming energy. Therefore, the performance of the current stage depends on the previous stages.

Incorporating the SIR coverage, the average total energy density in (12) is modified as

$$E(\lambda_a(K)) = \sum_{k=1}^{K} P_{cov}(k-1)\bar{t}_{tx}(k)\left[\lambda_a(k)P_C + \lambda_a(k)P(\lambda_a(k)) + \frac{\lambda_a(k)^2}{\lambda_a(k)}P_{LO}\right],$$  

(16)

where $P_{cov}(k)$ denotes the joint SIR coverage probability for $k$ stages and $P_{cov}(0) = 1$, where we do not consider the effect of SIR coverage probability as all the devices at stage $k = 1$ transmit even though they might not be successful. In Sect. [V] we consider sequential and parallel transmission modes with the assumption of inter-stage independence, i.e., the joint SIR coverage becomes $P_{cov}(k) = \prod_{i=1}^{k} P_i(T)$ for $k > 1$.

Next, we formulate the energy density optimization problem assuming that each device has unit SIR coverage probability. Note that $P_{cov}(k)$ for $k \in \{1, \ldots, K\}$ depends on the communication scheme, the network model parameters, such as device density, fading and shadowing distribution, and the path loss, and is determined independently from the energy model. This section mainly concentrates on the energy model, while the SIR coverage and rate models – and hence, $P_{cov}(k)$’s for various transmission schemes – will be discussed extensively in Sects. [V][VI]. In Sect. [VII], we combine the energy results of this section with the coverage characteristics for the successive and parallel transmission modes, to evaluate the energy efficiencies of the proposed models.

**C. Energy Density Optimization Problem**

The following optimization problem minimizes the average total energy density over $K$ stages:

$$\min_{\lambda_a(K)} E(\lambda_a(K))$$

s.t.  

$$\lambda_a(k) \leq \lambda_a(k-1), \quad k \in \{2, \ldots, K-1\}$$

$$\lambda_a(k) = \lambda_a(k-1), \quad k \in \{2, \ldots, K\}$$

$$\bar{t}_{tx}(k) = \bar{t}_{tx}(k-1)E[N_a(k-1)], \quad k \in \{2, \ldots, K\}$$

$$\lambda_a(1) = \lambda - \lambda \sum_{k=1}^{K-1} \gamma^k$$

$$\lambda_a(K) = \lambda_{BS},$$

(17)

where $\lambda_a(k) = \lambda \gamma^k$ for $k \in \{1, \ldots, K-1\}$, i.e., fixed fraction of aggregators for all stages.
TABLE II: Design parameters.

| $k$ | $\lambda_u(k)$ | $\lambda(1 - \bar{\gamma}_K)$ | $\lambda \gamma^{k-1}$ | $\lambda \gamma^{K-1}$ |
|-----|----------------|-------------------------------|------------------------|------------------------|
| 1   | $\lambda(1 - \bar{\gamma}_K)$ | $\lambda \gamma^{k-1}$ | $\lambda \gamma^{K-1}$ |
| $2 \leq k \leq K-1$ | $\lambda \gamma$ | $\lambda \gamma^{k}$ | $\lambda_{BS}$ |
| $k = K$ | $\mathbb{E}[N_u(k)]$ | $(1 - \bar{\gamma}_K)/\gamma$ | $\gamma^{-1}$ | $\lambda \gamma^{K-1}/\lambda_{BS}$ |
| $\bar{t}_{tx}(k)$ | 1 | $(1 - \bar{\gamma}_K)/\gamma^{k-1}$ | $(1 - \bar{\gamma}_K)/\gamma^{K-1}$ |

For the optimization formulation in (17), and for $K \geq 2$, we let $\bar{\gamma}_K = \sum_{k=1}^{K-1} \gamma^k$. The important design parameters of the total energy density are tabulated in Table II. Using (3) and the design parameters in Table II, the mean total uplink power for different stages is given by

$$P(\lambda_a(k)) = \begin{cases} \frac{(1 - \bar{\gamma}_K) \gamma^{(1+\alpha/2)a/2}}{(\pi \lambda)^{\alpha/2}} \Gamma \left( \frac{\alpha}{2} \right), & k = 1 \\ \frac{\gamma^{(1+ka/2)a/2}}{(\pi \lambda)^{\alpha/2}} \Gamma \left( \frac{\alpha}{2} \right), & 2 \leq k \leq K-1 \\ \frac{\lambda \gamma^{k-1} \alpha/2}{\lambda_{BS}^{1+\alpha/2}} \Gamma \left( \frac{\alpha}{2} \right), & k = K. \end{cases} \quad (18)$$

**Proposition 1.** Combining (16) with the constraints in (17) for unit SIR coverage probability, we define $c_k(\gamma)$ to be the total energy density at stage $k \in \{1, \ldots, K\}$ for $\lambda = 1$. Hence, $c_k(\gamma)$ is given by

$$c_k(\gamma) = \begin{cases} (1 - \bar{\gamma}_K) P_C + \gamma P(\gamma \lambda) + \frac{(1 - \bar{\gamma}_K)^2}{\gamma} P_{LO}, & k = 1 \\ (1 - \bar{\gamma}_K) \left[ P_C + \gamma P(\gamma^k \lambda) + \gamma^{-1} P_{LO} \right], & 2 \leq k \leq K-1 \\ (1 - \bar{\gamma}_K) \left[ P_C + \frac{\lambda_{BS}}{\lambda \gamma^k} P(\lambda_{BS}) + \frac{\lambda \gamma^{k-1}}{\lambda_{BS}} P_{LO} \right], & k = K. \end{cases} \quad (19)$$

Incorporating the constraints, and using (19), the objective function of (17) is equivalent to

$$E(\lambda_a(K)) = \lambda \sum_{k=1}^{K} \mathcal{P}_{cov}(k-1)c_k(\gamma). \quad (20)$$

Similarly, the energy upper bound $E_U(\lambda_a(K))$ can be minimized by evaluating (20) at $\mathcal{P}_{cov}(k) = 1 \forall k$, and taking its derivative with respect to $\gamma$.

**Proposition 2.** $c_k(\gamma)$ is an increasing function of $k$ for $1 \leq k \leq K-1$, and $c_K(\gamma)$ is a decreasing function of total number of stages $K$. 

Proof: Note that $\gamma^{-k}$ increases with $k$, and $P(\gamma^k \lambda)$ is an increasing function of $k$ for $2 \leq k \leq K - 1$. Using (19), it is easy to show that $c_k(\gamma) < c_{k+1}(\gamma)$ for $1 \leq k \leq K - 2$ for fixed $\gamma$. Note also that $P(\lambda_{\text{BS}})$ decreases, and $P(\lambda_{\text{BS}})/\gamma^{K-1}$ does not change with $K$, but $\gamma^{K-1} P_{\text{LO}}$ decreases with $K$, hence it is trivial to show that $c_K(\gamma)$ is decreasing in $K$.

