# STAC: Simultaneous Transmitting and Air Computing in Wireless Data Center Networks

Xiugang Wu, *Member, IEEE*, Shengli Zhang, *Member, IEEE*, and Ayfer Özgür, *Member, IEEE*

***Abstract*— The data center network (DCN), wired or wireless, features large amounts of many-to-one (M2O) sessions. Each M2O session is currently established using point-to-point (P2P) communications and store-and-forward (SAF) relays, and is generally followed by a certain computation at the destination, typically a weighted summation of the received information digits. Fundamentally different from this separate P2P/SAF-based-transmission and computation framework, this paper proposes simultaneous transmission and air computation (STAC), a novel physical layer scheme that achieves STAC in wireless DCNs. In particular, STAC builds on a number of distinguishing characteristics of DCs to take advantage of the superposition nature of electromagnetic signals. With STAC, multiple sources transmit in the same time slot with appropriately chosen parameters, such that the superimposed signal can be directly transformed to the desired summation at the receiver. To enable STAC, we propose an enhanced software-defined network architecture, where a wired low-bandwidth backbone provides wireless transceivers with external reference signals. We also discuss some new challenges that STAC brings to scheduling and routing. Theoretical analysis and simulation results show that STAC can significantly improve both bandwidth and energy efficiency in DCNs.**

***Index Terms*— Data center network, green communications, software defined network, wireless network.**

## I. Introduction

A MODERN Data Center (DC) typically consists of a large dedicated cluster of commercial computers (work nodes) that are housed together to store/process big files in a parallel manner. The parallel storage/processing of the files necessitates frequent communication among the work nodes, which is accomplished through the Data Center Network (DCN). Despite the maturity in deployment and their high bandwidth, there are a few critical problems associated with wired DCNs, such as limited flexibility, cabling complexity, device cost, over subscription, etc. These problems significantly limit the scalability of DCNs and are being exacerbated by the huge amount of data that needs to be stored/processed within a DC and exchanged through the DCN in today's age of big data.

To overcome these limitations, recent works [2]–[7] have explored the possibility of constructing wireless DCNs using high frequency electromagnetic (EM) signalling. Although the data rate of wireless links is still currently lower than that of wireline links, wireless transmission techniques have been under rapid development, and particularly, 60 GHz techniques have been proposed for realizing wireless DCN links with bandwidth comparable to wireline connections in [2] and [3], while the blockage and directivity problems associated with EM waves can be significantly mitigated by utilizing strategies like ceiling reflection and 3D beamforming [4]. Free-space optical DCN communication has also been investigated [5], and shown to achieve further improvements in bandwidth, as well as nearly perfect directivity which helps to significantly reduce interference in the wireless DCN and protect the up layer rate from being dramatically decreased. On the other hand, [6] considers augmenting the wired DCN with wireless flyways, and [7] demonstrates that a completely wireless DCN with a Cayley structure is feasible and performs even better than a wired DCN.

### *A. Challenging the P2P and SAF Paradigm*

This paper aims to further the research of wireless DCNs by challenging two of its fundamental assumptions, namely, Point-to-Point (P2P) communication and Store-and-Forward (SAF) relaying. These two paradigms form the basis of today's ubiquitous wireless networks, e.g., cellular and 802.11 networks, and are also widely adopted as the assumptions in the literature on wireless DCNs [8], [9]; however, here we argue that they may be highly suboptimal for DCNs due to a number of distinguishing characteristics of the traffic and the computations typically occurring within DCs.

First, the DCNs feature a large number of Many-to-One (M2O) sessions arising from various DC applications, e.g., Google File System (GFS) [10] and MapReduce [11]. Due to the limited transmission range of high frequency EM waves, these M2O sessions are established through multi-hopping over hierarchical multiple-access units [12], [13] as shown in Fig. 1, where each hop is based on P2P communication followed by SAF relaying onto the next hop. Specifically, with Time Division Multiple-Access (TDMA) inside the multiple-access unit, the source nodes $1, 2, \ldots, K$ successively transmit their information to the relay node 0 in different time slots via P2P transmissions, and the relay stores the received information in its buffer before forwarding them to the destination $d$. Since node 0's buffer and input/output

Manuscript received February 1, 2016; revised July 18, 2016; accepted August 23, 2016. Date of publication September 20, 2016; date of current version December 29, 2016. This work was supported in part the National 973 Project under Grant 2013CB336700, in part by the Chinese NSF Project under Grant 61372078, and in part by the Shenzhen NSF Project under Grant JCYJ20160226192223251. This paper was presented at the IEEE/CIC International Conference on Communications in China, Shenzhen, China, November 2015 [1]. *(Corresponding author: Shengli Zhang.)*

X. Wu and A. Özgür are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: x23wu@stanford.edu; aozgur@stanford.edu).

S. Zhang is with the Faculty of Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: zsl@szu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2016.2611838

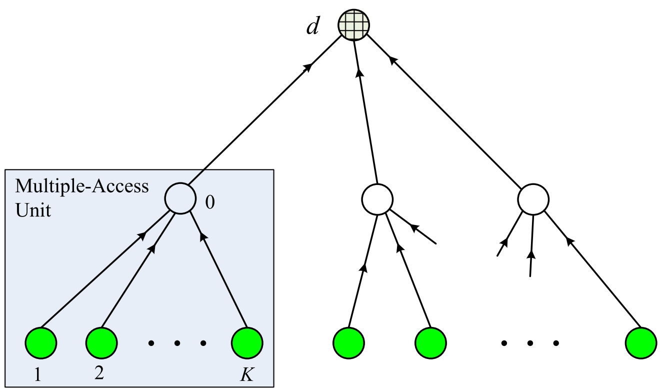



Fig. 1. Illustration of one M2O session. Source nodes (solid circles) transmit information to destination $d$ via relay nodes (hollow circles).

bandwidth are shared by all the $K$ source nodes, the performance of this approach can be poor, especially when $K$ is large. Indeed, the problem gets worse as we get closer to the destination, since the information to be transmitted accumulates along the way.

Another important characteristic of DCs is that the M2O sessions are generally followed by certain computations at the destination nodes. These computations normally satisfy the *commutative* and *associative* operational laws, with weighted summation being the typical case (e.g., in linear network coded storage [14] and MapReduce-based machine learning [15] applications). This opens up the possibility of dividing a large computation task into several sub-tasks that can be conducted at the intermediate relay nodes, rather than demanding the final destination do all the computation. In other words, instead of forwarding all the received digits, the relay could perform some intermediate computation and then forward only the output of the computation, thereby utilizing the bandwidth more efficiently. Considering that the bottleneck lies in the DCN scalability, not the compute capabilities of the works nodes, we believe that such a Compute-and-Forward (CAF) approach can be highly preferable to traditional SAF relaying in DCNs.

### B. A New Scheme: STAC

Based on the above observations, two natural suggestions to improve the performance of P2P and SAF based DCNs can be the following:

1) The sources can transmit their digits in the *same* time slot, and the receiver can recover the digits based on *multi-user detection*.
2) The relay can perform *computation* (weighted summation typically) of the recovered digits, and then forward the *computation result* onto the next hop.

In this paper, we go one step further. We propose a new physical layer scheme, dubbed STAC (Simultaneous Transmission and Air Computation), which takes advantage of the *superposition* nature of EM signals and can achieve the above described two functions simultaneously over the air.

