Smallholder Agriculture in the Information Age: Limits and Opportunities

Mariya Zheleva
Computer Science
University at Albany SUNY
mzheleva@albany.edu

Petko Bogdanov
Computer Science
University at Albany SUNY
pbogdanov@albany.edu

Daphney-Stravoula Zois
Electrical and Computer Engineering
University at Albany SUNY
dzois@albany.edu

Wei Xiong
Computer Science
University at Albany SUNY
wxiong@albany.edu

Ranveer Chandra
Microsoft Research
Redmond, WA
ranveer@microsoft.com

Mark Kimball
Essex Farm
Essex, NY
essexfarm@gmail.com

ABSTRACT
Recent projections by the United Nations show that the food production needs to double by 2050 in order to meet the nutrition demand of the world’s growing population. A key enabler of this growth are smallholder family farms, that form the backbone of agricultural (AG) production worldwide. To meet this increasing demand, smallholder farms need to implement critical advances in task management and coordination, crop and livestock monitoring and efficient farming practices. Information and Communication Technology (ICT) will play a critical role in these advances by providing integrated and affordable cyber-physical systems (CPS) that can longitudinally measure, analyze and control AG operations. In this paper we make headway towards the design and integration of such AG-CPS. We begin by characterizing the information and communication technology demand of smallholder agriculture based on traffic analysis of farm Internet use. Our findings inform the design and integration of an end-to-end AG-CPS called FarmNET that provides (i) robust control mechanisms for multi-sensor AG data collection and fusion, (ii) wide-area, heterogeneous wireless networks for ubiquitous farm connectivity, (iii) algorithms and models for farm data analytics that produce actionable information from the collected agricultural data, and (iv) control mechanisms for autonomous, proactive farming.

KEYWORDS
Smallholder agriculture, integrated AG-CPS, wireless networks, data analytics, control and automation.

1 INTRODUCTION
Smallholder farms rely predominantly on single-family labor. Such farms form the backbone of agricultural (AG) production and are essential to eradicate hunger in the face of a changing climate, while preserving our natural resources [60]. Recent estimates indicate that 80% of the food produced in the developing world comes from smallholder farms [78]. This number is far exceeded in the U.S., where USDA estimates that 97.6% of farms are smallholder enterprises and they are responsible for 85% of the nation’s AG production [88]. With the world’s booming population, the United Nations foundation estimates that the farm production needs to double by the year 2050 for society to be able to eradicate hunger and secure nutrition [52]. This creates an appealing market opportunity for smallholder farms to proliferate while solving one of humanity’s big challenges. Such growth mandates improved efficiency in current farm practices related to (i) task management and coordination, (ii) crop and livestock monitoring, data analytics and control, (iii) expansion of local markets and (iv) access and adoption of new farming practices [66].

Information and communication technology (ICT) will play a critical role in such advances, however, current technologies [9, 14, 17–19] are highly-specialized (e.g. focusing on soybean production), provide closed solutions in that farmers have no control over their data and corresponding analytics, and most importantly, are not affordable for smallholder enterprises. Thus, practical progress in ICT for smallholder agriculture hinges on the availability of technology that either does not exist or needs to be re-purposed from its predominantly urban context to fit the unique spatial, temporal and environmental characteristics of smallholder farming. Such technology includes (i) robust sensing infrastructures to measure farm state and operations, (ii) ubiquitous wireless network connectivity to transmit sensor and farmer data, (iii) domain-specific data models and analytics to extract actionable knowledge from the data, and (iv) adaptive control algorithms for efficient sensing and proactive control of farm processes towards autonomous farming. Beyond availability, it is critical for these technologies to be seamlessly integrated into an AG cyber-physical system (AG-CPS) that interacts with farmers, farm assets and processes to efficiently measure, analyze and control them, and inform decision-making, improved farming practices, distribution chains and consumer relations. To this end, there is a need of fundamental research and
Engineering of novel and integrated end-to-end mechanisms for farm sensing, network connectivity, data analytics and control.

Related research and technology integration can be subdivided into (i) wireless networks for under-served areas [37, 69, 72, 74, 75, 91, 107, 110, 116, 117, 121, 134, 135], (ii) data mining for AG applications [34, 48, 51, 59, 61, 64, 70, 80, 89, 90, 92, 101, 102, 104, 119, 122, 132, 137, 144], and (iii) estimation and control for decision making in uncertain environments under constraints [146–156]. Key limitations of existing solutions are that they are either developed for non-AG contexts or tackle connectivity, data analytics and control in isolation. To bring meaningful ICT innovation in agriculture, we need to adopt an integrated approach.

Our work informs such an integrated approach via systematic analysis of farm Internet use and ICT needs. Our analysis is based on a year-long and continuing collaboration with Essex Farm in upstate New York that has allowed us to learn first-hand about the ICT needs of farm operations. With farmers' permission, we have also been able to collect traces of farmer mobility and Internet use. Our analysis of these traces shows that farm traffic is a unique mix of farmer and IoT sensor activity with interlocking characteristics. In terms of volume and direction, the IoT sensor traffic is upload-intensive, while farmer traffic is download-intensive. While farmer traffic is bursty and unpredictable, that of IoT sensors is primarily periodic and, thus predictable. Finally, while farmer traffic is spatially-concentrated, IoT sensor traffic is distributed across the farm's territory.

These insights create a unique design space for (i) ubiquitous AG wireless network architecture and protocols, (ii) novel AG data analytics and (iii) AG control. We integrate these key components and present our vision of an end-to-end AG-CPS called FarmNET. FarmNET integrates four key components to collect longitudinal data, and utilize real-time data analytics, domain-specific models and control algorithms to enable increased quality and productivity of farm operations, with minimal footprint.