Proposition 2 implies that the energy density is higher for higher order stages. On the other hand, as the number of stages increases, the energy density at the last stage decreases. The number of stages cannot be increased arbitrarily. Next, to capture the energy consumption tradeoff between the stages, we investigate the effect of $\gamma$ on the total energy density.

Remark 2. Using (19), we can easily observe that $c_k(\gamma)$ is decreasing in $\gamma$ for $1 \leq k \leq K - 1$. Note also that $c_K(\gamma)$ is given by the following sum

$$(1 - \bar{\gamma}_K) \left[ P_C + \frac{\lambda_{\text{BS}}}{\gamma^{K-1}} P(\lambda_{\text{BS}}) \right] + (1 - \bar{\gamma}_K) \frac{\lambda^{K-1}}{\lambda_{\text{BS}}} P_{\text{LO}}.$$  

We observe that the first term decays with $\gamma$ as $1 - \bar{\gamma}_K$ decreases with $\gamma$, and $P(\lambda_{\text{BS}})/\gamma^{K-1}$ is invariant to $\gamma$. The second term is not strictly a decreasing function of $\gamma$. On the other hand, it is increasing, for $\gamma \in (0, a)$, where $a \in (0, 0.5)$. As the second term is scaled by $\lambda/\lambda_{\text{BS}}$, which is typically very high for the operating regime of M2M, it determines the trend of total energy.

Proposition 3. Let $\gamma^*(K)$ be the minimizer of the energy optimization problem in (20) with the assumption of $P_{\text{cov}}(k) = 1 \forall k$, and $c_k(\gamma)$ is increasing in $\gamma$. Then, $\gamma^*(K)$ is a decreasing function of the total number of stages $K$.

Proof: See Appendix B.

This section concentrates on the energy efficiency of M2M assuming perfect coverage. We now study the SIR coverage of the hierarchical M2M model for different transmission schemes.

V. SIR COVERAGE PROBABILITY

The data aggregation model in Sects. III and IV does not incorporate the fact that the transmissions are not always successful. In this section, we generalize this to a coverage-based model, assuming that the transmission is successful if the SIR of the device is above a threshold. We derive the probability of coverage in the uplink for the proposed data aggregation model.

---

Note that it is trivial to extend this result for $P_{\text{cov}}(k) < 1$ because $c_k(\gamma)$’s are invariant to the coverage probability.
in Sects. III and IV. We assume that the tagged aggregator and tagged devices experience both Rayleigh fading and lognormal shadowing. The transmit power at the typical node a distance \( r \) from its aggregator is \( S_r^\alpha \), and the received power is \( g \), where \( S \) is lognormal distributed large-scale fading parameter denoted as \( S \sim \ln \mathcal{N}(m, v) \), and the random variable \( g \) follows an exponential distribution with mean \( 1/\mu \), which we denote as \( g \sim \exp(\mu) \). In this case, the average received power \( P_R \) is \( 1/\mu \).

Orthogonal access is assumed in the uplink and at any given resource block, there is at most one device transmitting in each cell. Let \( \Psi_u \) be the point process denoting the location of devices transmitting on the same resource as the typical device. The uplink SIR of the typical device \( x \in \Psi_u \) on a given resource block is

\[
\text{SIR} = \frac{g}{\sum_{z \in \Psi_u \setminus \{x\}} g_z S_z R_z^\alpha S_z^{-1} D_z^{-\alpha}},
\]

(21)

where \( R_z \) and \( D_z \) denote the distance between the transmitter aggregator pair and the distance between the interferer and the typical aggregator, respectively. The random variable \( g_z \sim g \) is the small-scale iid fading parameter due to interferer \( z \), and \( S_z \) denotes the large-scale fading, i.e., shadowing, between the device \( z \) and its nearest aggregator, and \( S_{z,0} \) denotes the shadowing between the device \( z \) and the aggregator of the typical device located at the origin.

The uplink SIR coverage of the proposed system model is given by the following Lemma.

**Lemma 1.** The uplink SIR coverage \([17]\): With full power control and with minimum average path loss association\(^4\) the uplink SIR coverage is given by\(^5\)

\[
\mathcal{P}(T) = \mathbb{P}(\text{SIR} > T) \approx \exp \left( -\frac{2T}{\alpha - 2 C_{\alpha}(T)} \right),
\]

(22)

where \( C_{\alpha}(T) = {}_2F_1 \left( 1, 1 - \frac{2}{\alpha}; 2 - \frac{2}{\alpha}; -T \right) \), and \( {}_2F_1 \) is the Gauss-Hypergeometric function.

A. Coverage Probability and Number of Stages

The number of multi-hop stages \( K \) is mainly determined by the SIR coverage and the distance coverage, which we define as the probability that the distance between two neighboring devices

\(^4\)In “minimum average path loss association”, a device associates to an aggregator with minimum path loss averaged over the small-scale fading, i.e., the aggregator has minimum \( S_i R_i^\alpha \) product among all aggregators.

\(^5\)Note that the coverage results do not depend on the large-scale fading.
is below a threshold. In this section, we give lower and upper bounds on $K$ using these coverage concepts.

**Assumption 1. Interstage independence.** The proposed hierarchical aggregation model introduces dependence among the stages of multi-hop communication since each subsequent stage is generated by the thinning of the previous stage. For analytical tractability, we assume that the multi-hop stages are independent of each other. Hence, the transmission is successful if and only if all the individual stages are successful. The success probability over $K$ stages is

$$
\mathcal{P}_{cov}(K) = \mathbb{P}(\text{SIR}_1 > T, \ldots, \text{SIR}_K > T) = \prod_{k=1}^{K} \mathcal{P}_k(T),
$$

where $\mathcal{P}_k(T)$ denotes the coverage probability at stage $k$. With full channel inversion and minimum average path loss association, the uplink SIR coverage is independent of the infrastructure density as given in (22). Since $\mathcal{P}_k(T)$ is also independent of the device density, and only depends on the threshold $T$ and path loss exponent $\alpha$, it is identical for all stages, and denoted by $\mathcal{P}(T)$.

