Mirror Mirror on the Ceiling: Flexible Wireless Links for Data Centers

Xia Zhou, Zengbin Zhang, Yibo Zhu, Yubo Li[§], Saipriya Kumar,
Amin Vahdat[†], Ben Y. Zhao and Haitao Zheng
Department of Computer Science, U. C. Santa Barbara, USA
[§]Xi'an Jiaotong University, P. R. China [†]U. C. San Diego, USA
{*xiazhou,zengbin,yibo,saipriya,ravenben,htzheng*}*@cs.ucsb.edu*, *heatonlyb@gmail.com*, *vahdat@cs.ucsd.edu*

ABSTRACT

Modern data centers are massive, and support a range of distributed applications across potentially hundreds of server racks. As their utilization and bandwidth needs continue to grow, traditional methods of augmenting bandwidth have proven complex and costly in time and resources. Recent measurements show that data center traffic is often limited by congestion loss caused by short traffic bursts. Thus an attractive alternative to adding physical bandwidth is to augment wired links with wireless links in the 60 GHz band.

We address two limitations with current 60 GHz wireless proposals. First, 60 GHz wireless links are limited by line-of-sight, and can be blocked by even small obstacles. Second, even beamforming links leak power, and potential interference will severely limit concurrent transmissions in dense data centers. We propose and evaluate a new wireless primitive for data centers, *3D beamforming*, where 60 GHz signals bounce off data center ceilings, thus establishing indirect line-of-sight between any two racks in a data center. We build a small 3D beamforming testbed to demonstrate its ability to address both link blockage and link interference, thus improving link range and number of concurrent transmissions in the data center. In addition, we propose a simple link scheduler and use traffic simulations to show that these 3D links significantly expand wireless capacity compared to their 2D counterparts.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

Keywords

Data centers, 60 GHz wireless, wireless beamforming

1. INTRODUCTION

Modern distributed applications running in clusters and data centers can run at massive scale, with potentially tens of thousands of servers spread across hundreds of racks. The bandwidth requirements of these applications can range from the relatively modest (*e.g.*, hundreds of Mb/s per server [44]) to substantial (*e.g.*, tens of Gb/s per server for high-end scientific computing [24]). Delivering such bandwidth comes at substantial cost for the requisite switching infrastructure. As a result, a number of recent efforts have investigated techniques for deploying more efficient data center network topologies [11, 14, 20, 21, 22].

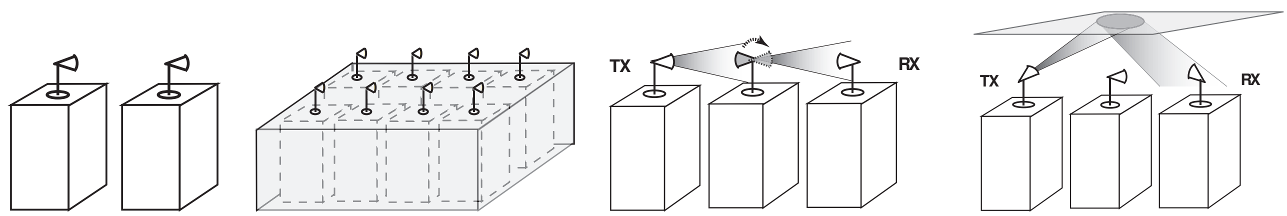
While these alternate topologies offer a range of benefits over the current state of the art, we argue that there are a number of inherent challenges with the deployment of *any* wired network technology. First, any large-scale network consists of multiple "stages," meaning a multiplicative factor in the number of fibers/wires required for every server in the cluster. The process of planning, routing, deploying, testing, and then repairing tens of thousands of fibers at the scale of a building incurs substantial cost (both in capital and operational expenditures). Based on our experience, this manual overhead often delays the operational time of large clusters by weeks or months. Lost machine depreciation and lost opportunity costs for leveraging additional capacity can produce millions of dollars in hidden cluster costs.