1 Sensing frontend. Each FarmNET sensor will be wirelessly connected and highly-reconfigurable. The wireless capability of sensors will allow for seamless farm data offload, whereas the reconfigurability will enable adaptive sampling to efficiently manage the tradeoffs between volume and periodicity of farm measurements versus accuracy of data analytics and control algorithms.

2 Communication network. To accommodate the unique farm traffic, we envision a heterogeneous wireless network that is comprised of a plug-and-play wireless backhaul realized over TV white spaces, and a three-modal last mile implemented over LTE or Wi-Fi. This architecture design poses fundamental challenges in (i) joint cross-layer optimization of last-mile and backhaul access that is informed by the properties of the heterogeneous farm traffic, and (ii) characterization, modeling and integration of power efficiency of AG-CPS.

3 Data analytics. A key challenge in FarmNET is to enable robust knowledge discovery from noisy and sparse sensor measurements and to employ them for analytic tasks such as anomaly detection, root cause analysis, historical trend detection and prediction based on statistical data-driven models and simulation. This problem lends itself for a graph-theoretical formulation with nodes representing farm entities (pastures, animals, arable fields, etc.), connections modeling interactions among entities, and multi-variate graph signals modeling temporal entity states (e.g., soil moisture, animal health and milk production). This dynamic heterogeneous graph framework will enable a holistic understanding of all sensed farm operations and enable modeling and control-enabled optimization of the global farm health.

4 Control. To enable high-output and efficient, controlled–environment AG technologies and systems, we envision a holistic controlled sensing framework that will integrate IoT–sensing capabilities with agriculture data collection, network structure and humans in the loop, to enable real–time accurate agricultural monitoring and control. To this end, we propose (i) a stochastic dynamic system model that fully describes the farm’s state (i.e., health and footprint) over time, while incorporating the effect of both cyber (e.g., control signals) and physical components (e.g., agricultural variables), (ii) recursive, structured farm state estimators, and (iii) strategies that control sensing and farm processes by optimizing farm state estimation accuracy and different operation costs (e.g., sensing, farm processes).

In what follows, we first describe the current state of AG ICTs ($\S 2$). Next, we provide analysis of farm ICT needs that is based on empirical evaluation of farm Internet access ($\S 3$). Our analysis informs the design of an end-to-end AG-CPS dubbed FarmNET that is presented in $\S 4$. Finally, we conclude our paper in $\S 5$.

2 LIMITS AND OPPORTUNITIES

In this section, we first describe the current state of ICTs for agriculture and detail their limitations. We survey existing solutions that can be harnessed to address these limitations. We also discuss several technological needs that cannot be met by re-purposing of existing technologies and require fundamentally new design.

2.1 Current State of AG ICTs

Multiple solutions that target precision agriculture and AG decision support exist both in industry and academia.

2.1.1 Industrial products. Industrial products can be largely subdivided in such that target (i) sensing [1, 2, 5, 7, 14], (ii) data analytics [4, 8, 10–13, 15–17, 17, 21] and (iii) consumer relations [3]. A large fraction of the sensing solutions perform single-modality sensing, i.e., only imagery [2, 5], moisture [7] or nitrates [1]. gThrive [14] is the only one that supports multi-modal sensing, and basic data fusion and analytics. The industry has largely focused on AG data analytics with a large number of start-ups [4, 8, 10–13, 15, 17, 21] and well-established AG corporations [16, 17] entering this business. All of these are closed-form, cloud-based solutions that do not allow flexibility in data management and do not provide farmers access to their own data. This, as found in recent research [82], has raised concerns around data privacy, security and control. A common limitation found by smallholder farmers with regards to existing industrial products is that they are prohibitively-expensive, and thus, not economically-feasible for smallholder farm operations. In addition, such systems are typically not open and thus do not provide opportunities for modular customizations for different smallholder farm operations.
**Summary of limitations.** Industrial products have several key limitations, related to cost, and flexibility of access and data management. Furthermore, none of the existing industrial products provide a wide-profile, end-to-end solution; instead, they focus either on sensing or data analytics and typically take network connectivity for granted. While decision support is the focus of some existing products, current efforts in control mechanisms for autonomous farming are limited, highly-specialized and out of financial reach of smallholder farms.

2.1.2 Academic research. Related research and technology integration can be subdivided into (i) wireless networks for underserved areas [37, 69, 72, 74, 75, 91, 107, 110, 116, 117, 121, 134, 135], (ii) data mining for AG applications [34, 48, 51, 59, 61, 64, 70, 80, 89, 90, 92, 101, 102, 104, 119, 122, 132, 137, 144] and (iii) estimation and control for decision making in uncertain environments under constraints [146–156]. Key limitations of existing solutions are that they are either developed out of the AG context or tackle connectivity, data analytics and control in isolation.

Recent technological advances in precision agriculture hinge on the availability of **wireless network connectivity**, however, all of them take connectivity for granted (e.g. products described in §2.1.1). At the same time, smallholder farmlands, with their extremely-low population density, often provide the least-appealing business case for commercial network deployments. As a result, farmlands are characterized with spotty, inconsistent, intermittent or all together lacking network coverage. Advances in wireless networks for under-served areas [37, 69, 72, 74, 75, 91, 107, 110, 116, 117, 121, 134, 135] bring promise for improved farm connectivity. Unfortunately, existing solutions are designed exclusively for human-generated traffic and are not readily applicable for farm connectivity that needs to accommodate a mix of human and IoT sensors traffic with varying delay constraints (detailed description and preliminary results in § 3). A large volume of prior work focuses on wireless sensor network connectivity (WSN) [22–25, 29, 31, 56, 62, 142] with some specializing in WSN for agriculture [33, 58, 111, 139–141]. A key limitation of these works is that they only accommodate sensor data and will not scale well for an integrated AG-CPS such as FarmNET, that is optimized to handle heterogeneous farmer-sensor traffic.