**Lemma 2. An Upper Bound on $K$.** Given a minimum probability of coverage requirement $\mathcal{P}_{cov}(K) > 1 - \varepsilon$ and an SIR threshold $T$, the number of stages is upper bounded by

$$
K_U = \left\lceil \log \left( \frac{1}{1 - (1 - \varepsilon)} \right) \frac{(\alpha - 2) - 2}{2} \right\rceil. \quad (23)
$$

**Proof:** The upper bound is obtained by combining (22) with the condition $\mathcal{P}_{cov}(K) > 1 - \varepsilon$.

**Corollary 1. Low SIR regime.** For $\alpha = 4$ and $T < 1$, the product $TC_4(T)$ is simplified to

$$
TC_4(T) = -\frac{1}{2} \sum_{n=1}^{\infty} \frac{(-T)^n}{(n - \frac{1}{2})} = \sqrt{T} \tan^{-1}(\sqrt{T}), \quad T < 1. \quad (24)
$$

Hence, for $\alpha = 4$, the number of stages is upper bounded by $K_U = \left\lceil \frac{\log(1/(1-\varepsilon))}{\sqrt{T} \tan^{-1}(\sqrt{T})} \right\rceil$, for $T < 1$.

In addition to the SIR outage, since the devices are randomly deployed, any device will be in outage when its neighboring devices are outside its maximum transmission range. Thus, we also aim to investigate the minimum number of required stages given a distance outage constraint.

A lower bound on the optimal number of multi-hop stages is given by the following Lemma.

---

6This assumption is required for the transmission modes described in detail in Sect. [VI] and is validated in Sect. [VII].
Lemma 3. A Lower Bound on $K$. The number of stages is lower bounded by

$$K_L = \left\lceil L(\lambda_a) \mathbb{E}\left[\frac{1}{N_a}\left(\frac{P_R}{P_{T,\max}}\right)^{\frac{1}{\alpha}}\right] \right\rceil,$$

where $L(\lambda_a)$ denotes the mean total length of the connections, $N_a$ is the number of devices in the Voronoi cell of a typical aggregator, and $P_{T,\max}$ is the maximum transmit power.

Proof: Given a maximum transmitter power constraint, $P_{T,\max}$ for each device, the maximum transmission range is given by $(P_{T,\max}/P_R)^{1/\alpha}$. Dividing $L(\lambda_a)$ by $N_a$ and taking its expectation with respect to the distribution of $N_a$, we obtain the mean length of connections $L(\lambda_a)$ and dividing this ratio by the maximum transmission range, we obtain the desired result.

Lemma 4. The mean total length of the connections in the Voronoi cell of an aggregator equals

$$L(\lambda_a) = \frac{\lambda_u}{2\lambda_a^{3/2}}.$$

Proof: The idea of the proof is similar to the proof of Theorem \[\text{T}\] where $f(x) = |x|$. Therefore,

$$L(\lambda_a) = \lambda_u \int_{\mathbb{R}^2} f(x) e^{-\lambda_a \|x\|^2} \, dx = 2\pi \lambda_u \int_0^\infty r^2 e^{-\lambda_a \pi r^2} \, dr = \frac{\lambda_u}{2\lambda_a^{3/2}}.$$

The following Lemma provides a lower bound for $K_L$ that is based on the mean total length of the connections to the typical aggregator and the fraction of aggregators.

Lemma 5. $K_L$ is lower bounded by

$$\left[ \frac{1}{\mathbb{E}[N_a]} \frac{(1 - \gamma)}{2\lambda_a^{1/2}} \left( \frac{P_R}{P_{T,\max}} \right)^{1/\alpha} \right] \leq K_L \leq K_U.$$

Proof: Using (26), we can lower bound $K_L$ as

$$K_L \geq \left[ \frac{1}{\mathbb{E}[N_a]} \frac{\lambda_u}{2\lambda_a^{3/2}} \left( \frac{P_R}{P_{T,\max}} \right)^{1/\alpha} \right],$$

where we use the convexity of $1/N_a$, i.e., $\mathbb{E}\left[\frac{1}{N_a}\right] \geq \frac{1}{\mathbb{E}[N_a]}$. Noting that $\lambda_u$ and $\lambda_a$ are functions of $K$, we have the following relation:

$$L(\lambda_a) = \frac{\lambda_u}{2\lambda_a^{3/2}} = \frac{\lambda_u^{K-1}(1 - \gamma)}{2(\lambda_a^{K})^{3/2}} = \frac{(1 - \gamma)}{2\lambda_a^{1/2} \lambda_a^{K/2} + 1}.$$

$L(\lambda_a)$ is increasing in $K$ since $\gamma \leq 0.5$. Plugging (29) into (28), we obtain the lower bound.

\[\text{For tractability, we take expectation over a PPP first to find } L(a), \text{ and then } \mathbb{E}\left[\frac{1}{N_a}\right] \text{ assuming independence.}\]
VI. TRANSMISSION RATE MODELS

For an interference limited network, the rate of the typical device is given by

\[ \text{Rate} = \frac{W}{N_a} \log (1 + \text{SIR}) \]

where \( W \) is the total bandwidth of the communication channel, and \( N_a \) is the load at the typical aggregator. The average number of devices served by the typical aggregator is denoted by \( \mathbb{E}[N_a] = \frac{\lambda_u}{\lambda_a} \). Rate coverage is defined as rate exceeding a given threshold, i.e.,

\[ P(\text{Rate} > \rho) = \sum_{l=0}^{\infty} P(SIR > 2^{\frac{\rho W}{N_a}} - 1|N_a = l) \mathbb{P}_{N_a}(l), \]

where \( \mathbb{P}_{N_a}(l) \) is the probability mass function (PMF) of \( N_a \), and is given by the following Lemma.

**Lemma 6.** The PMF of the number of devices served per aggregator \( N_a \) is given by

\[ \mathbb{P}_{N_a}(l) = \frac{G_{N_a}^{(l)}(0)}{l!} = \frac{3.5^{3.5} (\lambda_u/\lambda_a)^l \Gamma(3.5 + l)}{(3.5 + (\lambda_u/\lambda_a))^{3.5+l} \Gamma(3.5) \Gamma(l+1)}, \]

where \( \lambda_u(k) \) and \( \lambda_a(k) \) for \( k \geq 1 \) are given in Table II.

**Proof:** Normalized distribution function of Voronoi cell areas in 2D can be modeled by

\[ f(y) = \frac{3.5^{3.5}}{\Gamma(3.5)} y^{5/2} \exp \left( -\frac{7}{2} y \right). \]

The densities of the transmitting devices and the aggregators are given by \( \lambda_u \) and \( \lambda_a \), respectively. Hence, the probability generating function (PGF) of the devices in the random area \( y \) is

\[ G_{N_a}(z) = \mathbb{E}[z^{N_a}] = \mathbb{E}[\exp ((\lambda_u/\lambda_a) y(z - 1))], \]

where conditioned on \( y \), the PGF is of a Poisson random variable \( N_a \) with mean \( (\lambda_u/\lambda_a) y \).