Second, wired deployments typically cannot anticipate bandwidth requirements to every rack, and thus must distribute a fixed amount of fiber to every rack spot. This means that we must overdeploy fiber for the "worst case" of communication requirements rather than the average case, increasing the cost and maintenance of networking infrastructure.

Third, deployed networks are extremely costly and complex to modify, partially due to the characteristics of multi-stage network topologies. For example, doubling the number of top of rack servers in a Clos topology network [14, 21] either requires rewiring half the existing fiber (impossible since fibers are often bundled together) or pre-deploying twice the higher-stage network switches than otherwise required (an expensive proposition). Similarly, moving a rack later based on computation needs requires running additional fiber to match. Our experience has shown that adding fiber to a pre-existing cluster is a complex process that incurs substantial delays, and sometimes simply intractable.

Of course, the key benefit of a wired network, whether electrical or optical, is the tremendous amount of bandwidth it can deliver. While the highest end of data center computing is likely to continue to require wired network deployments, in this paper we focus on the subset of applications, and perhaps the majority of applications as borne out by recent measurement studies [15, 16, 25], with more modest or

SIGCOMM'12, August 13–17, 2012, Helsinki, Finland.



(a) Rack-based DC (b) Container-based DC (c) 2D Beamforming (d) 3D Beamforming

Figure 1: Radio transceivers are placed on top of each rack (a) or container (b). Using 2D beamforming (c), transceivers communicate with neighboring ones directly, but forward traffic in multiple hops to non-neighboring racks. Using 3D beamforming (d), the ceiling reflects the signals from each sender to its receiver, avoiding multi-hop relays.

more localized/bursty bandwidth requirements. That is, we focus on the subset that do not require (near) non-blocking all-to-all communication at data center scale.

In particular, we focus on high-throughput, beamforming wireless links in the 60 GHz band. The unlicensed 60 GHz band provides multi-Gbps data rates and can be implemented with relatively low-cost hardware. Because 60 GHz signals attenuate quickly with distance, multiple wireless links can be deployed in a single data center. In our efforts to expand the effective bandwidth of 60 GHz links, we hope to create a new primitive that can be used to either augment existing networks with *on-demand* network links, or potentially replace wired links in data centers with modest bandwidth requirements. We build on pioneering efforts of earlier work that proposed 60 GHz links to alleviate hot spots in the data center [23, 26].

However, earlier efforts face a number of limitations. First, even beamforming directional links will experience *signal leakage*, and produce a cone of interference to receivers near or behind the intended target receiver. This limits the number of links that can be active concurrently in densely occupied data centers, and reduces the aggregate throughput offered by these wireless links.

Second, these links *require direct line-of-sight* (LOS) between sender and receiver, and can be blocked by even small objects in the path. This limits the effective range of 60 GHz links to neighboring top-of-rack radios. Since hotspots occur regularly at both edge and core links [15], augmenting core links would require multiple hops through a line-of-sight 60 GHz network. Half-duplex, directional antennas mean that these multi-hop links will suffer at least a 50% throughput drop, higher-levels of potential congestion, and additional delays required to frequently adjust antenna orientation.

To address these issues, we investigate the feasibility of 60 GHz 3D beamforming as a flexible wireless primitive in data centers. In 3D beamforming, a top-of-rack directional antenna forms a wireless link by reflecting a focused beam off the ceiling towards the receiver. This reduces its interference footprint, avoids blocking obstacles, and provides an indirect line-of-sight path for reliable communication. Such a system requires only beamforming radios readily available today, and near perfect reflection can be provided by simple flat metal plates mounted on the ceiling of a data center.

3D beamforming has several distinctive advantages over prior "2D" approaches. First, bouncing the beam off the ceiling allows links to extend the reach of radio signals by avoiding blocking obstacles. Second, the 3D direction of the beam significantly reduces its interference range, allowing more nearby flows to transmit concurrently. Third, the reduced interference extends the effective range of each link, allowing our system to connect any two racks using a single hop, and mitigating the need for multihop links.