In **AG data analytics**, data mining and machine learning techniques have been applied to extract high-level knowledge of the farm state [119]. Of central interest are crop [122, 137] and animal health [48, 51, 59], soil properties [35, 100], animal tracking and behavior inference [59, 70, 80, 90, 102, 104, 122, 132, 137, 144]. These techniques are designed for offline processing of previously-collected data, however, they are not suitable for real-time tracking of multiple interacting entities (e.g. animals, pastures, feed and weather) as an evolving network. Tracking of such multi-entity, longitudinal interactions requires a dynamic-network-mining approach and is critical to enable anomaly detection, root-cause analysis and realistic simulation for “what-if” analysis. Dynamic network mining is an emerging research field that has produced scalable methods for anomalous temporal subnetwork detection [44, 98, 99, 130], prediction of the network’s global state based on local properties [54, 55], information and disease propagation [42, 43] applied to transportation, biological and social networks. Such methods, however, are not readily-adoptable in the AG context, as AG processes incur different interaction dynamics. For example, the temporal interaction between grazing herd animals and pasture paddocks requires novel definitions of anomalies and novel predictive models for pasture productivity in the presence of grazing animals, varying weather and nutrients within the spatial network of paddocks.

Prior work on monitoring and control for precision agriculture and farm monitoring such as [28, 30, 36, 39, 40, 45, 46, 53, 58, 67, 81, 103, 143] has considered either ad-hoc or static optimization approaches. However, real-time and cost-efficient monitoring and control requires rigorous dynamic farm system modeling, which jointly considers the cyber and physical components and precisely defines their interactions, optimization and control system theories. Similar approaches have been successfully applied to other applications (e.g., environmental monitoring [20, 49, 94–97, 115, 127–129], target tracking [32, 47, 50, 65, 71, 73, 83, 86, 106, 124, 125], physical activity tracking [145, 147, 149, 152]), however, the proposed solutions are not readily applicable in the AG context, since AG-CPS (i) require more complex dynamic models, (ii) require functions that control both sensing and farm processes, and (iii) exploit the unique AG system characteristics (e.g., humans–in–the–loop, network–induced constraints, model structure).

**Summary of limitations.** While the academic community has made a substantial headway towards data-driven agriculture, no work focuses on providing an end-to-end solution to enable sustainable, proactive and autonomous agriculture. Furthermore, a key limitation of existing data-driven approaches is that they focus on single-time, offline analytics and are thus not suited for longitudinal, real-time and actionable analysis of AG data. Similarly, AG control has considered static optimization or ad hoc approaches, however, further development is necessary to enable real-time, dynamic control. Finally, in terms of wireless networks, existing solutions focus on accommodating human-generated traffic and will not scale well for the heterogeneous demand on farm networks.

2.2 Limits and Opportunities Specific to Smallholder Farming

Varying seasonal workforce; limited connectivity in remote rural locations; lack of accessible systems for quality control, planning and operational analytics; and maintaining close working relationships with (possibly multiple) end-customers are among the main challenges for smallholder family-operated farms. Thus, automation of the common monitoring tasks via low-cost sensing and connectivity solutions, data-driven planning and control as well as offering measurable sustainability/quality statistics for end customers open tangible opportunities for improving smallholder farming enterprises of varying scales. These challenges transcend to developing world smallholder farms, although the tradeoffs between cost and system utility in these scenarios need to be further considered.

In order to embrace the above opportunities, emerging AG-CPS need to employ longitudinal data, and utilize real-time data analytics, domain-specific models and control algorithms to enable increased quality and productivity with minimal footprint. These challenges create an appealing research agenda for (i) robust multi-sensor data collection and fusion, (ii) wide-area, heterogeneous wireless networks for ubiquitous farm connectivity, (iii) algorithms and
models for farm data analytics that produce actionable information from raw sensor data, and (iv) novel estimation and control mechanisms for autonomous and proactive farming. These components will need to be seamlessly-integrated in a holistic and modular AG-CPS. In the design of such integrated AG-CPS, of central importance should be the trade-off between the economic feasibility of the sensing and communication infrastructure and the accuracy and efficiency of data analytics, monitoring and control to promote proactive farming practices. This tradeoff can be tackled by developing and adopting algorithms, models, hardware and software that leverage open-source and highly-reconfigurable components.

3 ANALYSIS OF FARM ICT NEEDS

Our analysis of farm ICT needs is based on a measurement campaign we executed in Essex Farm between July and December of 2016. In what follows we provide background on Essex Farm, our methodology and objectives, results, and design implications.