Combining (32) and (33), we obtain the PGF of \( N_a \) as

\[ G_{N_a}(z) = \int_0^\infty \exp \left( \frac{\lambda_u}{\lambda_a} y(z - 1) \right) \frac{3.5^{3.5}}{\Gamma(3.5)} y^{5/2} \exp \left( -\frac{7}{2} y \right) \, dy = \frac{3.5^{3.5}}{(3.5 + (\lambda_u/\lambda_a)(1-z))^{3.5}}. \]

Then, the PMF of \( N_a \) is recovered by taking derivatives of \( G \).

The key assumption in our analysis is that there is one active device per resource block in each Voronoi cell. Using (31) and \( \lambda_u(k)/\lambda_a(k) = \mathbb{E}[N_a(k)] \), the probability of not finding any device in the Voronoi cell of the typical aggregator at stage \( k \) is

\[ \mathbb{P}_{N_a}(0) = G_{N_a}(0) = 3.5^{3.5} (3.5 + \mathbb{E}[N_a(k)])^{-3.5}. \]

The number of devices \( N_a \) served by the typical aggregator is mainly determined by the aggregator fraction \( \gamma \). When \( \gamma \) is high, i.e., \( \mathbb{E}[N_a(k)] \) is low, the probability that there is no
transmitting device within the Voronoi cell of the typical aggregator is not negligible. Let $p_{\text{th}}(k)$ be the probability that there is at least a device in the Voronoi cell of the typical aggregator in the $k^{\text{th}}$ stage. Therefore, the interference field of stage $k$ is thinned by $p_{\text{th}}(k)$, and the effective density of the interfering devices at stage $k$ is $p_{\text{th}}(k)\lambda_u(k)$. Using (35), $p_{\text{th}}(k) = 1 - 3.5^{3.5}(3.5 + \mathbb{E}[N_u(k)])^{-3.5}$. From Table II, we can determine $p_{\text{th}}(k)$, which has the same value for $2 \leq k \leq K - 1$, and different values for $k = \{1, K\}$.

The uplink SIR coverage for successive stages in (22) is independent of the device density and $p_{\text{th}}(k)$. However, the rate coverage results for the parallel mode depend on the device density, and the interference field of stage $k$ is thinned by $p_{\text{th}}(k)$, as described in Sect. VI-B.

**Corollary 2.** The average rate of the typical device is

$$\bar{R}(\alpha) = \int_{t>0} P(\text{Rate} > t) \, dt \approx \sum_{l=0}^{\infty} P_{N_u}(l) \int_{t>0} \exp \left( -\frac{2\alpha^2}{\alpha^2 - 1} C_{\alpha}(2^{\frac{\alpha^2}{\alpha^2 - 1}} - 1) \right) \, dt. \tag{36}$$

We consider two main transmission protocols, namely i) a successive transmission protocol where the stages are activated sequentially, i.e., a half-duplex sequential mode, and ii) a parallel transmission mode, which is either a full-duplex protocol where all stages are simultaneously active, or a half-duplex protocol with alternating active stages. We investigate their energy efficiencies and provide numerical comparisons for the rate-energy-delay tradeoffs in Sect. VII.

A. Rate Distribution for Successive Stages

In this mode, each transmission cycle consists of the stages operating in succession. Stages may not be repeated before a cycle is completed. This mode may provide low rate, but it has low interference since multiple stages are not active simultaneously. Let $\mathcal{K} = \{1, \ldots, K\}$ denote the set of stages and $\mathcal{R} = \{\text{Rate}_1, \ldots, \text{Rate}_K\}$ be the set of transmission rates at each stage. The transmission rate in successive mode is given by

$$R_S = \frac{1}{K} \min_{k \in \mathcal{K}} \text{Rate}_k. \tag{36}$$

**Remark 3.** The hierarchical levels are not independent from each other, and hence, it is not tractable to analyze the joint rate distribution for successive stages. Instead, we define the rate outage as in (37) where transmission rates are assumed to be independent. Without tracking the path of the bits (payload) transmitted, we only consider if the hierarchical transmission process
is successful. Transmission from any device is successful if its payload is delivered to the BS at the end of $K$ stages. A performance comparison for the independence assumption and the actual simulation results is given in Sect. VII.

With the independence assumption in Remark 3, the rate coverage for successive stages is

$$\mathbb{P}(R_S > \rho) = \prod_{k \in \mathcal{K}} \mathbb{P}(\text{Rate}_k > K\rho) \approx \prod_{k \in \mathcal{K}} \sum_{l=0}^{\infty} \exp \left( - \frac{2^{\frac{K_{al}}{W}} - 1}{\alpha - 1} C_\alpha (2^{\frac{K_{al}}{W}} - 1) \right) \mathbb{P}_{N_a}(k)(l).$$

**B. Rate Distribution for Full-Duplex Parallel Stages**

In full-duplex parallel mode, transmissions are not interrupted during a transmission cycle unlike the successive transmission mode. All multi-hop stages operate in parallel, the 1st stage devices only transmit, and the rest of the devices both transmit and aggregate simultaneously, during all stages of the multi-hop transmission. Therefore, this mode is a full-duplex model.

Due to the simultaneous transmissions at all levels of the hierarchical model, the interference at each stage is due to i) the interferers of that stage, i.e., the intra-stage interference, and ii) the remaining transmitting devices of the other stages, i.e., the inter-stage interference. The hierarchical levels are determined in the same manner similar to the successive mode, and also dependent in this mode, and hence, the inter-stage interference is correlated. Although full-duplex parallel mode offers high transmission rate compared to successive mode, it has higher interference since all the stages are active. The intra-stage interference in the parallel mode can be obtained in the similar manner as in the successive mode. The following lemma provides the analytical expressions for the Laplace transforms of intra-stage and inter-stage interference.

**Lemma 7.** Given the active transmission stage $k$, the Laplace Transforms of the intra-stage interference and the inter-stage interference are given as follows:

(a) The Laplace transform of the intra-stage interference at stage $k$ is

$$\mathcal{L}_{I_k}(s) \approx \exp \left( - \frac{2s}{\alpha - 2} \frac{p_{th}(k)\lambda_u(k)}{\sum_{j=1}^{K} p_{th}(j)\lambda_u(j)} C_\alpha(s) \right).$$

(b) The Laplace transform of the total inter-stage interference from stages $\{ j \mid j \in \mathcal{K}, j \neq k \}$ is

$$\mathcal{L}_{I_{k\setminus}}(s) \approx \exp \left( - 2\pi \lambda^*_l(k) \int_{v>0} g_1(v, s, \alpha) g_2(v) \, dv \right).$$
where \( g_1(v, s, \alpha) = \int_{z > 0} \frac{z}{1+(sv\alpha z)^{-1}e^z} \, dz \), \( g_2(v) = \pi \lambda_0^u(k) e^{-\lambda_0^u(k)p v} \) and \( \lambda_0^u(k) = \sum_{j \neq k} \alpha \Psi_{th}(j) \lambda_u(j) \) is the total density of the transmitting devices from all stages excluding the active stage \( k \).