With computation task division, it suffices to illustrate STAC for a particular multiple-access unit as depicted in Fig. 1. Suppose that the receive node 0 is only interested in the weighted summation $s_0$ of the $K$ source digits $s_1, s_2, \ldots, s_K$,

$$s_0 = \sum_{i=1}^{K} w_i s_i, \qquad (1)$$

where $w_1, w_2, \ldots, w_K$ are the weight coefficients which are assumed to be real integers throughout this paper. In STAC, the $K$ source nodes transmit their digits in the same time slot with appropriately chosen transmit powers, frequencies, phases and times, such that their information bearing EM signals arrive at node 0 in a desired *superimposed* form that can be transformed to $s_0$ directly. We will show that STAC significantly improves bandwidth and energy efficiencies over the separation based approach for transmission and computation as it effectively allows to communicate only the desired computation and not the constituent digits to the receiver. Note that given $s_1, s_2, \ldots, s_K$, we can always find $s_0$ but not vice versa, so the information content of $s_0$ can be much smaller than the total information contained in $s_1, s_2, \ldots, s_K$. Additionally, in the general case when node 0 needs to fully recover the original $K$ source digits, e.g., for performing some computation other than weighted summation, one can still apply STAC by properly designing a set of *pseudo* coefficients $\{w_1, w_2, \ldots, w_K\}$ such that the original digits $s_1, s_2, \ldots, s_K$ can be extracted from the received $s_0$. (In other words, STAC subsumes multi-user detection as a special case.) We will show that in this pseudo coefficients case, for fixed SER (Symbol Error Rate) and bandwidth efficiency, STAC achieves better energy efficiency than the separation based approach.

To enable STAC, accurate channel state information (CSI) and perfect frequency/time synchronization among the transceivers are needed, both of which may be difficult to obtain in general wireless networks. Thanks to the indirect ceiling-reflected Line of Sight (LoS) channel [4] and fixed locations of the transceivers, however, the CSI in a DC is nearly time-invariant and can be accurately estimated. To accomplish the synchronization, on the other hand, this paper proposes to use wired connections among all the work nodes to provide the wireless transceivers with external reference signals (e.g., a high quality external clock signal) [16], based on an enhanced Software Defined Network (SDN) architecture [17].

It should be pointed out that, the wired connections here are distinguished from the information transmission links in a wired DCN. The former are dedicated and solely responsible for control signals, not requiring the high bandwidth as in the latter, and thus will not cause the aforementioned problems encountered by wired DCNs. We also remark that to build up such a wired control network in DCs is plausible considering that the work nodes are usually compactly piled up in a dedicated room of limited size. As a by-product, it will also reduce the DCN operation cost by eliminating the need of using individual oscillators at the transceivers.

Of course, as a fundamental physical layer change, introducing STAC will bring new challenges to the network layer scheduling and routing. In fact, as we will see, with STAC, the optimal M2O routing becomes a minimum Steiner tree problem. Since this problem is known to be NP-hard, we

**Algorithm 1** Network Coded Recovery

1: $s_0 = \sum_{i=1}^{K} w_i s_i$
2: $s_0 \leftarrow s_0 \bmod 2^q$

develop a greedy single-M2O session routing algorithm, and use it in the case of multiple M2O sessions to demonstrate the network gain brought by STAC.

The remainder of the paper is organized as follows. First, Section II presents some motivating examples, followed by a detailed description of the STAC system architecture in Section III. Then, Sections IV and V discuss the technical issues and performance improvements at the physical layer and the network layer, respectively. Finally, Section VI concludes the paper and proposes some future research directions.

## II. Motivating Examples

Two major DC applications are i) distributed file storage, e.g., GFS [10] and Hadoop Distributed File System (HDFS) [18], and ii) parallel big data processing based on the MapReduce model [11]. We now present three detailed DC application examples mentioned in Section I that motivate STAC, where the first two correspond to GFS and MapReduce, respectively, and the last one shows the flexibility of STAC for general applications. Again, with task division, we can concentrate our discussion on the multiple-access unit depicted in Fig. 1.

### A. Network Coded Storage

Due to the non-negligible node failures in a DC [10], in distributed storage systems, a big file is usually divided into many fixed-length data blocks that are further protected by multiple replicas stored at different work nodes.

For storage efficiency, a network code can be applied [14], [19], [20], where each node stores network coded blocks instead of raw data. When a coded block is lost due to a node failure, it can be reconstructed at another node by performing Algorithm 1 on a digit-by-digit basis, where $s_0$ denotes a digit from the lost coded block requiring recovery, $s_1, s_2, \ldots, s_K$ are digits from the data blocks stored at the other nodes, $w_1, w_2, \ldots, w_K$ are the network coding coefficients, and the modulo operation is due to the finite field size $2^q$. Clearly, with STAC, we can achieve Step 1 of the algorithm directly.

### B. MapReduce Based Data Processing

In MapReduce model, when the map nodes finish the processing, their outputs with the same key will be sent to a specified reduce node for the final computations. Such computations are also typically in the form of weighted summations [11], [21], e.g., for all machine leaning algorithms fitting the statistical query model [15], scientific processes [22], [23], documents similarity comparisons [24], etc. Again, STAC can be applied to achieve simultaneous transmission and computation here.

**Algorithm 2** Source Digits Extraction

```
i ← 1
while i ≤ K do
    s_i ← s_0 mod 2^q
    s_0 ← (s_0 − s_i)/2^q
    i ← i + 1
end while
```

### C. General Case

When the receive node 0 needs the original source digits, one can appropriately design a set of pseudo coefficients $\{w_1, w_2, \ldots, w_K\}$ such that the source digits $s_1, s_2, \ldots, s_K$ can be extracted from $s_0$. In particular, suppose for each $i = 1, 2, \ldots, K$,

$$0 \le s_i \le 2^q - 1,$$

then choosing $w_i = 2^{q(i-1)}$ yields

$$s_0 = \sum_{i=1}^{K} 2^{q(i-1)} s_i,$$

based on which all the source digits can be extracted with Algorithm 2.

To conclude this section, we point out that STAC can have a wide range of applications in addition to the above examples. For example, although our motivation for proposing STAC mainly comes from M2O sessions that include the above examples as special cases, it can also be applied to P2P or M2M (Many-to-Many) situations with the idea of combining or splitting sessions as discussed in [25]. Also, note that although this paper focuses on wireless DCNs, the application of STAC may not be restricted to DCs. In fact, it can be applied to any general wireless networks with MapReduce or Hadoop applications [26], or wireless sensor networks where the sensor nodes need to finish some computation task in a cooperative way [27].

## III. System Architecture With STAC

This section describes the system architecture with STAC. We first introduce the principle of a basic STAC unit, and then propose an enhanced SDN architecture that enables the functioning of DCNs with STAC units.

### A. A Basic STAC Unit

STAC is a general physical layer scheme that can be applied to wireless DCs with any structure, carrier frequency, etc. For illustration, consider a typical layout of the wireless DC as shown in Fig. 2, where each rack contains multiple work nodes and has an antenna array mounted on its top to communicate with other racks (communications within a rack are accomplished with intra-rack connections) [7]. As in [4], ceiling-reflecting and 3D beamforming techniques are adopted to achieve an indirect LoS link between any two antenna arrays without causing interference to others.

Suppose $K$ work nodes (in $K$ different racks) need to transmit their digits $s_1, s_2, \ldots, s_K$ to node 0 for computing

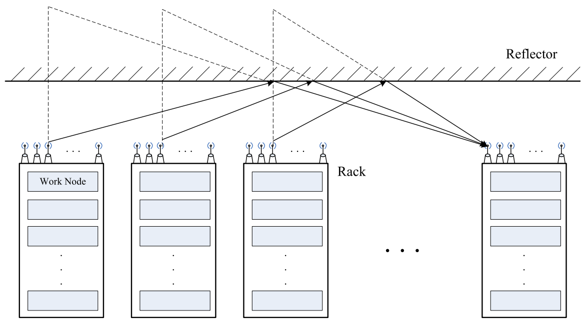



Fig. 2. A typical wireless DC layout.