In this paper, we propose a 3D beamforming system for 60 GHz wireless transmissions in data centers. The 3D beamforming idea was first introduced by Zhang et al. in [46]. In this paper, we greatly extend the prior work, and use measurements of a local 60 GHz testbed to quantify and compare the performance of 3D and 2D beamforming links. We find that 3D wireless beamforming works well in practice, and experiences zero loss in signal or throughput from reflection. We also describe a link scheduler for 3D beamforming systems that maximizes concurrent links while also taking into account accumulative interference and antenna alignment delays. Finally, we use a detailed simulation of data center traffic hotspots to quantify the performance of 3D beamforming systems. Our results show that while 2D links can only support a small portion of hotspot traffic links, 3D beamforming can connect all rack pairs in a single hop, and can significantly reduce overall data completion time for wired networks across a range of bisection bandwidths.

While wired networks will likely remain the vehicle of choice for the high-end of distributed computing, we believe that efforts such as 3D beamforming can expand the applicability and benefits of wireless networking to a broader range of data center deployments.

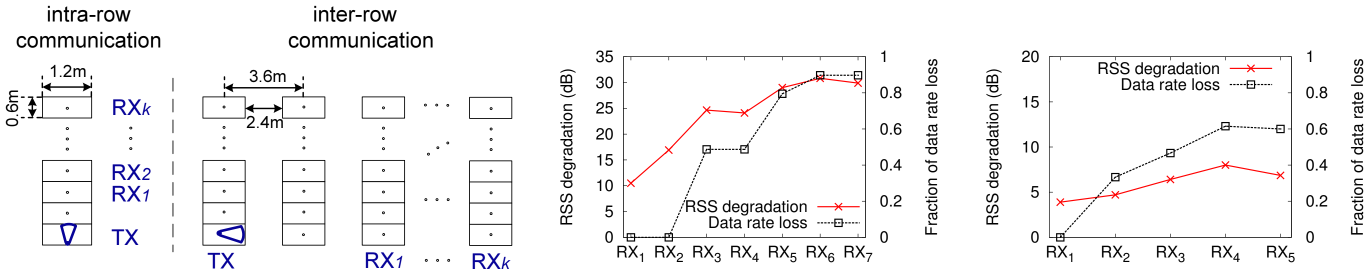
2. 60 GHZ: LIMITATIONS AND SOLUTIONS

While modifying the topology of wired data centers is costly, complex, and sometimes intractable, administrators can introduce flexible point-to-point network links with the addition of wireless radios. Prior work has proposed the use of 60 GHz links to augment data center capacity [23, 26, 35, 38]. Figures 1(a)-(b) show a common deployment scenario, where wireless radios are placed on top of each rack or container to connect pairs of top-of-rack (ToR) switches.

In practice, however, data center managers remain skeptical on deploying wireless links despite their potential benefits [1]. In this section, we summarize prior work in this space, and use detailed experiments on a 60 GHz testbed to identify and quantify key limitations of current proposals.

2.1 60 GHz Links in Data Centers

Existing designs [23, 26, 27, 38] adopt 60 GHz wireless technologies for several reasons. *First*, the 7GHz spectrum



(a) Experiment Setup (b) Intra-row Blockage (c) Inter-row Blockage

Figure 2: Impact of link blockage using 2D beamforming. (a) Our experiment setup to examine antenna blockage for intra- and inter-row links. TX represents the transmitter, and RX_k is the target receiver which is blocked by k antennas. (b-c) The RSS degradation and data rate loss for intra- and inter-row scenarios.

available in this band can deliver the multi-Gbps data rates required by data centers. *Second*, 60 GHz links operate at a high carrier frequency, which limits the interference they generate [45], and is highly beneficial to data centers with dense rack deployments. *Third,* 60 GHz links can use beamforming to enhance link rate and further suppress interference [23, 27]. Beamforming is a physical layer technique to concentrate transmission energy in a specific direction. Today, 60 GHz beamforming radios are readily available and affordable, either as directional (horn) antennas [23] or antenna arrays [8]. They use either mechanical or electronic mechanisms to achieve fine-grain directional control [23, 40].