**Proof:** Part (a). Note that the intra-stage interference in the proposed model can be found using the fraction of active transmitting devices of stage \( k \), which is given by \( \frac{\alpha \Psi_{th}(k) \lambda_u(k)}{\sum_{j=1}^{\infty} \alpha \Psi_{th}(j) \lambda_u(j)} \) using the interference thinning approach. The analysis follows from [17] and its proof is skipped. Note that as the fraction of transmitters goes to unity, the intra-stage interference becomes independent of the device density as in the sequential model since the interference is due to only the intra-stage devices, and there is no inter-stage interference.

Part (b). Let \( I_{kc} \) be the total inter-stage interference at stage \( k \), and \( \Psi_{kc} = \bigcup_{j \neq k} \Psi_{u(j)} \) be the point process denoting the location of inter-stage devices transmitting on the same resource as the typical device. Note that the transmitter processes for all stages are determined by independent thinning of the initial device process \( \Psi \) modeled as PPP. Hence, \( \Psi_{kc} \) is the superposition of the inter-stage transmitter processes, and is a PPP with density \( \lambda_c^u(k) = \sum_{j \neq k} \alpha \Psi_{th}(j) \lambda_u(j) \), and is also independent of \( \Psi_{u(k)} \) since the random variables \( \{R_z\} \) are assumed independent [20]. Then, the Laplace transform of \( I_{kc} \), i.e., \( \mathcal{L}_{I_{kc}}(s) \), is given by

\[
\mathcal{L}_{I_{kc}}(s) = \mathbb{E} \left[ \exp \left( -s \sum_{z \in \Psi_{u,j}} g_z R_z^\alpha D_z^{-\alpha} \right) \right] = \mathbb{E} \left[ \prod_{z \in \Psi_{u,j}} \mathbb{E}_{g_z} \left[ \exp \left( -s g_z R_z^\alpha D_z^{-\alpha} \right) \right] \right]
\]

\[
\overset{(a)}{=} \mathbb{E} \left[ \prod_{z \in \Psi_{u,j}} \frac{1}{1+sR_z^\alpha D_z^{-\alpha}} \right] \overset{(b)}{=} \mathbb{E} \left[ \prod_{z \in \Psi_{u,j}} \mathbb{E}_{R_z} \left[ \frac{1}{1+sR_z^\alpha D_z^{-\alpha}} \right] \right]
\]

\[
\overset{(c)}{=} \exp \left( -2\pi \lambda_0^c(k) \int_{z > 0} \left( 1 - \mathbb{E}_{R_z} \left[ \frac{1}{1+sR_z^\alpha D_z^{-\alpha}} \right] \right) z \, dz \right) ,
\]

(40)

where (a) follows from the independence of \( g_z \), (b) follows from the independence of \( \{R_z\} \), and (c) follows from the definition of probability generating functional (PGFL) of the PPP [21].

**Assumption 2.** The actual distribution of \( R_z \) is very hard to characterize due to the randomness both in the area of the Voronoi cell of the aggregator and in the number of the devices it serves. Therefore, we approximate the distribution of \( R_z \) by the distance of a randomly chosen point in \( \mathbb{R}^2 \) to its closest aggregator and hence it can be approximated by a Rayleigh distribution [20]:

\[
f_{R_z}(r_z) = 2\pi \lambda_0^c(k) r_z \exp \left( -\lambda_0^c(k)p r_z^2 \right), \quad r_z \geq 0.
\]

(41)

Expanding the term \( \mathbb{E}_{R_z} \left[ \frac{1}{1+sR_z^\alpha D_z^{-\alpha}} \right] \) in (40) as an integral using the density in (41), a change
of variables \( r^2 = v \), and rearranging the order of integrals in (40), \( \mathcal{L}_{I_{k,c}}(s) \) is given by
\[
\mathcal{L}_{I_{k,c}}(s) = \exp \left( -2\pi \lambda_u^c(k) \int \left( \int_{z > 0} \frac{z}{1 + (sv^2)^{-\alpha}} \, dz \right) g_2(v) \, dv \right).
\]

**Corollary 3.** For \( \alpha = 4 \), the Laplace transform of the total inter-stage interference is given by
\[
\mathcal{L}_{I_{k,c}}(s) = \exp \left( -\frac{\pi \sqrt{s}}{2} \right) \forall k,
\]
and is independent of the device density. Thus, the inter-stage interference is the same for all \( k \).

**Lemma 8.** The uplink SIR coverage probability for the full-duplex parallel mode is given by
\[
\mathbb{P}(\text{SIR}_P > T) = \prod_{k \in \mathcal{K}} \mathbb{P}(\text{SIR}_k > T),
\]
where \( \mathbb{P}(\text{SIR}_k > T) \approx \exp \left( -\frac{2T}{\alpha - 2} \sum_{j=1}^{\mathcal{K}} \frac{p_{eh(j)} \lambda_u(j)}{p_{eh}(j) \lambda_u(j)} C_\alpha(T) \right) \mathcal{L}_{I_{k,c}}(T). \)

**Proof:** The proof is similar to that of Theorem I in [17]. However, instead of one serving stage, the serving stages change sequentially. Since the total inter-stage interference is independent from the intra-stage interference, its Laplace transform is incorporated as a multiplicative term.

**Lemma 9.** The rate coverage probability for the full-duplex parallel transmission mode is
\[
\mathbb{P}(R_P > \rho) = \prod_{k \in \mathcal{K}} \mathbb{P}(\text{Rate}_k > \rho) = \prod_{k \in \mathcal{K}} \sum_{l=0}^{\infty} \mathbb{P}(\text{SIR}_k > 2^{\frac{\rho l}{\alpha}} - 1 | N_a(k) = l) \mathbb{P}_{N_a(k)}(l).
\]

**Proof:** Proof follows from the combination of (43) with the definition of rate.

The rate coverage expression for the full-duplex parallel mode is validated in Sect. VII.

**Remark 4.** Although the full-duplex parallel transmission strategy is probably not feasible for M2M communication, it is helpful to have a comparison of the rate-energy tradeoffs of both models, and provided for completeness. The models described above, i.e., the sequential mode which is half-duplex by design, and the full-duplex parallel mode, are the two principle design schemes that mainly differ in terms of their total rate coverages and energy consumptions.