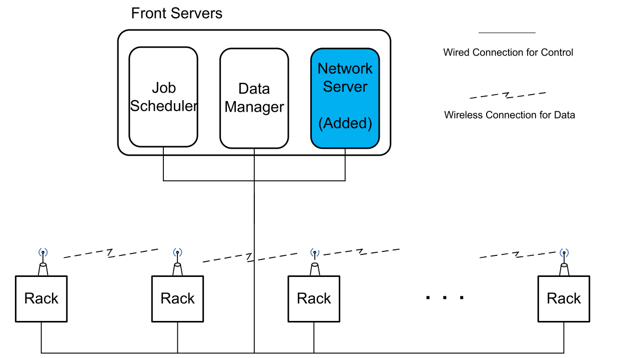



Fig. 3. An enhanced SDN architecture.

the weighted summation as in (1). The operating principle of STAC is illustrated in the following.

Each source node $i$ maps its digit $s_i$ to a baseband modulated complex symbol $d_i$, and then up converts the symbol $d_i$ to a passband signal given by

$$\sqrt{P_i}e^{-j\theta_i}d_i(t)e^{-jf_ct},$$

where $\theta_i$ and $\sqrt{P_i}$ are the pre-equalizing phase and amplitude coefficients, respectively. Suppose each node $i$ transmits at time $t_i$ using 3D beamforming, then the received passband signal $y(t)$ can be expressed as

$$\sum_{i=1}^{K} h_i e^{j\theta'_i}\sqrt{P_i}e^{-j\theta_i}d_i(t-t_i-\tau_i)e^{-jf_c(t-t_i-\tau_i)}+n(t)$$

where $h_i e^{j\theta'_i}$ is the equivalent complex channel coefficient from node $i$ to 0, $\tau_i$ is the propagation delay for node $i$, and $n(t)$ is a Gaussian noise of variance $\sigma^2$ for both the real and imaginary dimensions. With accurate CSI, one can set

$$\theta'_i = \theta_i \text{ and } t_i = t_0 - \tau_i, \tag{2}$$

such that the received signal simplifies to

$$y(t) = \sum_{i=1}^{K} h_i\sqrt{P_i}d_i(t-t_0)e^{-jf_c(t-t_0)}+n(t),$$

which, after down conversion and sampling at time $t = t_0$, yields the baseband symbol[1]

$$y = \sum_{i=1}^{K} h_i\sqrt{P_i}d_i + n, \tag{3}$$

where $n \sim \mathcal{N}(0,\sigma^2)$. Clearly, if each node $i$ sets

$$P_i = (w_i/h_i)^2, \tag{4}$$

then after eliminating the noise, node 0 can construct the desired digit $s_0$ as in (1) from the symbol $y$ in (3).

[1]The $h_i$ in (3) are real variables, so that the real and imaginary parts of symbol $y$ can be separated. In this paper, we only consider the real part for simplicity. In general, complex signals are more complicated when superimposed. However, we assume phase synchronization in STAC, where all the complex signals are superimposed in a synchronized version. In other words, the real parts and the imaginary parts of the signals are summed separately. By precoding the imaginary part and real part with their own calculated coefficients, two summed symbols can be obtained at the receiver, from the real and imaginary parts respectively.

With the above described principle, one can see that the time/frequency synchronization and pre-equalization, such as (2) and (4), are essential for STAC, and synchronization failure and channel impairments may severely degrade the performance of STAC. These synchronization and pre-equalization issues may not be easy to resolve in general [28], [29], however in DCs the transceiver and environment are typically static, and there is only LoS (Line-of-Sight) propagation path, i.e., no reflection, no multi-path and no interference. Under this circumstance, one can relatively easily achieve the synchronization and pre-equalization by using an enhanced SDN architecture as we discuss next.

### *B. An Enhanced SDN Architecture*

The DC generally work based on centralized control, where the front servers, including the job scheduler and data manager, control all the work nodes. In current DCNs, control signals and data traffic share the same network. Here, we propose to use a dedicated low bandwidth wired control network with an added network server as shown in Fig. 3, based on an enhanced SDN architecture. As mentioned in Section I, the wired control network is feasible due to limited DC size and fixed node locations.

Our SDN architecture is an enhanced one in the sense that, it not only accomplishes networking control as in general SDNs, but also also provides the wireless transceivers the physical and upper layer configurations to enable STAC, including the synchronization information, the physical layer parameters such as powers, frequencies, phases and times, and the scheduling/routing information.

Synchronization with External Reference Signals: External reference signals are provided to all the transceivers for synchronization. These include a high quality external clock signal, with which individual crystal oscillators at the transceivers are no longer needed and the operation cost can be thereby reduced. These reference signals can also help calibrate the wireless transceivers, e.g., reduce the errors induced from the device hardware differences [16].

Physical Layer Parameters: The network server maintains a connection information table that stores important physical layer parameters for each connection, such as the transmission delay $\tau$, channel coefficient $he^{-j\theta}$ and the steering vectors required for 3D beamforming. When a transceiver fails (or a new one comes in), it informs the network server through the control network and the connection information table is updated.

Scheduling/Routing: Also maintained by the network server is a table storing the scheduling/routing information. When a current task finishes or a new one needs to start, the job scheduler informs the network server to update the scheduling/routing information table. Then the network server initiates the needed coordinations between the work nodes involved.

*Difference From Traditional SDN:* Note that the traditional SDN, which centralizedly control the routings and switches in the network, has already been deployed in current wired DCN to manage the data flows at the network layer. In STAC, the physical layer transmissions are much more complex and important, and we extend the concepts of SDN to the physical layer. Specifically, there are two fundamental extensions. i) The OpenFlow [30] interface between the controller and the node (end node and routing node) should be extended to include the physical transmission parameters, such as the transmission time, coefficients, beamforming direction and so on. The physical parameters can be initially obtained when a wireless transceiver joins the network with measuring the channels between the transceiver and its neighbors (there is a direct link between them). Note that the complexity of this procedure is fixed for any size of network as the neighbor number is almost fixed and the channel conditions are almost static. In OpenFlow, the flows are first identified by matching to their head information before they are forwarded accordingly. Both the head information and the forward rules are pre-determined by the controller. In our scheme, we need to add the physical layer head information, such as transceiver address and channel into the match field. We also need to add the physical transmission parameters into the forward rule. ii) The controller should provide synchronization signals, such as an external clock signal, to provide a master-slave type synchronization service to all the wireless transceivers. Specifically, the external clock signal can help to guarantee synchronized frequency and clock steps. The controller can also periodically send time synchronize packets in a master-slave manner as in [31] to obtain initial time synchronization and to deal with other synchronization errors.

Another feature of our enhanced SDN is that we use wireless channels for the data transmission and wired channel for the control network. This is because control information is typically light but periodical and static, and therefore we can simply connect the transceivers directly with wirelines without using expensive and high-energy-cost switches. On the other hand, wireless transmissions, with increasing rate and unlimited bandwidth inside the DC, are more appropriate for heavy and random data traffics.

## IV. Physical Layer Issues

### *A. Modulation-Demodulation Mapping*

The modulation for STAC is the same as that for P2P channels. However, their demodulation mappings are subtly different: STAC demodulation maps a superimposed symbol, which may even not belong to the transmit symbol sets, to the summation of the digits, whereas the P2P channel demodulation maps a particular symbol from the transmit symbol set to the corresponding digit.