A 60GHz Testbed. As an experimental platform for 60 GHz wireless projects, we have built a small 60 GHz wireless testbed. To build a bidirectional 60 GHz link, we set two Gigalink 6451 radios from HXI Millimeter Wave Products [5] to operate using frequency division on two frequency bands near 60 GHz, centered at 58.2 GHz and 61.9 GHz. A simple On-Off-Keying modulation method delivers up to 1.25Gbps of bandwidth,but actual throughput is limited to 1 Gbps by the radio's fiber-optic Gigabit Ethernet interface.

These radios were originally designed for long distance wireless links, and each is equipped with a 60 GHz transceiver and a 41dBi directional antenna transmitting at 10dBm power. To modify them for short distance data center links in our experiments, HXI engineers modified the radio amplifier to reduce output power to 0dBm. We also replaced 41dBi antennas with a 25dBi standard gain horn antenna from Flann Microwave [3], with 3dB beamwidth of 10°. For our measurements, we read RSS values from an SNMP-based software monitor running on the radio. Since currently available 60 GHz radios all transmit at a single fixed data rate, we derive link throughput by converting RSS values to data rates according to the 802.11ad standard [6].[1]

2.2 Current Limitations

Despite the many tangible benefits of adding 60 GHz links to a data center, there are two notable limitations with current designs. Here we discuss each in detail and use data from our testbed to quantify its impact.

Link Blockage. Link blockage is a limiting factor for 60 GHz links. The 5mm wavelength of these links means that any object larger than 2.5 mm can effectively block signals or reflect them, producing multipath fading and degrading transmission rates [37]. In today's data centers, this is problematic because racks are organized in a grid, and transceivers and antennas on one rack can easily block transmissions on another rack. This has led to current designs limiting themselves to connecting neighboring racks [23, 27] (see Figure 1(c)).

We measure the severity of the problem by placing multiple antennas between two 60 GHz endpoints, and observing performance degradation on the link. We examine cases for both intra- and inter-row rack communications (see Figure 2(a)). In our intra-row case, antennas are separated by a fixed distance of 0.6m (roughly one rack width). Thus, a link blocked by k antennas, *i.e.* $TX \rightarrow RX_k$, has a distance of $(k+1)\cdot 0.6$m. We vary the number of blocking antennas in the experiment, and show the difference in performances with and without blocking antennas. Figure 2(b) shows the RSS degradation and data rate loss. Clearly, even when there is only one blocking antenna (link distance = 1.2m), RSS degradation can be as high as 10dB. Since the signal strength is still high, RSS degradation does not reduce data rates. As the number of antennas increases, the RSS degradation becomes more severe and data rates drop quickly: 3 antennas (link distance = 2.4m) can cause a 25dB RSS degradation and 50% data rate loss, while 6 blocking antennas (link distance = 4.2m) can cause a 30dB RSS degradation and nearly 90% data rate loss.

For inter-row communications, antennas are separated by a fixed distance of 3.6m (one rack length of 1.2m plus one row separation of 2.4m) [23]. Results in Figure 2(c) show similar trends as the intra-row case, except that the impact of blockage is slightly lower. This is because the RF beam emitted by the horn antenna propagates in a cone-shape. The closer the first blocking antenna is to the transmitter, the more the signal it blocks from the receiver. Overall, our measurement results clearly demonstrate that link blockages cause severe problems for 60 GHz transmissions.

To reduce link blockage, one option is to intelligently place radios on each racks, which might be effective for some rack pair connections. To connect racks dynamically, however, the radios must tune to different directions and still

[1]The 802.11ad standard defines a set of data rates (385Mbps-6.76Gbps). Following its receiver sensitivity table, which guarantees 1% packet error rate and assumes 10dB noise figure and 5dB additional implementation loss, we derive the required SNR for each data rate. We then compute the measured SNR from each measured RSS and noise, and use the above mapping to derive the data rate supported by each measured RSS value.