Next, we introduce the half-duplex parallel transmission, which is a feasible scheme for M2M.
C. Rate Distribution for Half-Duplex Parallel Stages

The full-duplex parallel mode can be transformed into a half-duplex communication scheme. In this mode, at a particular time slot, only the even or odd stages are active. Analysis of this model is quite similar to the parallel mode analysis, using only the active stages when calculating the inter-stage interference. The SIR coverage of the half-duplex mode can be characterized in a fashion similar to full-duplex mode SIR coverage in (43), and is given by the following lemma.

**Lemma 10.** The uplink SIR coverage probability for the half-duplex parallel mode is given by

\[
P(\text{SIR}_P > T) = \prod_{k \in \mathcal{K}, \text{odd}} P(\text{SIR}_k > T),
\]

where \( P(\text{SIR}_k > T) \approx \exp\left(-\frac{2T}{\alpha-2} \sum_{j \text{ odd}} p_{th}(j) \lambda_u(j) C_\alpha(T) \right) \mathcal{L}_{k,c}(T). \)

For the half-duplex mode, the rate coverage expression in (44) changes in accordance with (45), which can be found by following the steps in Lemma 9. Since only half of the stages are active simultaneously, the overall rate is half the rate of the active stages. Thus, to achieve a rate threshold of \( \rho \), the rate threshold for the active stages should be set to \( 2\rho \). Note also that \( p_{th}(k) \)'s are modified in accordance with the updated values of \( \mathbb{E}[N_a(k)] \).

D. Expected Communication Delay

Average communication delay depends on the number of hops in the transmission and the communication rate. Using the bounds (23) and (25) on the total number of hops \( K \) in Sect. V-A, and denoting the delay of the typical device due to the transmission of a payload of \( M \) bits at rate \( R \) over \( K \) stages by \( D(R, K) \), it satisfies \( D(R, K) = t(R, K) - t(R, 1) \), where \( t(R, 1) \) is the direct transmission duration as a function of the transmission rate \( R \) in the unit of bits/s and the payload \( M \) in the unit of bits, and it is given as \( t(R, 1) = MR^{-1} \), and \( t(R, K) \) is the transmission duration over \( K \) hops. We calculate the average transmission delay conditioned on the devices that satisfy the minimum rate threshold \( \rho \) over \( K \) hops, which is obtained as

\[
\mathbb{E}[D(R, K) | R > \rho] = \mathbb{E}[t(R, K) | R > \rho] - ME\left[\frac{1}{R} | R > \rho\right],
\]

where \( \mathbb{E}\left[\frac{1}{R} | R > \rho\right] = \int_0^{1/\rho} \left[1 - \mathbb{P}\left(R \geq \frac{1}{T}\right)\right] dt \) is given by

\[
\mathbb{E}\left[\frac{1}{R} | R > \rho\right] \approx \sum_{l=1}^{\infty} \int_0^{1/\rho} \left[1 - \exp\left(-\frac{2\pi \tau}{\alpha-1} - 1\right)\right] dt \mathbb{P}_{N_a}(l),
\]
which can be calculated using the rate distribution given in (30).

The expected delay depends on the transmission protocol. As discussed in Sect. VI, we mainly focus on two different transmission protocols, i.e., successive and full-duplex parallel transmission modes as described in detail in Sects. VI-A and VI-B, respectively.

**Successive Transmission.** Using the distribution of the sequential mode transmission rate $R_S$ given in (37), the expected communication duration for $K$ stages can be obtained as

$$
\mathbb{E}[t(R_S, K) | R_S > \rho] = MK \mathbb{E} \left[ \frac{1}{R_S} | R_S > \rho \right] \approx MK \int_0^{1/\rho} \left[ 1 - \prod_{k \in K} \sum_{l=0}^{\infty} \exp \left( - \frac{2^{k l / \frac{1}{2} - 1}}{2^{k l / \frac{1}{2} - 1}} C_\alpha (2^{k l / \frac{1}{2} - 1})^{\gamma_{opt}} \right) \mathbb{P}_{N_a(k)}(l) \right] dt.
$$

**Parallel Transmission.** The distribution of the full-duplex parallel mode transmission rate $R_P$ is given in (44). Therefore, the expected communication duration is obtained as

$$
\mathbb{E}[t(R_P, K) | R_P > \rho] = MK \mathbb{E} \left[ \frac{1}{R_P} | R_P > \rho \right] \approx MK \int_0^{1/\rho} \left[ 1 - \prod_{k \in K} \sum_{l=0}^{\infty} \lambda_{pu}(k) \lambda_{pu}(j) \sum_{j=1}^{p_{th}(j)} \mathbb{P}_{N_a(k)}(l) \exp \left( - \frac{2^{k l / \frac{1}{2} - 1}}{2^{k l / \frac{1}{2} - 1}} \frac{p_{th}(k) \lambda_{pu}(k)}{\sum_{j=1}^{p_{th}(j)} \lambda_{pu}(j) C_\alpha (2^{k l / \frac{1}{2} - 1})} \right) \right] dt.
$$

The expected delay for the sequential (or full-duplex parallel) mode can be found by combining (47) (or (48)) with the delay expression in (46).

**VII. Numerical Results**

For the performance evaluation of the aggregator-based M2M communication model in Sects. III-IV and the rate coverage models in Sect. VI we incorporate the power-dissipation factors of...
IEEE 802.11b (WiFi) technology [14]. The simulation parameters are shown in Table III.

**Optimal number of hops.** We illustrate the trend between the number of required stages versus the outage probability constraint for different thresholds $T = \{0.01, 0.1, 1\}$ in Fig. 2. For low outage probabilities, i.e., small $\varepsilon$ values, $K_U$ should be also low as the average transmission range is long. As $\varepsilon$ increases, the maximum transmission range decreases and $K_U$ increases. The variation of the lower bound on the number of hops, $K_L$, for varying $\varepsilon$ and $K_U$ is also shown.