*1) STAC Modulation:* Specifically, writing node $i$'s digit $s_i$ into the bit sequence form yields

$$[s_i(1), s_i(2), \ldots, s_i(l), \ldots, s_i(L)]$$

where $s_i(l)$ is the $l$-th bit, $L$ is the sequence length, and

$$s_i = \sum_{l=0}^{L-1} 2^l s_i(l).$$

For modulation, throughout this paper we assume BPSK (Binary Phase Shift Keying)[2] is used without error correction coding. At node $i$, each bit $s_i(l)$ is modulated to a symbol $d_i(l) \in \{-1, +1\}$ as $d_i(l) = 1 - 2 \times s_i(l)$.

*2) STAC Demodulation:* After the $l$-th transmission and the removal of noise with signal detection, the received superimposed symbol can be written as

$$y(l) = \sum_{i=1}^{K} h_i \sqrt{P_i} d_i(l) \tag{5}$$

By setting the transmit power[3] $P_i = (w_i/h_i)^2$, one has

$$y(l) = \sum_{i=1}^{K} w_i d_i(l), \tag{6}$$

which, through the operation

$$\frac{1}{2}\left(\sum_{i=1}^{K} w_i - y(l)\right),$$

yields the summation $\sum_{i=1}^{K} w_i s_i(l)$.

Finally, the desired digit can be constructed as

$$\sum_{l=0}^{L-1} 2^l \sum_{i=1}^{K} w_i s_i(l) = \sum_{i=1}^{K} w_i \sum_{l=0}^{L-1} 2^l s_i(l) = \sum_{i=1}^{K} w_i s_i.$$

### *B. Signal Detection*

We now present a simple signal detection scheme for removing the noise in (3) to obtain (5), and analyze its corresponding SER. It suffices to consider only one of the $L$ transmissions, and hence the index $l$ as in the last subsection will be omitted.

Specifically, view the symbol $\sum_{i=1}^{K} w_i d_i$ in (6) as a point of a non-standard PAM (Pulse Amplitude Modulation) constellation that results from the weighted superposition of the transmit BPSK constellations and hence may have unequal distance between different adjacent constellation points. A simple detection scheme is to quantize the $y$ in (3) to its nearest constellation point. Let $\pi$ be a permutation on $\{1, 2, \ldots, K\}$

[2] STAC also applies with other modulations such as QPSK, QAM, OOK, OFDM, etc. This paper only considers the simplest BPSK due to the reason mentioned in Footnote 1.

[3] With the unit power of $d_i$ in BPSK, the transmit power $P_i|d_i|^2$ simply equals $P_i$.

such that $w_{\pi(j_1)} \leq w_{\pi(j_2)}$, $\forall j_1 \leq j_2$. We have the following theorem regarding the SER with such detection.

*Theorem 1:* The SER with the nearest point detection is upper bounded by

$$\mathrm{SER}_{\mathrm{STAC}} \leq \left(1 - \frac{1}{2^K}\right) \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right) \tag{7}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt$ is the complementary error function, $\sigma^2$ is the variance of the noise, and the equality in (7) holds when the distance between any two adjacent constellation points is equal to 2, e.g., when $w_{\pi(j)} = 2^{j-1}$ or 1, $\forall j = 1, \ldots, K$.

*Proof:* Recall that we assume BPSK modulation, and $d_i \in \{-1, 1\}, \forall i = 1, 2, \ldots, K$, denotes the transmit symbol of node $i$. The receive constellation for the superimposed symbol $\sum_{i=1}^{K} w_i d_i$ results from the weighted superposition of the transmit BPSK constellations of the $K$ nodes, which depends on the choice of $\{w_1, w_2, \ldots, w_K\}$.

Since $w_i$ are all real integers, one can easily check that the distance between any two adjacent points in the receive constellation is at least 2. When this distance is exactly 2, e.g., when $w_{\pi(j)} = 2^{j-1}$ or $1, \forall j = 1, \ldots, K$, the SER attains its maximum value. In this case, the SER for the detection of the leftmost or rightmost point is

$$\begin{aligned} \mathrm{SER}_1 &= \Pr(n \geq 1) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_1^\infty e^{-t^2/(2\sigma^2)} \, dt \\ &= \frac{1}{2} \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right) \end{aligned}$$

where $n$ is the receive noise as in (3). For the detection of the other points in the receive constellation, the SER is given by

$$\begin{aligned} \mathrm{SER}_2 &= \Pr(n \leq -1) + \Pr(n \geq 1) \\ &= 2 \times \Pr(n \geq 1) \\ &= \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right). \end{aligned}$$

Finally, noting that the leftmost (or rightmost) constellation point happens if and only if all $d_i$ take value $-1$ (or $+1$), which is with probability $\frac{1}{2^K}$, we obtain the following upper bound on the average SER of STAC:

$$\begin{aligned} \mathrm{SER}_{\mathrm{STAC}} &\leq \frac{2}{2^K} \times \mathrm{SER}_1 + \left(1 - \frac{2}{2^K}\right) \times \mathrm{SER}_2 \\ &= \left(1 - \frac{1}{2^K}\right) \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right), \end{aligned}$$

which proves the theorem. ■

Note that the nearest point detection is in general *suboptimal* due to the possible different weight coefficients. Nevertheless, in the next we will show that even with this suboptimal detection, STAC outperforms the separate P2P-based transmission and computation strategy, where the latter will also be referred to as separate strategy for convenience in the sequel discussion.

### C. Performance of STAC

The performance of STAC is a tradeoff among SER, energy efficiency and bandwidth efficiency, and is clearly dependent of the weight coefficients. The *air computation* essence of STAC and its advantage over the separate strategy can be best illustrated in the ideal case of $w_1 = w_2 = \cdots = w_K = 1$, where we will show that for fixed energy efficiency, STAC always achieves better SER in the high SNR regime and significantly improved bandwidth efficiency.

On the other hand, to show that STAC uniformly outperforms the separate strategy, we will consider the pseudo coefficients case as mentioned in Section II, i.e., $w_{\pi(j)} = 2^{j-1}, \forall j$. The argument here is that by applying STAC with the pseudo coefficients, one can recover the original $K$ source digits, based on which summation with any weight coefficients can be computed. We will show that in this case, STAC achieves better energy efficiency for fixed SER and bandwidth efficiency.

*1) The Ideal Case:* Suppose $w_1 = w_2 = \cdots = w_K = 1$. The SER of STAC is given in Theorem 1, i.e.,

$$\mathrm{SER}_{\mathrm{STAC}} = \left(1 - \frac{1}{2^K}\right) \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right).$$

Note that in this case, the resultant receive PAM constellation has only $K + 1$, instead of $2^K$, points,[4] where the decrease of the constellation size is due to the "air computation". Or equivalently, viewed from the energy perspective, this advantage is reflected by the fact that the needed transmit power now attains the minimum $P_i = 1/h_i^2$ for each node $i$.

For the separate strategy, assume each node $i$ transmits with the same power $P_i = 1/h_i^2$ as in STAC. The SER for node $i$ is a standard result, given by

$$\frac{1}{2} \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right).$$

Combining all the detected $K$ symbols, the receiver computes $\sum_{i=1}^{K} w_i d_i$, and the resultant $\mathrm{SER}_{\mathrm{SEP}}$ can be lower bounded as in the following theorem.