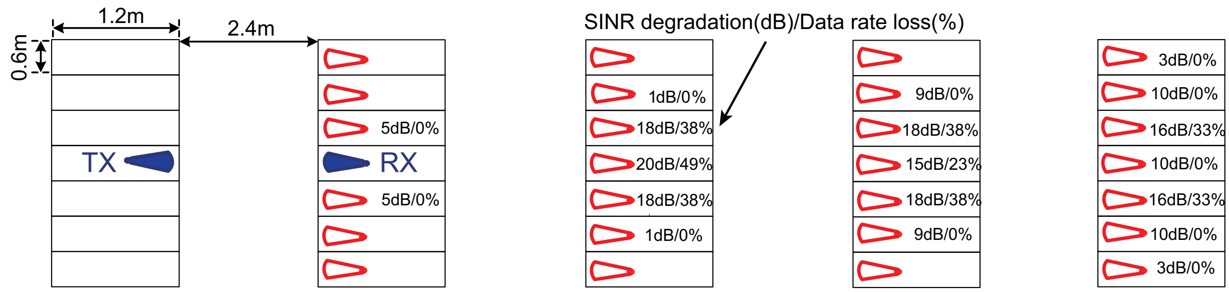


Figure 3: Our experiment to examine the impact of link interference using 2D beamforming. We place a transmitter TX in the middle of a rack row as the interfering transmitter, a receiver RX in the next rack row as its receiver, and four additional rack rows as victims. Each victim radio, in red, points its antenna to the rack of its neighboring row to the left, representing the receiver of an inter-row link of link distance 3.6m. In each rack we list the SINR degradation and data rate loss due to the interference from TX. We omit the results at racks whose SINR degradation is less than 1dB.

block transmissions. Another option is to place racks in hexagons [41]. It leads to inefficient space use, and yet still does not solve the fundamental link blockage problem. Finally, one can build multi-hop connections between non-neighboring racks. This, however, increases end-to-end delay, dramatically reduces throughput, and produces potential bottlenecks at racks that are congested by forwarding traffic from multiple links.

Radio Interference. Despite the use of beamforming to bound the transmission energy in a "narrow" direction, radio interference remains an issue for these systems. Radio design artifacts will still produce signal leaks outside of the intended direction [29, 40]. When placed in a dense rack formation, leakage produces harmful interference between nearby links and limits the density of concurrent links.

Using our testbed, we measure the impact of interference produced by a single transmission, in the presence of antenna blockage. As shown in Figure 3, we place a 60GHz transmitter (TX) in the middle of the data center, and measure the RSS at 27 racks (in red) located in four neighboring rows. At each of these 27 racks, the radio points its antenna to the rack of its immediately left row, representing the receiver of an inter-row link with link distance of 3.6m and SNR of 31dB. This experiment allows us to measure the interference experienced by each of the 27 inter-row links when TX is transmitting, from which we compute the SINR degradation and data rate loss due to interference. Results in Figure 3 show that despite the fact that TX's interference signal is blocked by various antennas, 15 inter-row links behind the destination still observe 5-20dB degradation in their SINR. 8 of these links suffer 20-49% data loss.

The spread of radio interference significantly limits the number of concurrent wireless links in a data center. One option is to separate the links in the frequency domain. But this reduces the per-link capacity, since the total available bandwidth is fixed across the frequency range. Alternatively, data center managers can increase the spacing between racks to reduce interference. But this leads to inefficient space and power usage, and weakens long-distance links.

2.3 Solution: 3D Beamforming

To address these limitations, we propose *3D beamforming*, a new beamforming approach that leverages ceiling reflections to connect racks wirelessly. An example is shown in Figure 1(d), where a transmitter bounces its signal off of the ceiling to the receiver. This creates an indirect line-of-sight path between the sender and receiver, bypassing obstacles[2] and reducing interference footprint.

To align its antenna for a transmission, the sender only needs to know the physical location of the receiver rack, and point to a position on the ceiling directly between the two racks. This is because all racks (and their 60 GHz radio antennas) are of the same height.