3.1 Essex Farm

Essex Farm1 was established in 2004 by Mark and Kristin Kimball and is a unique diverse-profile family-operated farm in Upstate New York that spans an area of 1,100 acres. It operates as a farm-to-door CSA (Community Supported Agriculture), but unlike classical CSAs, it provides a full, all-you-can-eat diet, year-round to its members. The farm specializes in a diverse profile of agricultural activities from vegetable and fruit production to grains, eggs, dairy and wide spectrum of meats. Beyond production, the farm also collaborates with a local enterprise called The Hub on the Hill2 to make preserves from the seasonal produce to maintain its supplies year-round. The farm employs anywhere between 5 and 20 additional farmers throughout the year. These farmers are typically young professionals, who come from different parts of the U.S. and Europe and are looking to get training and hands-on experience with farming. Thus Essex Farm provides them with a unique opportunity to (i) learn in a farm with a diverse activity profile and (ii) interact with cutting-edge IT innovation that is undergoing on the farm. Overall, the diverse activity profile of the farm, its farm-to-door operation, its employment of young farming professionals and its collaboration with other enterprises in AG sustainability makes for a unique ecosystem to understand AG ICT needs and opportunities.

3.2 Methodology

The farm Internet access is currently provided over a 5MBps microwave wireless terrestrial link that beams over lake Champlain to connect the farm with their ISP in Vermont. The gateway link is then locally-distributed through three Wi-Fi access points connected to the gateway via an Ethernet LAN. Besides the farm Wi-Fi, there is also limited coverage provided by commercial mobile carriers. The goals of our measurements were to (i) characterize the commercial network availability and quality on the farm, and (ii) understand the volume, direction and spatio-temporal characteristics of farm traffic demand. For the first task, we developed an Android application to collect geo-tagged network performance information every 30 seconds and submit it to our server for storage and analysis. Two phones running the app were carried by different farmers over the course of a week in high season (July). For the second task, we collected longitudinal pcap traces at the farm gateway. Once collected, we post-processed these pcap traces using tstat3 in order to extract individual TCP and UDP flows and study the per-flow performance and inter-arrival rate.

3.3 Analysis Results

The farm Internet demand is generated by a mix of sensors and farmers. Farmers use applications that require real-time Internet access, whereas sensors are a mix of real-time access and delay-tolerant nodes. In the remainder of our analysis, we split the collected traces into farmer-generated and sensor-generated, and apply the same analysis methodology to the two trace subsets. This leads to unique insights into the characteristics of farm ICT demand.

We begin our analysis by focusing on the traces collected by our Android application. Using these traces, we study the spatial characteristics of traffic demand, and the network availability and quality. We find that farmer traffic is highly-localized, whereas sensor traffic is spatially-distributed. We also find that the current Wi-Fi network and commercial cellular network are not able to meet the offered demand. We then focus on pcap trace analysis. We find that farmer and sensor traffic have opposing characteristics in terms of traffic volume, direction and predictability. In what follows, we detail our results.

3.3.1 Spatial distribution of supply and demand.

– Network availability. Figure 1 presents our results for RSSI measurements of Wi-Fi and one of the major U.S. mobile carriers4. Reliable Internet access is available only in the area around the farm office, house, shop and barn and is extremely poor (cellular) or lacking (Wi-Fi) in the rest of the farm.

– Spatial traffic characteristics. We split the farm territory (1,100 acres) into 10x10 meter squares and analyze the frequency of farmer visits of each square. In the course of a week only 2.5% of all the 10x10 grid squares were visited by farmers. This indicates high spatial concentration of farmer traffic demand, which means that a majority of farmer Internet access can be accommodated with several stationary always-on access points. Unlike farmers’ traffic, the IoT sensor traffic is spatially-distributed across the farm due to the need for ubiquitous farm measurements.

3.3.2 Traffic volume, direction and predictability.

– Temporal traffic characteristics. We collected longitudinal pcap traces that capture all the farm traffic (IoT sensors and farmers). We compare the flow inter-arrival time (IAT) of farmer and sensor traffic. According to Figure 2, the farmer traffic arrives at a wide range of intervals (from 100ms to 100s) and is unpredictable. At the same time, 70% of the IoT sensor traffic is characterized with an IAT of 50s, thus, IoT farm traffic is predictable.

– Volume and direction of traffic. Lastly, we are interested in characterizing the intensity of traffic (volume) in the uplink and downlink.

1http://www.essexfarmscsa.com/
2http://thehubonthehill.org/
3http://tstat.tlc.polito.it/
4Interactive maps available at https://goo.gl/yvEwZu
Figure 1: Cellular (left) and Wi-Fi (right) coverage on Essex Farm. No cellular access was detected in roughly 20% of the measurements. Even when available, the cellular network signal strength rarely exceeded -100dBm, which does not suffice for meaningful Internet access. Wi-Fi access is available only in the office area and is lacking anywhere else on the farm.

Figure 2: Flow inter-arrival time (IAT) of sensor (left) and farmer traffic (right). IoT sensor traffic is predictable with 70% of flows having IAT of 50s. Farmer traffic’s IAT varies between 100ms and 100s and is not predictable.

Figure 3: Volume and direction of sensor (left) and farmer (right) traffic. Sensor traffic is predominantly in uplink direction, averaging 100KB in size per flow. Farmer traffic is downlink-intensive with up to 500KB-sized flows.

<table>
<thead>
<tr>
<th>Spatial</th>
<th>IoT-Sensor traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Highly-localized</td>
</tr>
<tr>
<td>Periodicity</td>
<td>Real-time+delay-tolerant</td>
</tr>
<tr>
<td>Direction</td>
<td>Downlink</td>
</tr>
</tbody>
</table>

Table 1: Summary of findings and design outlook.

3.4 Design Implications

Our analysis, summarized in Table 1, shows that IoT sensor and farmers traffic have opposing characteristics across all evaluation criteria. This creates a unique design space for novel (i) AG wireless network architecture and protocols, (ii) AG data analytics and (iii) AG monitoring and control.