**Optimal fraction of receivers decay with the number of stages.** We evaluate the performance of the energy model in Sects. III and IV for $K \in \{1, \ldots, 10\}$. In Figs. 3-4, we observe the variation of the average total energy density with respect to $\gamma$ for different $P_{LO}$, where $E_D$ (solid line) corresponds to the energy density of direct transmission to the BS. The optimal values of $\gamma$ that minimize the total energy density, i.e., $\gamma^*$ (marked), are decreasing with $K$.

| Parameter               | Value/Range           | Parameter               | Value/Range           |
|-------------------------|-----------------------|-------------------------|-----------------------|
| Total bandwidth         | $W = 10^5$ Hz         | Power consumption of RX (TX) | $P_{RX} = 200$ ($P_{TX} = 100$) mW |
| Path loss exponent      | $\alpha = 4$         | Network size            | $2R \times 2R$ sq. km, $R = 2.5$km |
| Device (BS) density     | $\lambda = 10^3$ ($\lambda_{BS} = 1$) per sq. km | Large-scale fading $S$ | Lognormal (8 dB std. dev.) |
| Payload                 | $M = 100$ bits       | Small-scale fading $g$  | $\exp(1)$             |

**TABLE III:** Simulation parameters.
The simulation setup for the verification of analytical rate models developed in Sect. VI is as follows. Device locations are distributed as PPP over a square region of size $5 \times 5$ sq. km. Note that the density of aggregators at stage $k$ is $\lambda \gamma^k$, i.e., the number of aggregators decay geometrically. Therefore, for high $K$ values we need a region with much larger area for the validation of the model, but scaling the region increases the computational complexity exponentially. Therefore, we restrict ourselves to $K = 1 : 4$, which is also consistent with the plot characteristics on the number of hops illustrated in Fig. 2 and investigate the performance of the proposed model. Furthermore, we approximate (37) by truncating $N_a$ over the range $l = 0, 1, \ldots, L$, where $L$ is set to 20. This is a reasonable range as the fraction of aggregators is chosen to be $\gamma = 0.1$, i.e., there are 9 devices per aggregator on average.

**Rate coverage for half-duplex and full-duplex modes.** The rate coverage for the sequential and the full-duplex parallel modes for different number of hops are illustrated in Figs. 5 and 6. The analytical results almost match the simulations, and the difference is due to the *interstage independence assumption* in Sect. V-A and the *independent power control assumption* in [20]. Comparing Figs. 5 and 6 we observe that the full-duplex parallel mode does not offer higher rate coverage compared to sequential mode, which is due to *inter-stage interference* as detailed in Sect. VI-B. Therefore, in terms of rate coverage, sequential mode is preferable over full-duplex mode for the proposed setup. The rate coverage for the half-duplex parallel mode for $K = \{2, 4, 6\}$ hops is illustrated in Fig. 7. This mode has higher coverage compared to full-
Fig. 8: A rate outage-total energy density tradeoff for sequential and parallel modes, $K = 3$.

Fig. 9: A rate outage-total energy density tradeoff comparison of half-duplex parallel model for $K = 6$.

duplex parallel mode for the same number of stages and its rate coverage performance is close to the one of the sequential transmission mode.

**Rate-energy tradeoffs.** We compare the average total energy density for the sequential and full-duplex parallel schemes in Fig. 8. The rate coverage decays as the hop number increases, which reduces the chance of successful transmissions, and the overall energy consumption. To make a relevant comparison based on the energy consumptions of sequential and parallel modes, we consider multiple transmission cycles equal to the number of stages, i.e., $K = 3$. On the plot legend, $E_P$ and $E_S$ stand for the total energy densities, and $Q_P$ and $Q_S$ stand for the rate outage probabilities, i.e., complements of the respective rate coverage probabilities, of the full-duplex parallel and sequential transmission modes, respectively. In full-duplex parallel transmission, all devices are always active and transmitting, but the communication rate is low due to higher interference, hence, its energy consumption is less than the successive mode. We also investigate the average total energy density for the half-duplex parallel scheme with $K = 6$ in Fig. 9. In the figure legend, HD stands for half-duplex. This scheme has higher coverage probability as only half of the stages are operating simultaneously, and higher transmit distance as the aggregator to device fraction is $\gamma^2 \ll \gamma$, which yield higher energy consumption per transmission.

Different transmission modes have revealed the tradeoffs between the coverage and the energy requirements. We observe that full-duplex parallel mode has low energy density, but low rate coverage and high delay. Sequential mode also has low energy density, high rate coverage and low
delay. On the other hand, half-duplex parallel mode has higher energy density (×2) and higher rate coverage compared to full-duplex parallel mode. Unlike the full-duplex parallel mode, which is not convenient for M2M communication despite being energy efficient, the half-duplex parallel mode is a favorable method with higher energy consumption. However, the energy density of the model can be reduced by readjusting (optimizing) the aggregator fractions for alternating transmission stages. Since M2M is delay tolerant, sequential transmission mode is also feasible and a low-cost technique. Considering the operating regime for M2M devices and the simplicity of their design, sequential mode is preferable as it has low energy density and high coverage.

VIII. CONCLUSIONS

We study a general multi-hop-based uplink communication scheme for M2M communication, and develop a new data aggregation model for M2M devices, using tools from stochastic geometry. To the best of our knowledge, this is the first work on power-limited communication providing a unified energy consumption model with coverage in cellular networks. Our results show that the uplink coverage characteristics dominate the trend of the energy consumption for the proposed transmission modes. Considering the operating regime of M2M devices, sequential and half-duplex parallel modes are more feasible compared to full-duplex mode.

Interesting extensions would include the minimization of the energy expenditure through joint optimization of the optimal number of multi-hop stages and fraction of aggregators. Better energy models can be developed to incorporate the different states of the transceiver, i.e., on, idle, sleep and off states, which will provide a more accurate energy model for the proposed communication scheme. Strategies for synchronization of transmissions is also important to prevent multi-hop delays and save receiver energy consumption. Multi-slope path loss models [22] would more accurately model the effect of interference and noise for networks with increasing device density.

ACKNOWLEDGMENT

The authors thank Sarabjot Singh for helpful feedback.
APPENDIX A
PROOF OF THEOREM 1

Let $\Psi_a = \sum_n \delta_X_n$, $\Psi_u = \sum_n \delta_X_n$, and $V_0$ be the Voronoi cell of the typical aggregator $X_0 \in \Psi_a$ located at the origin. Notice that $x \in V_0$ if and only if $\Psi_a(B^o(x, \|x\|)) = 0$. Hence,

$$
E^0_{\Psi_a} \int_{\mathbb{R}^2} f(x)1_{x \in V_0} \, dx = \int_{\mathbb{R}^2} f(x)P^0_{\Psi_a}[\Psi_a(B^o(x, \|x\|)) = 0] \, dx,
$$

which is the expectation with respect to the conditional (Palm) probability conditioned on $x \in V_0$. As $\Psi_a$ is PPP due to thinning, from Slivnyak’s theorem [15], the RHS can be rewritten as

$$
\int_{\mathbb{R}^2} f(x)P^0_{\Psi_a}[\Psi_a(B^o(x, \|x\|)) = 0] \, dx = \int_{\mathbb{R}^2} f(x)P[\Psi_a(B^o(x, \|x\|)) = 0] \, dx.
$$