*Theorem 2:* The SER with the separate strategy is lower bounded by

$$\mathrm{SER}_{\mathrm{SEP}} \geq \frac{1}{2} - \frac{1}{2}\left(1 - \mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right)\right)^K.$$

*Proof:* With separate symbol detection, the SER of each symbol is $\frac{1}{2}\mathrm{erfc}(\frac{1}{\sqrt{2}\sigma})$ for BPSK modulation. Since the number of erroneous symbols is a binomial random variable with parameters $(K, \frac{1}{2}\mathrm{erfc}(\frac{1}{\sqrt{2}\sigma}))$ and the computation result is wrong if there are odd number of erroneous symbols, we have

[4] In particular, the resultant receive PAM constellation is $\{-K, -K+2, \ldots, -1, 1, \ldots, K-2, K\}$ when $K$ is odd, and $\{-K, -K+2, \ldots, -2, 0, 2, \ldots, K-2, K\}$ when $K$ is even. Note that in either case, the size of the constellation is $K + 1$.

that $\mathrm{SER}_{\mathrm{SEP}}$ can be lower bounded by

$$\begin{aligned}\mathrm{SER}_{\mathrm{SEP}} \geq & \sum_{i=0}^{\lfloor K/2\rfloor-1} \binom{K}{2i+1}\left(\frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right)\right)^{2i+1} \\ & \times\left(1-\frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right)\right)^{K-2i-1} \\ = & \frac{1}{2}-\frac{1}{2}\left(1-\mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\right)\right)^{K}.\end{aligned}$$

■

We now show that in the ideal case, for fixed energy efficiency, STAC always achieves better SER than the separate strategy for sufficiently large SNR, in particular, when $\mathrm{erfc}(\frac{1}{\sqrt{2}\sigma}) < \frac{1}{2}$.

*Theorem 3:* $\mathrm{SER}_{\mathrm{SEP}} > \mathrm{SER}_{\mathrm{STAC}}$ for any $K \geq 2$ provided $\mathrm{erfc}(\frac{1}{\sqrt{2}\sigma}) < \frac{1}{2}$.

*Proof:* We prove the theorem by mathematical induction. First consider the case of $K = 2$. Denoting $\mathrm{erfc}(\frac{1}{\sqrt{2}\sigma})$ by $p \in (0, 0.5)$, we have

$$\mathrm{SER}_{\mathrm{SEP}} \geq \frac{1}{2}-\frac{1}{2}(1-p)^2,$$

and

$$\mathrm{SER}_{\mathrm{STAC}} = \frac{3}{4}p.$$

Therefore,

$$\begin{aligned}\mathrm{SER}_{\mathrm{SEP}} - \mathrm{SER}_{\mathrm{STAC}} &\geq \frac{1}{2}-\frac{1}{2}(1-p)^2-\frac{3}{4}p \\ &= \frac{1}{4}p(1-2p) \\ &> 0.\end{aligned}$$

Now assume that

$$\frac{1}{2}-\frac{1}{2}(1-p)^K > \left(1-\frac{1}{2^K}\right)p. \tag{8}$$

Note that this has been verified for the case when $K = 2$ and we will now prove that given (8) holds, one must also have

$$\frac{1}{2}-\frac{1}{2}(1-p)^{K+1} > \left(1-\frac{1}{2^{K+1}}\right)p.$$

For this, consider the following:

$$\begin{aligned}&\frac{1}{2}-\frac{1}{2}(1-p)^{K+1}-\left(1-\frac{1}{2^{K+1}}\right)p \\ &= \frac{1}{2}-\frac{1}{2}(1-p)^K(1-p)-\left(1-\frac{1}{2^K}+\frac{1}{2^K}-\frac{1}{2^{K+1}}\right)p \\ &= \frac{1}{2}-\frac{1}{2}(1-p)^K-\left(1-\frac{1}{2^K}\right)p+\frac{1}{2}(1-p)^K p \\ &\quad -\left(\frac{1}{2^K}-\frac{1}{2^{K+1}}\right)p \\ &> \frac{1}{2}(1-p)^K p-\left(\frac{1}{2^K}-\frac{1}{2^{K+1}}\right)p \qquad (9) \\ &= \frac{1}{2}(1-p)^K p-\frac{1}{2^{K+1}}p \\ &> 0, \qquad (10)\end{aligned}$$

where (9) follows from the assumption (8), and (10) follows since $p < \frac{1}{2}$. Since we already prove that (8) holds for $K = 2$, it can be concluded that (8) holds for any $K \geq 2$, and thus $\mathrm{SER}_{\mathrm{SEP}} > \mathrm{SER}_{\mathrm{STAC}}$. ■

Therefore, STAC achieves a better SER in the high SNR regime and simultaneously improves the bandwidth efficiency by a factor of $K$. Especially, note that as $K \to \infty$, $\mathrm{SER}_{\mathrm{STAC}} \to \mathrm{erfc}(1/\sqrt{2}\sigma)$ whereas $\mathrm{SER}_{\mathrm{SEP}} \to 1/2$! Fig. 4 plots $\mathrm{SER}_{\mathrm{STAC}}$ and $\mathrm{SER}_{\mathrm{SEP}}$ for $K = 4$.

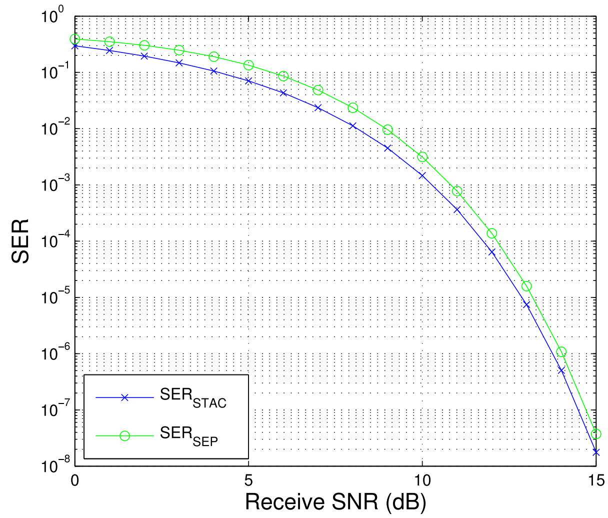



Fig. 4. SER comparison ($K = 4$).

*2) Pseudo Coefficients Case:* Consider a set of pseudo coefficients $w_{\pi(j)} = 2^{j-1}, \forall j$. To minimize the total transmit power $\sum_{i=1}^{K}(w_i/h_i)^2$ with STAC, we allocate these coefficients among the $K$ nodes such that $h_{\pi(j_1)} \geq h_{\pi(j_2)}, \forall j_1 \leq j_2$. Assuming STAC is completed within unit time, the total transmit energy $E_{\mathrm{STAC}}$ is given by

$$E_{\mathrm{STAC}} = \sum_{j=1}^{K}\left(2^{(j-1)}/h_{\pi(j)}\right)^2. \tag{11}$$

We now calculate the total energy needed $E_{\mathrm{SEP}}$ for the separate strategy assuming that each node transmits 1 bit to the receiver within $1/K$ time to maintain the same bandwidth efficiency as STAC. For the separate strategy to achieve the similar SER as STAC, the distance between any adjacent receive constellation points also needs to be 2, in which case node $i$'s transmit power is given by

$$P_i = \sum_{j=1}^{K}\left(2^{(j-1)}/h_i\right)^2.$$

Therefore, the total energy needed is

$$E_{\mathrm{SEP}} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(2^{(j-1)}/h_i\right)^2 \tag{12}$$

where the factor $1/K$ accounts for the transmission time of each node.

*Theorem 4:* $E_{\mathrm{SEP}} \geq E_{\mathrm{STAC}}$, where the equality holds only when $h_i$ are the same for all $i$.