- **Implications on wireless networking.** The joint spatial and timeliness characteristics of farm traffic have direct implications on network architecture design, as they permit successful accommodation of majority of the farm traffic, without having to deploy a dense, always-on wireless network throughout the entire farm. Instead, we envision a plug-and-play wireless backhaul that maintains a number of always-on stationary hot-spots and allows opportunistic access for the swath of delay-tolerant sensor traffic. Our findings on traffic types, periodicity, volume and direction call for a cross-layer protocol design, that targets rapid transfer of heterogeneous farm traffic and is informed by the volume, direction and predictability of this traffic.

- **Implications on data management and mining.** The necessity to support farmer and sensor traffic will raise an important trade-off question for data analytics and mining using IoT sensor readings: What is the minimum temporal resolution for different sensing modalities that ensure high quality (e.g., correctly identified anomalies, accurate animal tracking, etc.), while minimizing the rate of sensor readings and thus, not overloading the backhaul with unnecessary sensor data? In addition, real time analytics will have to incorporate delay tolerance and possibly missing values, while still providing maximally useful results to the end users.

- **Implications on monitoring and control.** The statistical characteristics of IoT–sensor–generated traffic and the structure of the proposed network architecture suggests that measurement and control signal information will be communicated probabilistically. As a result, the dynamic farm state model will need to incorporate the specific network–induced constraints, while the monitoring and control processes models will need to account for the delayed arrival (or missing) of measurement and control signals. Real-time, cost–efficient controlled–environment agriculture requires the design of appropriate estimation and control strategies for under-served areas to ensure the unobstructed operation of the AG-CPS system.

4 FARMNET

FarmNET, as illustrated in Figure 4, is an integrated architecture for real-time agricultural data collection, analytics and farm control. The farm ecosystem consists of farmers, IoT sensors and farm operations pending optimization. FarmNET integrates sensing and communication with data analytics and control to facilitate longitudinal
data collection, and analysis to produce actionable information and enable proactive farming. In what follows, we detail our vision of the design of each of FarmNET’s three key components and discuss their integration into an end-to-end AG-CPS.

4.1 Sensing and Communications

Background and motivation. Connectivity for IoT sensors and farmers is central to an AG-CPS, however, sparsely populated smallholder farmlands are not economically appealing for commercial network deployments, and thus commercial coverage is spotty, inconsistent or lacking altogether. We identify several unique characteristics of smallholder farmlands that lead to such poor commercial connectivity. First of all, smallholder farmlands are characterized with extremely-low population density. An average smallholder farm spans a large territory: between 87 and 148 acres [87], rendering its population density orders of magnitude smaller than classical rural scenarios. Our preliminary results (§3) also demonstrate that network demand of smallholder farms is a unique mixture of high-volume, spatially-localized and bursty human-generated traffic and low-volume, spatially distributed and predictable IoT sensor traffic. High-volume (and high-revenue) demand that comes from farmers’ Internet access is often spatially-localized, however, low-volume demand of IoT sensors is spatially-distributed across the large farmland. To accommodate such demand for low population density over large territory, typical commercial cellular providers face low return of investments and thus have low economic incentives.

FarmNET network architecture. To accommodate the heterogeneous farm demand, we envision a hybrid, multi-modal last mile access and a plug-and-play wireless backhaul network architecture, illustrated in the Communications pane of Figure 4. The hybrid last mile features three operation modalities (i) stationary, always on hot-spots, (ii) bring-your-own (BYO) opportunistic hot-spots and (iii) drone-mounted hot spots for target dispatch. To enable resource sharing, extensibility and open design, farmers’ devices and IoT sensors will be Wi-Fi capable. To provide on-demand Internet access anywhere on the farm, we envision that last mile modalities will connect to a plug-and-play, wide-area backhaul network. Such plug-and-play network can be supplied over TV White Spaces (TVWS) in order to bridge the farm traffic to the Internet and the on-farm cloud without requiring the establishment of on-farm LANs or mesh networks, each of which incurs high deployment and operational cost. To connect to the TVWS network, each last mile modality will need to be equipped with a TVWS client device (TVWS CPE). For the stationary and BYO last mile solutions, the CPE can be collocated with the access point, however, the target dispatch modality would require a delay-tolerant connection, since a TVWS CPE would exceed the drone’s load constraints. FarmNET’s target dispatch last-mile will be scheduled by control mechanisms (§4.3) to offload sensor data from areas without always-on coverage.

4.1.1 Research challenges. The architecture and operations of our sensing and communications need to be informed by the unique nature of farm traffic demand. The system design objective should be minimization of the traffic distribution time from the network users (i.e. sensors and farmers) to the network gateway, with simultaneous optimization of the power efficiency of key architecture components. These objectives pose fundamental research challenges in cross-layer protocol design, joint last-mile and wireless backhaul optimization and energy-efficient communications.

Cross-layer protocol design. Four major factors affect the traffic distribution time in our architecture: (i) medium access control in the Wi-Fi last mile, (ii) queuing at the Wi-Fi access point (Wi-Fi AP) and the TVWS client, (iii) medium access control in the

Figure 4: FarmNET architecture.
TVWS wireless backhaul, and (iv) transmission of redundant sensor data. Thus, we need to speed up the last-mile medium access, allow adaptive queueing and processing of heterogeneous traffic at the last-mile access points, and provide adaptive traffic scheduling over the plug-and-play TVWS backhaul.