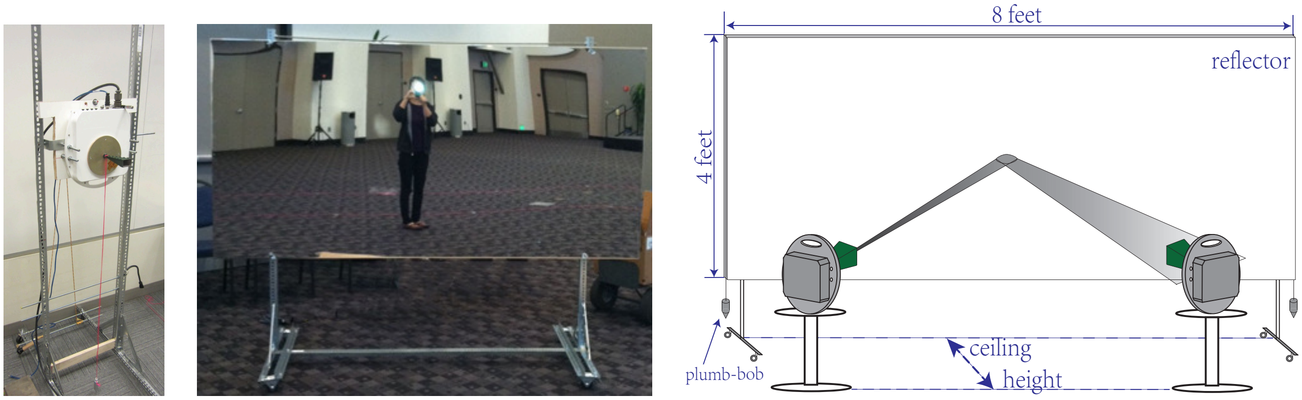
3D beamforming requires three hardware components:

- *Beamforming Radios*: We reuse beamforming radios [8, 23] and adjust beam directions in both azimuth and elevation by placing the horn antennas on rotators. Existing rotators can achieve an accuracy of 0.006°-0.09° [4, 7].
- *Ceiling Reflectors*: Reflectors on the ceiling act as *specular mirrors* to reflect signals. Our experiments confirm prior work [12, 36] showing that flat metal plates offer perfect specular reflection without degrading energy or changing path loss characteristics.
- *Electromagnetic Absorbers*: We place electromagnetic absorbers [12] near each antenna to prevent any local reflection and scattering. These inexpensive absorbers require no maintenance.

3D beamforming largely addresses both of the main limitations with existing 2D 60 GHz proposals. First, by bouncing beams off a reflective ceiling, it dramatically reduces the interference region for wireless links and allows deployment in densely packed data centers. Second, the reflective path avoids obstacles and creates effective line-of-sight paths between most or all rack pairs in a common 250-rack data center. Addressing these issues means we can connect most or all rack pairs using single-hop 60 GHz links, thus maximizing bandwidth and eliminating forwarding delays. It also means a large number of links can be active in a small area without causing mutual interference and limiting performance.

In this paper, we present first steps in building flexible wireless links using 3D beamforming. We identify and address practical issues in the physical and link layers, and describe experience and experimental results from a local 60

[2]Here we assume that there are no obstacles between top of racks (or containers) and the ceiling. For instance, this might require mounting the radios above cable trays, though we leave a detailed study of physical deployment to future work.



(a) Radio (b) Metal Reflector (c) Testbed Overview

Figure 4: Our 3D beamforming testbed. (a) The 60GHz radio with horn antenna, mounted on top of a mobile platform with adjustable height. (b) A 4ft×8ft, mirror-quality stainless steel reflector mounted vertically on a mobile platform. (c) An illustration of the experiment configuration.

GHz 3D beamforming testbed. We limit our discussion of 3D beamforming as a general link-layer primitive, and leave for future work other issues such as routing, traffic management, and wired/wireless co-scheduling. We believe this work addresses a few of the key concerns associated with large-scale wireless data center deployments, principally enabling substantially more bandwidth to be delivered more flexibly in the data center. However, many open questions remain before we expect to see large-scale deployments.

3. MICROBENCHMARK RESULTS

Using detailed hardware experiments, we now examine the key properties of 3D beamforming, and compare them to 2D systems. We focus specifically on physical performance characteristics of our approach, and its sensitivity to factors such as radio density, rotator accuracy, and reflection material.