*Proof:* The proof utilizes the important fact that $w_{\pi(j_1)} \le w_{\pi(j_2)}$ and $h_{\pi(j_1)} \ge h_{\pi(j_2)}$, $\forall j_1 \le j_2$, and is detailed as follows.

$$\begin{aligned}
K E_{\text{SEP}} \quad & \quad (13)\\
&= \sum_{i=1}^{K} 1/h_i^2 \sum_{j=1}^{K} \left(2^{(j-1)}\right)^2 \\
&= \sum_{i=1}^{K} 1/h_{\pi(i)}^2 \sum_{j=1}^{K} \left(2^{(j-1)}\right)^2 \\
&= \sum_{i=1}^{K} \sum_{j=1}^{K} 1/h_{\pi(i)}^2 \left(2^{(j-1)}\right)^2 \\
&= \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j \ne i} \frac{2^{2(j-1)}}{h_{\pi(i)}^2} \\
&= \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=1}^{i-1} \frac{2^{2(j-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \frac{2^{2(j-1)}}{h_{\pi(i)}^2} \quad (14)\\
&= \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \left[\frac{2^{2(i-1)}}{h_{\pi(j)}^2} + \frac{2^{2(j-1)}}{h_{\pi(i)}^2}\right] \quad (15)\\
&\ge \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \left[\frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \frac{2^{2(j-1)}}{h_{\pi(j)}^2}\right] \quad (16)\\
&= \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \frac{2^{2(j-1)}}{h_{\pi(j)}^2} \\
&= \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} (K-i) \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} \sum_{j=i+1}^{K} \frac{2^{2(j-1)}}{h_{\pi(j)}^2} \\
&= \sum_{i=1}^{K} (K-i) \frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \sum_{i=1}^{K} i \frac{2^{2(i-1)}}{h_{\pi(i)}^2} \\
&= K \sum_{i=1}^{K} \frac{2^{2(i-1)}}{h_{\pi(i)}^2} \\
&= K E_{\text{STAC}}
\end{aligned}$$

In the above, (14) follows from splitting the indices $j \ne i$ into two categories: $j \in \{1, 2, \ldots, i-1\}$ and $j \in \{i+1, i+2, \ldots, K\}$, and (15) follows from switching the roles of $i$ and $j$ and then changing the summation order in the second term of (14). Finally, inequality (16) can be verified by noting that $\frac{2^{2(j-1)}}{h_{\pi(i)}^2} + \frac{2^{2(i-1)}}{h_{\pi(j)}^2}$ is always greater than or equal to $\frac{2^{2(i-1)}}{h_{\pi(i)}^2} + \frac{2^{2(j-1)}}{h_{\pi(j)}^2}$, with the equality holding only when $h_i$ are the same for all $i$'s. This also shows that the more diverse the channel coefficients $h_i$ are, the more energy can be saved by STAC. ■

From Theorem 4, it can be concluded that STAC performs uniformly better than the separate strategy for any set of weight coefficients. This is because even requiring STAC to fully recover the original $K$ source digits leads to better energy efficiency than the separate strategy, for fixed SER and bandwidth efficiency.

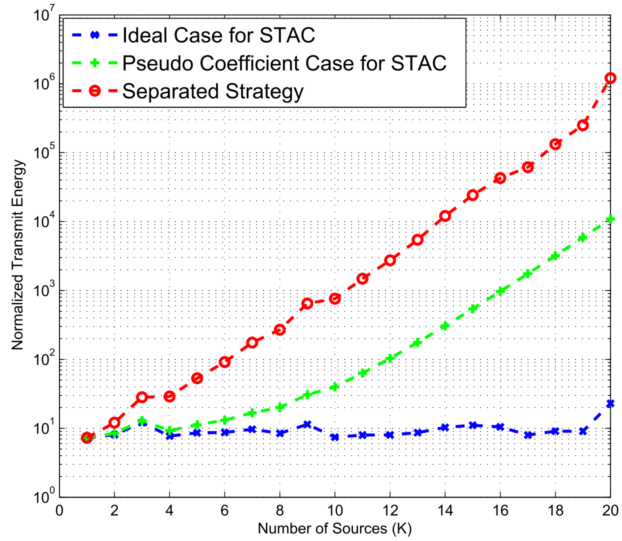



Fig. 5. Energy performance of STAC and the separate strategy.

*3) Discussion:* The above analyzes two extreme cases of the weight coefficients. In general, depending on the specific weight coefficients, one has the freedom of dividing the $K$ nodes into $M$ groups ($1 \le M \le K$), and letting each group transmit using STAC separately, to achieve a tradeoff between the bandwidth efficiency and energy efficiency.

Now, it should be clear that our STAC scheme includes both the separate P2P transmissions and the simultaneous transmission combined with multi-user detection as special cases: Choosing $M = K$ and viewing the P2P transmission as a degraded STAC lead to the former, while the latter is simply equivalent to applying STAC to the pseudo coefficients case.

Fig. 5 plots the normalized transmit energy per node of STAC and the separate P2P transmissions, where the channel coefficients $h$ are assumed to be randomly generated under the standard Rayleigh distribution. It can be seen that as the number of sources $K$ increases, the normalized transmit energy of STAC remains roughly the same for the ideal case, and increases exponentially for the pseudo coefficients case. Compared to the separate P2P transmissions, STAC can save about 95% energy when $K > 15$ even in the pseudo coefficients case.

## V. Network Layer Issues

### A. Network Model

The racks in a DC are normally deployed regularly. This motivates us to consider a 2D regular network as shown in Fig. 6, where each point is a half-duplex transceiver and wireless links exist only between lateral/diagonal neighbours due to the limited transmission range of high frequency EM waves.

Denote this network by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ being the set of vertices (work nodes) and $\mathcal{E}$ the set of edges (wireless links).

### B. Single M2O Session Routing

Consider one M2O session, where a set $\mathcal{S}$ of sources transmit information to the destination $d$ through some edges

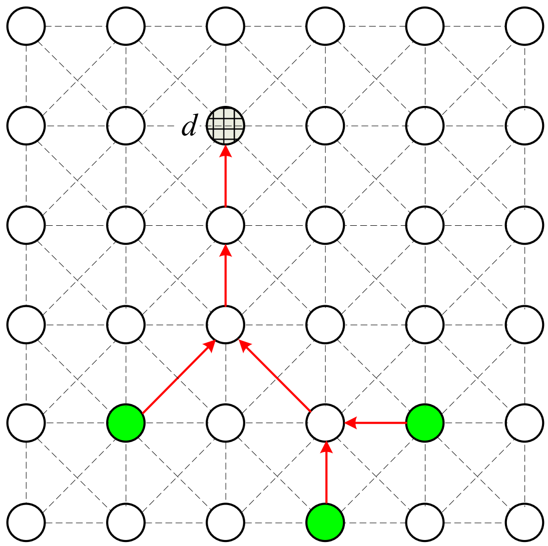



Fig. 6. A 2D regular network. Source nodes (solid circles) transmit information to destination $d$ via relay nodes (hallow circles).

**Algorithm 3** Single M2O Session Routing

$\mathcal{U}_1 = \{d\}$, $\mathcal{U}_2 = \mathcal{S}$, $\mathcal{T} = \emptyset$
**while** $\mathcal{U}_2 \neq \emptyset$ **do**
  Find path $\mathcal{P}$ with minimum cost from $\mathcal{U}_2$ to $\mathcal{U}_1$
  Add all the nodes on $\mathcal{P}$ to $\mathcal{U}_1$
  Remove the source node on $\mathcal{P}$ from $\mathcal{U}_2$
  $\mathcal{T} \leftarrow \mathcal{T} \cup \mathcal{P}$
**end while**
Output $\mathcal{T}$

from $\mathcal{E}$. The (single) M2O routing problem is to find out the tree $\mathcal{T}^*$ from the $|\mathcal{S}|$ sources to the destination $d$ with the minimum cost.

The M2O routing problem with STAC used fundamentally distinguishes from that with only P2P/SAF-based-transmissions, since one has to take into account the possibility of combining multiple flows into a single one with STAC. In fact, M2O routing with STAC used is dual to wireless multicast routing [32]—simply reversing the direction of each hop along a multicast tree yields an M2O tree and vice versa. Since the general multicast routing is an NP-hard minimum Steiner tree problem [32], we have the following theorem.