Adaptive scheduling of last-mile traffic. Efficiently transmitted multi-user data from the last mile need to be further propagated over the TVWS CPE to the TVWS link and then to the local network gateway. Additional delay components we need to manage in this step are due to (i) data queueing and (ii) scheduling of transmissions over the TVWS link. Current TVWS network standards assume a point-to-multi-point (P2MP) architecture and include IEEE 802.11af [63] and IEEE 802.22 [133]. The former uses Wi-Fi-like contention-based schemes for random medium access control, limiting the effective range of this standard [108]. In the context of TVWS it has been demonstrated for ranges up to 1km [85]. IEEE 802.22 uses a TDD-based channel partitioning that achieves up to 100km communication range [85]. Our AG-CPS context calls for wide-area communication networks, and thus, AG-CPS wireless backhauls should be 802.22-centric.

Traditionally, packet scheduling in TDD networks takes a reactive approach: as queues fill up, the clients request uplink resources in order to transmit back-logged data. This approach incurs delays as traffic is queued waiting for the TDD link to cycle through its frame structure. Our prior observations of sensor data volume, direction and predictability, open space for design of proactive and adaptive resource allocation schemes that schedule uplink slots for sensor traffic at the time of traffic arrival. As a result, the TVWS backhaul will provide P2MP communications, while efficiently accommodating sensor traffic fluctuations and scaling back its resources to handle farmer traffic.

Power-efficient sensing and communication. Along with commercial network connectivity, power is the next most scarce resource on farmlands. While electricity may be available in some key locations on a farm, such as the farmhouse, office and barns, it is largely unavailable on the remainder of the farmland. This mandates that AG-CPS designs should optimize power consumption and provide uniform services despite non-uniform availability of power across the farmland. The unique integrated design of FarmNET posits novel research questions that will further our understanding of AG power consumption, and enable power-efficient design of AG-CPS. Power efficiency of communication systems has been studied in multiple domains from cellular network infrastructures [38, 74, 76, 93, 109] to end user devices [105, 112, 113, 118, 126]. Power efficiency has also been considered for wireless sensor networks for agriculture [57, 68], however, existing approaches are limited to the sensing infrastructure and do not consider the effects of system inter-dependencies of an integrated data-driven and control-enabled design. An overarching approach to power efficiency would rely on adaptive duty-cycling through the ON and OFF state of communicating devices, informed by communication demand and the inherent power consumption of the underlying hardware. To this end, two urgent questions emerge: WHEN to turn power “off” and “on” and HOW to do it. To answer these questions we need to (i) characterize the power needs of individual components both in isolation and jointly, and (ii) integrate components in an energy-aware manner.

4.2 Data Management, Mining and Analytics

Extracting actionable information from farm operations sensed data is key to enabling real-time control and decision making for AG-CPS [101]. FarmNET will sense a variety of data (e.g., animal location and behavior, farming machines and farmers’ location and activity, soil state, crops and pasture growth, environmental conditions) and output, e.g., dairy, meat, grains, vegetables, etc. These diverse modalities open novel data management challenges for data acquired by noisy measurements at heterogeneous temporal scales. Moreover, it should enable analytics capabilities in a scalable manner, enabling prioritization according to the delay tolerance of specific farm operations. Another challenge is, that FarmNET will have to assist farmers in answering “what-if” questions by employing simple and accurate data-driven models that are incrementally re-trained based on new evidence. The novel aspects of our envisioned FarmNET system data analytics include (i) knowledge extraction that handles sensor uncertainty and sparse and economical sensor deployments suitable for smallholder farms (ii) support for anomaly detection, root-cause analysis and simulation using dynamic network models and (iii) flexible middle layer between sensing and control that relies on high-level states as opposed to raw sensor readings, thus completing the architecture of an end-to-end CPS.

To manage the tracking data stream and provide analytics in a centralized on-the-farm data cloud, we will adopt Apache Spark [6]. It supports various data sources (SQL, streams, etc.) and fast writes for IoT sensors’ streams, and works well with various analytics platforms (MLIB, GarphX). Beyond sensor data, we will import other data sources, such as real-time weather information and forecasts, operational data and manually collected ground truth data. The data repository will support (i) analytics tasks such as temporal trends, year-to-year comparison, anomaly detection and root cause analysis; (ii) simulation and prediction of the farm operations; and (iii) necessary data for the control component of FarmNET to enable optimized communications, sensing and operational outcomes.

Tracking uncertain mobile objects and their activity. Tracking and localization of moving objects using WSNs in various environments have seen tremendous advances, with a variety of sensors, data collection and algorithms proposed [26, 84]. In the agricultural context, such objects include animals, farmers and farming machinery. Successful and affordable tracking of the above will enable effective labor distribution and identification of bottlenecks, precision rotational grazing [80] and sick animal detection and improved farm fleet management. The case of tracking livestock presents its own challenges: large areas to cover and large number of individuals [77, 103]. There are, however, regularities in the herd behavior, e.g., shared preference for specific grass species and tendency to stay together [138], that allow for the cost-effective heterogeneous sensing WSN architecture of FarmNET. In particular, the majority of the herd members (passive) will be equipped with low-energy Bluetooth beacons (BLE), whose presence can be sensed by a small sub-sample of active herd members equipped with Bluetooth transponders, GPS, and radio to offload sensed data to available sinks.
Beyond position, we need to also track the activity of mobile objects. Activity classification of farm animal behavior (standing, grazing, laying, inactive, etc.) has been considered using accelerometers [59, 70, 89, 90, 132, 144], GPS sensors and signal strength to a gateway within a wireless sensor network [102, 104]. In particular, methods for classifying feeding behavior have achieved high accuracy: precision and sensitivity exceeding 80% for 5-10 minute windows [59]. Beyond instantaneous classification, FarmNET will incorporate predictions of the best performing techniques in a temporal smoothing framework for continuous behavioral smoothness and thus limit the effect of instantaneous mis-classification.