Using the Poisson law, we have $E^0_{\Psi_a} \sum_n f(X_n)1_{X_n \in V_0} = \lambda_n \int_{\mathbb{R}^2} f(x)e^{-\lambda_n \|x\|^2} \, dx$. We let $f(x) = \|x\|^{\alpha}$ so that the received power at the aggregator is unity. Then, we have the following final result for the mean total uplink power of the devices in the Voronoi cell of the typical aggregator:

$$
P(\lambda_a) = 2\pi \lambda_u \int_{r>0} r^{\alpha-1}e^{-\lambda_u \pi r^2} r \, dr = \frac{\lambda_u \alpha/2}{\lambda_a^{1+\alpha/2} \pi^{\alpha/2}} \Gamma\left(\frac{\alpha}{2}\right).
$$

APPENDIX B
PROOF OF PROPOSITION 3

Note that $E(\lambda_a(K)) = \lambda \sum_{k=1}^K c_k(\gamma)$, where $c_k(\gamma)$ is given in Proposition 1. Note also that

$$
\min\{E(\lambda_a(K))\} + \lambda \min\{c_{K+1}(\gamma)\} \leq \min\{E(\lambda_a(K+1))\}.
$$

Optimal solutions of $E(\lambda_a(K))$ and $E(\lambda_a(K+1))$ are $\gamma^*(K) = \arg \min E(\lambda_a(K))$, and $\gamma^*(K+1) = \arg \min E(\lambda_a(K+1))$, which yield

$$
\sum_{k=1}^K c_k(\gamma^*(K)) + \lambda \min\{c_{K+1}(\gamma)\} \leq \sum_{k=1}^{K+1} c_k(\gamma^*(K+1)).
$$

It is also clear that

$$
\sum_{k=1}^K c_k(\gamma^*(K)) \leq \sum_{k=1}^K c_k(\gamma^*(K+1)).
$$

Using (49) and (50), we have $\min\{c_{K+1}(\gamma)\} \leq c_{K+1}(\gamma^*(K+1))$. Assume that $\gamma^*(K+1) > \gamma^*(K)$. Then, from Remark 2, $c_k(\gamma^*(K)) > c_k(\gamma^*(K+1))$ for $k \in \{1, \ldots, K-1\}$, and $c_K(\gamma^*(K)) < c_K(\gamma^*(K+1))$. Using (50), and since $c_{K+1}(\gamma^*(K+1))$ is increasing in $\gamma^*(K+1)$, we can reduce it by decreasing $\gamma^*(K+1)$ to $\gamma^*(K)$, which gives a contradiction. Thus, $\gamma^*(K+1) \leq \gamma^*(K)$. 

REFERENCES

[1] “Cisco visual networking index: Global mobile data traffic forecast update, 2014-2019,” white paper, Cisco, Feb. 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/whitepaper_c11-520862.pdf

[2] H. S. Dhillon, H. Huang, and H. Viswanathan, “Wide-area wireless communication challenges for the Internet of Things,” submitted, arXiv preprint arXiv:1504.03242, 2015.

[3] S.-Y. Lien, K.-C. Chen, and Y. Lin, “Toward ubiquitous massive accesses in 3GPP machine-to-machine communications,” IEEE Communications Magazine, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[4] New WI proposal: Further LTE Physical Layer Enhancements for MTC, 3GPP Std. TDocs-RP-141 660. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg--RP-65--30566.htm

[5] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, “Power-efficient system design for cellular-based machine-to-machine communications,” IEEE Trans. Wireless Commun., vol. 12, no. 11, pp. 5740–5753, Nov. 2013.

[6] ——, “Fundamentals of throughput maximization with random arrivals for M2M communications,” IEEE Trans. Commun., vol. 62, no. 11, pp. 4094–4109, Nov. 2014.

[7] A. Laya, L. Alonso, and J. Alonso-Zarate, “Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives,” IEEE Communication Surveys and Tutorials, vol. 16, no. 1, pp. 1–13, 2014.

[8] D. Niyato, P. Wang, and D. I. Kim, “Performance modeling and analysis of heterogeneous machine type communications,” IEEE Transactions on Wireless Communications, vol. 13, no. 5, pp. 2836–2849, May 2014.

[9] A. Aijaz, M. Tshangini, M. Nakhai, X. Chu, and H. Aghvami, “Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees,” IEEE Trans. Commun., vol. 62, pp. 2353–2365, 2014.

[10] V. Rodoplu and T. Meng, “Minimum energy mobile wireless networks,” IEEE J. Sel. Areas Commun., vol. 17, no. 8, pp. 1333–1344, Aug. 1999.

[11] F. Baccelli and S. Zuyev, “Poisson-Voronoi spanning trees with applications to the optimization of communication networks,” Operations Research, vol. 47, no. 4, pp. 619–631, 1999.

[12] S. J. Baek, G. de Veciana, and X. Su, “Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation,” IEEE J. Sel. Areas Commun., vol. 22, no. 6, pp. 1130–1140, Aug. 2004.

[13] M. Haenggi and D. Puccinelli, “Routing in ad hoc networks: A case for long hops,” IEEE Commun. Mag., vol. 43, no. 10, pp. 93–101, Oct. 2005.

[14] A. Y. Wang and C. Sodini, “On the energy efficiency of wireless transceivers,” in Proc., IEEE ICC, 2006, pp. 3783–3788.

[15] F. Baccelli and B. Blaszczyszyn, Stochastic Geometry and Wireless Networks, Volume I — Theory, ser. Foundations and Trends in Networking. NoW Publishers, 2009, vol. 3, no. 3–4.

[16] S. M. Yu and S.-L. Kim, “Downlink capacity and base station density in cellular networks,” in Proc., 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), May 2013, pp. 119–124.

[17] S. Singh, X. Zhang, and J. G. Andrews, “Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets,” submitted, IEEE Trans. Wireless Comm. Available: http://arxiv.org/pdf/1412.1898, 2014.

[18] J.-S. Ferenc and Z. Neda, “On the size-distribution of Poisson Voronoi cells,” Physica A: Statistical Mechanics and its Applications, vol. 385, no. 2, p. 518526, Nov. 2007.

[19] S. Singh, H. Dhillon, and J. Andrews, “Offloading in heterogeneous networks: Modeling, analysis, and design insights,” IEEE Trans. Wireless Commun., vol. 12, no. 5, pp. 2484–2497, 2013.
[20] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, “Analytical modeling of uplink cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2669–2679, Jun. 2013.

[21] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed. John Wiley and Sons, 1996.

[22] X. Zhang and J. G. Andrews, “Downlink cellular network analysis with multi-slope path loss models,” *to appear, IEEE Trans. Commun.*, 2015.