*Theorem 5:* The M2O routing with STAC used is an NP-hard minimum Steiner tree problem.

Due to the NP-hardness, in Algorithm 3 we propose a greedy routing algorithm based on Dijkstra's [33], which results in a tree $\mathcal{T}$ with cost at most $\lceil \log(|\mathcal{S}| + 1) \rceil$ times the cost of the optimal tree [34]. This algorithm will be used in designing the multiple-M2O session route.

## C. Multiple-Session Scheduling/Routing

Given a set of fixed routing trees, one for each M2O session, the task of a scheduler is to find out the minimum number of time slots (cycle) $T$ during which each session transmits at least once. Greedy algorithms can be used to accomplish this [35]. Note that the number $T$ is lower bounded by the maximum node degree in graph $\mathcal{G}$, i.e., $T \geq \max_{v \in \mathcal{V}} D(v)$, where the degree $D(v)$ of a node $v$ is defined as $D(v) = D_1(v) + 2D_2(v)$, with $D_1(v)$ being the number of trees in which node $v$ is a source or destination and $D_2(v)$ the number of trees in which node $v$ is a relay. The minimum throughput among all the sessions is equal to $1/T$.

Now consider the multiple-M2O session routing. Due to the consecutive jobs and the stringent configuration time, we adopt the online routing where the route of each new session is computed without changing the exiting sessions' routes. Specifically, the network server stores the schedules and routes of the current sessions; when a new session needs to establish, the network server uses Algorithm 3 to find out its route, where the link/tree cost is determined by the existing sessions. After this new route is determined, the schedules will be updated accordingly.

With properly defined cost values, one hopes to balance the total transmit power and network throughput. For this, we propose to use the following tree cost

$$c(e_1, e_2, \ldots, e_N) = g_1 \sum_{i=1}^{N} t_i + g_2 \sum_{i=1}^{N} 2^{D(v(e_i))} \tag{17}$$

where $e_i$ is an edge in the given routing tree with transmission cost $t_i$, $v(e_i)$ denotes the input node of $e_i$, and $g_1, g_2$ are the normalizing factors.

## D. Performance Evaluation

We now use Matlab to run simulations to compare the network performances with the separate strategy and with STAC.

Observing that a job is usually allocated to nearby nodes, we consider a district with $10 \times 10$ nodes. Assume that the transmission cost is equal to 2 for a lateral link, and 3 for a diagonal link. The normalizing factors in (17) are chosen as $g_1 = 1$ and $g_2 = \frac{150}{\sum_{i=1}^{100} 2^{D(i)}}$. Assume that for each session, there is one receive node and 10 source nodes, all of which are randomly selected within the district.

Fig. 7-(a) and (b) plot the the maximum node degree in the network and the average session transmit power against

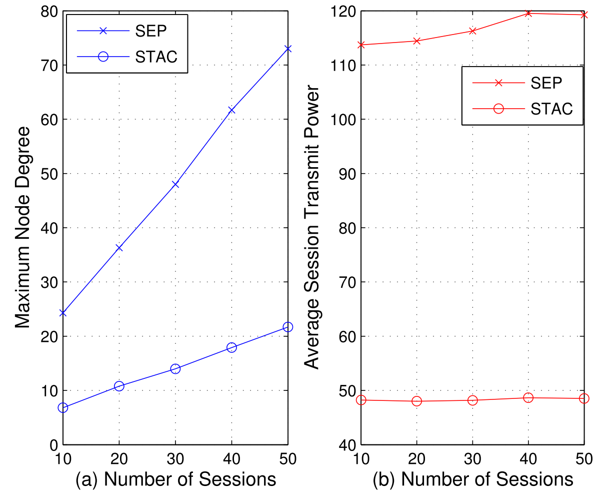



Fig. 7. STAC versus the separate strategy.

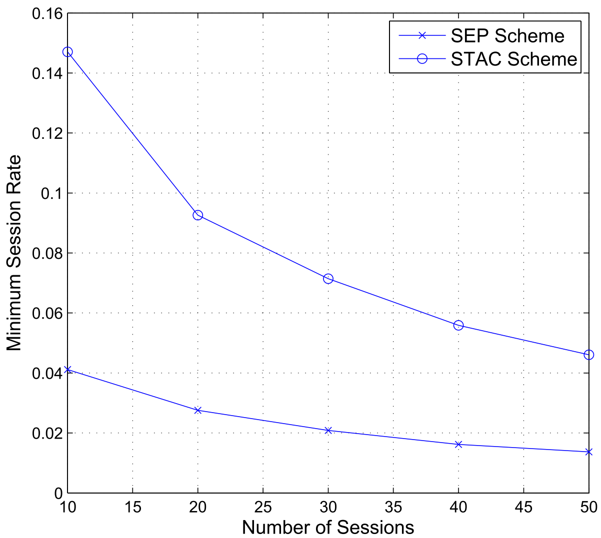



Fig. 8. Minimum session rate comparison.

the number of sessions, respectively. As can be seen, the maximum node degree, whose reciprocal upper bounds the minimum session throughput, increases linearly with the session number for both strategies, but the slope for the separate strategy is much larger than that for STAC; while the average session transmit power is roughly independent of the session number, and that for the separate strategy is about three times of that for STAC. A comparison of the minimum session throughput is given Fig. 8.

## VI. Conclusion

The wireless DCN differs from general wireless networks in that it has large amounts of M2O sessions, which are normally followed by further computations at the destinations, with weighted summation being the typical case. Building on this observation and several distinguishing characteristics of DCs like limited and controlled physical space with static channels, we have proposed a novel physical layer scheme STAC that achieves simultaneous transmissions and computations over the air, and an enhanced SDN architecture to enable it. We showed that STAC can significantly improve both bandwidth and energy efficiencies. STAC also opens a number of exciting future research directions:

### A. Effect on MapReduce Algorithms

Since STAC inherently integrates transmissions and computations, its application can impact the design of computation algorithms in MapReduce. For instance, if the desired computation at the reduce node is in a product form: $s_0 = \prod_{i=1}^{K} s_i$, then to apply STAC, one can make a log operation on the map node's output: $s_i' = \ln s_i$, so that the reduce node can easily transform its received $s_0' = \sum_{i=1}^{K} s_i'$ to $s_0$ by taking $s_0 = e^{s_0'}$.

### B. Advanced Transmission Techniques

This paper only focuses on introducing the basic idea of STAC, while it is possible to combine STAC with other advanced techniques. For example, observing the broadcast traffic [36] in DCs, one can consider exploiting the broadcast nature of wireless medium when designing transmission schemes.

We can also enable inter-session packet scrambling with the idea of network coding [37]. By allowing inter-session network coding, especially physical-layer network coding for the bi-directional traffic [38], the system performance can be further improved.

Additionally, although this paper assumes no error correction coding for simplicity, channel coding can be indeed used in STAC to combat the noise. Linear channel codes as in (1) are directly applicable, and new channel decoding schemes can be devised as in [39].

### C. Theoretical Challenges

STAC brings new challenges to both communication and networking theories. For the one-hop communication model, the capacity and the trade-off between bandwidth and energy efficiencies need further investigations; for the multi-hop multi-session networking model, it would be of interest to find effective polynomial time routing algorithms for some special edge costs.

## References


[1] S. Zhang, X. Wu, and A. Ozgur, "STAC: Simultaneous transmitting and air computing in wireless data center networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–7.

[2] S. Kandula, J. Padhye, and P. Bahl, "Flyways to de-congest data center networks," in *Proc. ACM HotNets Conf.*, 2009, pp. 1–6.