Tracking the state of crops and pastures. Crops, pasture grass species and their growth state comprise another important potential domain for precision farming [138]. FarmNET will predominantly rely on affordable UAV-image sensing as opposed to high-density individual plant sensors. Beyond arable land crop monitoring, which has received the highest attention from both industry and AG ICT research, FarmNET will support precision rotational grazing that requires quantifying the temporal interactions among the pasture biomass, grazing animals, water, nutrients and the effect of weather conditions [80]. Exact estimation of biomass and nutrients over time typically involves labor-intensive sampling and expensive laboratory testing [80]. Low-cost sensors have been predominantly applied to arable land [122, 137] and only recently employed for grassland [120]. Although less precise, the utility of sensing approaches involving images [34, 64, 92] laser [114] and ultrasonic [123] sensors have recently been studied for pasture measurement. Kabir et al. [79] compared the precision for grass growth estimation of ultrasonic, CCD and reflectance sensors mounted on a tractor and concluded that a camera sensor at 90 degrees angle had a dominating performance. Although capable of ensuring a good quality, continuous monitoring using tractor-mounted sensors is costly and infeasible for the temporal resolution (at least once a day) we are envisioning. Instead, FarmNET will utilize a combination of image sensors mounted on an UAV and the sensed behavior of the grazing animals, farmer activity and farming machine-mounted sensors to track the crops and pasture growth over time. Research challenges include: scheduling flight plans (altitude and spatially adaptive) to ensure a matching tracking quality to that of on-the-ground tractor-mounted sensors.

Anomaly detection and root-cause analysis. We will cast our tracking data as a dynamic graph of entities (e.g., fields, crops, animal herds, farmers, machines) with associated multivariate time series as their state over time, and edges corresponding to dependencies. Example entity signals will include: grazing over time, occupancy of animals over time, water due to precipitation and grass level over time. We will support various dynamic network analytics task building on our previous work on global network state classification [54, 55], detection of outlier regions in a single time snapshot [41, 130] and in time periods [44, 98, 99]. The multivariate nature of graph signals comprise the novel research challenges in this task. Unlike existing methods defining anomalies as simple additive score functions [44, 98, 99], we will consider dependencies among signals, leading to novel graph optimization formulations and corresponding algorithms. FarmNET will also support historical trend queries over subgraph regions of interest, de-trending from seasonal weather patterns and detection of unhealthy animals.

Anomaly detection in our dynamic graph setting [44, 98, 99] will direct attention to management zones that behave differently from the rest of the network (e.g., overgrazing, regions with soil erosion). Beyond the identification of such zones in time, we will also be able to perform root-cause analysis by examining anomalies preceding the anomalous outcome, e.g., was the weather abnormal, were animals left grazing longer than prescribed, or were they concentrated in only one location which got overgrazed? Once a suspected cause is at hand, farmers will be able to take the corresponding interventions to alleviate the situation, e.g. add fertilizer, decrease the grazing period in the enclosing paddock and others, schedule tillage or weed removal.

Supporting “What-if” questions via simulations. Detailed dynamic models for the interaction between animal herds, pastures and effects of nutrient cycles and climate have been proposed recently [131, 136]. Such models are characterized by dozens of parameters and typically make assumptions about the homogeneity of grazing behavior and grass and legume growth. Calibrating all parameters for a specific farm operation is not a trivial task. We will employ an alternative data-driven modeling based on high-resolution (both temporal and spatial) tracking data for grazing, crops and pasture growth, weather conditions and nutrient supply. We will model each field node as a queue [27] in which organic matter (grass, vegetables, legumes, etc.) arrive in the queue at a rate dependent on its current state (e.g., grass length), weather conditions and fertilizers. We will learn this independent growth rate using simple linear regression models from past observations. The herd’s grazing or crop harvesting within the same node act as a server for the queue and the rate of those processing (i.e. consuming the queue) can also be regressed based on tracking observations. Equipped with (i) the rate functions of growth and (ii) consumption we can simulate the dynamic system under different configurations. For example, we can answer question such as: What will be the state of a pasture if we doubled the herd; or shorten the grazing interval in the rotational schedule; or if we reconfigure the paddocks? What will happen to a specific crop if there is a consecutive drought period of 20 days? Being able to simulate realistically such scenarios can enable better decisions as the model will be configured based on observations from the specific farm. In addition, our models can only get more accurate if we incorporate more sensing modalities and increase sensor density.

4.3 Real–Time Farm Monitoring and Control
Knowledge of the state of pastures and crop fields in agriculture is crucial for farmers. As weather patterns change, crops mature, and cattle graze pastures, farmers rely on a combination of experience, visual observation, and intuition as to when to irrigate, apply fertilizer, or move cattle to another pasture. Their decisions are far from optimal. To enable high–output and efficient controlled–environment agriculture technologies and systems, we need a novel, holistic controlled sensing framework that will integrate IoT–sensing capabilities with agriculture data collection, network structure and humans in the loop, to enable real–time accurate monitoring and control. This objective requires (i) a stochastic dynamic system model that completely describes the farm’s state (i.e., health
and footprint) over time, while incorporating the effect of both cy-
ber (e.g., control signals) and physical components (e.g., agri-
cultural variables), (ii) recursive structured state estimators, and (iii) control
strategies that optimize estimation accuracy and costs. Prior work
on control for precision agriculture and farm monitoring systems
has separately considered the cyber and physical components and
proposed ad-hoc [28, 30, 36, 39, 40, 45, 46, 53, 58, 81, 103, 143] or
static optimization approaches [67]. An integrated approach will,
instead, be based on rigorous stochastic dynamic farm system
modeling, optimization and control system theories.

**Model and Estimators.** CPS-assisted farm management relies
on realistic modeling of its constituent components and the avail-
ability of accurate farm state estimates. We propose to model the
health and footprint evolution of a farm as a discrete-time sto-
chastic dynamical system. We will define the farm health as the
collective set of a variety of variables (e.g., soil moisture, bright-
ness temperature, vegetation biomass, soil chemical composition,
climate, production per unit of harvested area, product feed value)
that characterize the production ability of the farm and the quality
of its products. In contrast, we will define the farm footprint as a set
of variables quantifying the effect on the environment (e.g., carbon
dioxide, water consumption, demand for land to feed, soil nutrients
quality). Our model will exploit (i) the physics-based models of
farm variables’ evolution from the literature and §4.2, (ii) the struc-
tural and statistical properties of the IoT sensor models, and (ii)
the form of control mechanisms in various levels (i.e., adaptation
of sensor rates, dispatch of network infrastructure, humans in the
loop, farm processes) and the effect of network traffic’s statistical
properties (§3) and network architecture (§4.1). Beyond modeling,
we will develop low-complexity recursive Bayesian estimators of
the farm state that account for the network architecture charac-
teristics and traffic’s statistical properties, the model structure and
humans in the loop.

**Joint Optimization of Sensing, Estimation, and Control.**
FarmNET’s system design requires a controlled sensing framework
that considers all relevant information to maximize the farm’s
health, while minimizing its footprint in a dynamically changing
environment. Leveraging the dynamic farm model, the system
will automatically decide which sensor data to collect and network
infrastructure to dispatch to continuously estimate and monitor
the farm’s health and footprint. Simultaneously, it will automati-
cally control the different farm processes (e.g., tillage scheduling
and configuration, herd location scheduling, fertilizer use and opti-
mal lighting) to improve the farm’s state. Determining and imple-
menting the exact solution of the outlined optimization problem is
a computationally-expensive task. Thus, we plan to design low-
complexity, near-optimal sensing, estimation and control strategies
that will exploit properties of the AG–CPS, optimal solution
properties and farmers’ domain knowledge and feedback to enable
real-time farm monitoring and control.

5 DISCUSSION AND CONCLUSION

Today, one of the critical problems that humanity is facing is *how
to secure nutrition in the face of a changing climate?* To tackle this
problem, we need to increase agricultural production, while dramat-
ically reducing its environmental footprint. Smallholder farms that
form the backbone of agricultural production, are thus faced with
a tremendous opportunity to expand and proliferate, while solving
one of the world’s most pressing issues. This opportunity, how-
ever, comes with a list of challenges related to understanding and
optimization of farm operations. Information and communication
technology will play a critical role in such understanding by design-
ing agricultural cyber-physical systems to measure, model, analyze,
evaluate and dynamically control the farm state and operations.

Multiple solutions from industry and academia attempt to solve
this problem. Each of these two categories suffers its own inherent
limitations. Industrial products are for-profit and often strive to pro-
vide closed-form solutions. Such solutions are highly-specialized
and are not compatible with each other. This requires farmers to
purchase various non-extensible systems to provide full-profile AG
ICT support, which quickly becomes intractable and financially-
feasible in the smallholder AG context. Academic solutions for
AG ICT are often data- or sensor-centric. While such solutions
make important progress towards modeling of AG processes, they
typically address AG-CPS components in isolation and are not well-
fitted for real-time, longitudinal analysis and farm simulations.

While our preliminary analysis and corresponding design is
US-centric, we believe that the proposed solution has far-reaching
implications on smallholder agriculture in the international context.
A key factor to investigating the applicability of the proposed solu-
tions to smallholder farms in the developing world is to follow a
modular design approach both (i) horizontally: independent end-to-
end infrastructure for various operations within a policulture farm
such as animals, vegetables, grain, etc.; and (ii) vertically by ensur-
ing interoperability of varying cost/quality sensing, connectivity
and analytics solutions. We see such a design central to further
understand the technological, farmer utility and cost challenges in
various farmer enterprises. This tradeoff between cost and utility as
well as in-depth understanding of the specific needs of farmers in
rural areas of the developing world is of key importance when cus-
tomizing AG-CPS systems for those settings. Our modular design
and evaluation of the utility of individual modules in collaboration
with Essex Farm will shed light to the above questions of feasibility
to the developing world.

Our research makes important headway towards integrated, end-
to-end AG-CPS by providing in-depth analysis of farm ICT de-
mand. Our findings indicate that farm demand is a unique mixture
of farmer- and sensor-generated traffic with interlocking charac-
teristics. In terms of spatial distribution, farmer traffic is highly-
localized, whereas sensor traffic is distributed. Furthermore, while
farmer traffic is bursty, downlink-intensive and unpredictable, sen-
stor traffic is periodic, uplink-intensive and predictable. These find-
ings create a unique design space for an integrated AG-CPS dubbed
FarmNET. FarmNET consists of three interdependent components
including (i) sensing and communication, (ii) data analytics and
(iii) AG monitoring and control. In this vision paper, we outlined
key functional blocks of each component, surveyed the state of the
art and outlined an agenda for future development.

An integrated system such as FarmNET is essential to enable
practical ICT innovation in agriculture. Such system requires an
interdisciplinary approach that brings expertise from sensing and
wireless networks, data science, and estimation and control with a
constant practitioner feedback in the loop.