[3] K. Ramachandran, R. Kokku, R. Mahindra, and S. Rangarajan, "60 Ghz data-center networking: Wireless ⇒ worryless?" NEC Lab. Amer. Inc., Princeton, NJ, USA, Tech. Rep., Jul. 2008, pp. 1–11.

[4] X. Zhou *et al.*, "Mirror mirror on the ceiling: Flexible wireless links for data centers," in *Proc. ACM SIGCOMM Conf.*, New York, NY, USA, 2012, pp. 443–454.

[5] N. Hamedazimi *et al.*, "Firefly: A reconfigurable wireless data center fabric using free-space optics," in *Proc. ACM SIGCOMM Conf.*, New York, NY, USA, 2014, pp. 319–330.

[6] D. Halperin, S. Kandula, J. Padhye, P. Bahl, and D. Wetherall, "Augmenting data center networks with multi-gigabit wireless links," in *Proc. ACM SIGCOMM Conf.*, New York, NY, USA, 2011, pp. 38–49.

[7] J.-Y. Shin, E. G. Sirer, H. Weatherspoon, and D. Kirovski, "On the feasibility of completely wirelesss datacenters," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1666–1679, Oct. 2013.

[8] Y. Cui, H. Wang, X. Cheng, and B. Chen, "Wireless data center networking," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 46–53, Dec. 2011.

[9] Y. Cui, H. Wang, and X. Cheng, "Channel allocation in wireless data center networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1395–1403.

[10] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," in *Proc. 19th ACM Symp. Oper. Syst. Principles (SOSP)*, New York, NY, USA, 2003, pp. 29–43.

[11] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[12] L. Mucchi, L. Ronga, and G. Chisci, "Noise-loop multiple access," *IEEE Trans. Veh. Technol.*, to be published.

[13] R. Liu, "Autocorrelation division multiple access systems (ADMA)," in *Proc. Int. Conf. Commun., Circuits Syst.*, Jul. 2009, p. 1.

[14] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

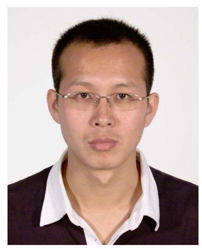
[15] C.-T. Chu *et al.*, "MapReduce for machine learning on multicore," in *Proc. Neural Inf. Process. Syst. Conf. (NIPS)*, Apr. 2006, p. 281.

[16] C. Shepard *et al.*, "Argos: Practical many-antenna base stations," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2012, pp. 53–64.

[17] N. McKeown *et al.*, "OpenFlow: Enabling innovation in campus networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, Apr. 2008.
[18] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. IEEE 26th Symp. Mass Storage Syst. Technol. (MSST)*, May 2010, pp. 1–10.
[19] Y. Wu, "Existence and construction of capacity-achieving network codes for distributed storage," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 2, pp. 277–288, Feb. 2010.
[20] D. S. Papailiopoulos, J. Luo, A. G. Dimakis, C. Huang, and J. Li, "Simple regenerating codes: Network coding for cloud storage," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2801–2805.
[21] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, and C. Kozyrakis, "Evaluating mapreduce for multi-core and multiprocessor systems," in *Proc. 13th IEEE Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2007, pp. 13–24.
[22] S. Seo, E. J. Yoon, J. Kim, S. Jin, J.-S. Kim, and S. Maeng, "HAMA: An efficient matrix computation with the mapreduce framework," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Technol. Sci. (CloudCom)*, Nov./Dec. 2010, pp. 721–726.
[23] C. Liu, H. Yang, J. Fan, L.-W. He, and Y.-M. Wang, "Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce," in *Proc. 19th Int. Conf. World Wide Web*, New York, NY, USA, 2010, pp. 681–690.
[24] T. Elsayed, J. Lin, and D. W. Oard, "Pairwise document similarity in large collections with MapReduce," in *Proc. 46th Annu. Meeting Assoc. Comput. Linguistics Human Lang. Technol., Short Papers*, Stroudsburg, PA, USA, 2008, pp. 265–268.
[25] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, 2006, pp. 358–365.
[26] E. E. Marinelli, "Hyrax: Cloud computing on mobile devices using mapreduce," M.S. thesis, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2009.
[27] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, Aug. 2008.
[28] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.
[29] D. R. Brown, III, and H. V. Poor, "Time-slotted round-trip carrier synchronization for distributed beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5630–5643, Nov. 2008.
[30] W. Braun and M. Menth, "Software-defined networking using openflow: Protocols, applications and architectural design choices," *Future Internet*, vol. 6, no. 2, pp. 302–336, 2014.
[31] D. Cox, E. Jovanov, and A. Milenkovic, "Time synchronization for ZigBee networks," in *Proc. 37th Southeastern Symp. Syst. Theory (SSST)*, Mar. 2005, pp. 135–138.
[32] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks," in *Proc. 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2. Mar. 2000, pp. 585–594.
[33] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.
[34] A. Goel and K. Munagala, "Extending greedy multicast routing to delay sensitive applications," *Algorithmica*, vol. 33, no. 3, pp. 335–352, Jul. 2002.
[35] F. Gavril, "Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph," *SIAM J. Comput.*, vol. 1, no. 2, pp. 180–187, 1972.
[36] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of data center traffic: Measurements & analysis," in *Proc. 9th ACM SIGCOMM Conf. Internet Meas. Conf.*, New York, NY, USA, 2009, pp. 202–208.
[37] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
[38] M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, and A. Vahdat, "Hedera: Dynamic flow scheduling for data center networks," in *Proc. 7th USENIX Conf. Netw. Syst. Design Implement.*, Berkeley, CA, USA, 2010, p. 19.
[39] S. Zhang and S.-C. Liew, "Channel coding and decoding in a relay system operated with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 788–796, Jun. 2009.

**Xiugang Wu** (S'08–M'15) received the B.Eng. degree (Hons.) in electronics and information engineering from Tongji University, Shanghai, China, in 2007, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2009 and 2014, respectively.

He is currently a Post-Doctoral Research Fellow with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. His research interests include information theory and wireless networks.



**Shengli Zhang** (M'10) received the B.Eng. degree in electronic engineering and the M.Eng. degree in communication and information engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2002 and 2005, respectively, and the Ph.D. degree from the Department of Information Engineering, The Chinese University University of Hong Kong (CUHK), in 2008. From 2002 to 2005, he was with the Personal Communication Network and Spread Spectrum Laboratory, USTC, as a Research Engineer, where he was involved in several National 863 research projects including the Beyond-3 Generation of Mobile System in China (FUTURE Plan). From 2002 to 2005, he was also a Research Engineer with the UTStarcom Wireless Soft Research Center, Hefei, where he was involved in the research and implementation of the WCDMA communication systems. From 2008, he was a Research Associate with CUHK. He is currently an Associate Professor with the Communication Engineering Department, Shenzhen University, China. His current research interests include wireless networks, wireless communication, physical layer network coding, and cooperative wireless networks.

**Ayfer Özgür** (M'06) received the B.Sc. degree in electrical engineering and physics and the M.Sc. degree in communications from Middle East Technical University, Turkey, in 2001 and 2004, respectively, and the Ph.D. degree from the Information Processing Group, EPFL, Switzerland, in 2009. From 2001 to 2004, she was a Hardware Engineer with the Defense Industries Development Institute, Turkey. From 2010 to 2011, she was a Post-Doctoral Scholar with the Algorithmic Research in Network Information Group, EPFL. She is currently an Assistant Professor with the Electrical Engineering Department, Stanford University. Her research interests include network communications, wireless systems, and information and coding theory. She received the EPFL Best Ph.D. Thesis Award in 2010 and the NSF CAREER Award in 2013.