3D Beamforming Testbed. As shown by Figure 4, our local testbed consists of two 60GHz beamforming radios from HXI (described in Section 2), a 4ft×8ft metal reflector, and RF absorbers from ETS-Lindgren [2]. We test two types of reflectors: commercial-grade mirror-quality stainless steel plates and off-the-shelf cheap galvanized steel sheets from our local home improvement store. To assist with rapid experimentation, we mount the reflector vertically on a mobile platform that stands in parallel to a line connecting the center of the two radio transceivers. We vertically align platform using multiple hanging plumb-bobs. The corresponding ceiling height h is the perpendicular distance between the reflector and the line. To prevent reflected signals from producing more reflections at the receiver side, we place RF absorbers under the antenna. The absorber is a surface tiled with small pyramids 7.5cm thick. It does not block 3D transmit/reflection paths, but eliminates additional reflections. Finally, instead of using rotators, we manually calibrate the orientations of the horn antennas, using high precision laser pointers for guidance. We also manually introduce calibration errors to emulate the use of rotators of different precision (Section 3.3).

We performed detailed experiments in two indoor environments: a 10m×10m conference room and a 20m×26m pavilion room.

3.1 Validating Physical Properties

Our first question is a basic one: "does 3D beamforming work, and what is the impact of reflection on signal strength and interference characteristics?"

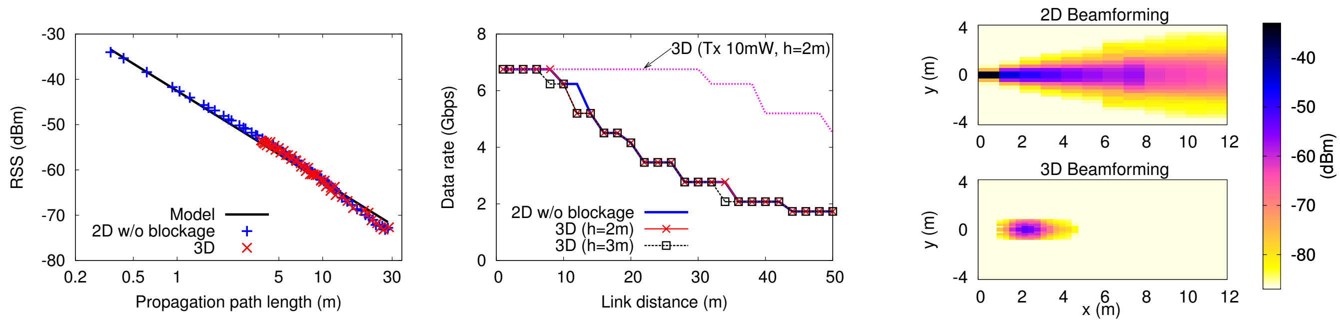
Property 1: Extended Link Connectivity. Our first experiment looks at link connectivity. Intuitively, using ceiling reflection, 3D beamforming will bypass obstacles in the horizontal plane, eliminating the antenna blockage problem of its 2D counterpart. More importantly, since ceiling reflectors should produce no loss [12, 36], it should produce an indirect LOS path following the free-space propagation model [23, 32]:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 (L^2 + 4h^2)} \quad (1)$$

where P_t and P_r are the transmit and receive power, G_t and G_r are the transmit and receive beamforming antenna gains, λ is the radio wavelength, L is the distance between the sender and receiver, and h is the distance from the antenna to the ceiling. To verify our hypothesis, we measure RSS at different link distances for both 2D (with no obstacles) and 3D beamforming. We also vary the ceiling height h between 2m and 3m.

The results confirm our hypothesis. Figure 5(a) plots the measured RSS as the function of the propagation path length, *i.e.* L for 2D and $\sqrt{L^2 + 4h^2}$ for 3D beamforming. As a point of reference, we also plot the free-space model in (1). We make three key observations. *First*, our measurement results match the model, confirming that both beamforming methods follow the free-space propagation model, and that the reflector introduces no energy loss. *Second*, a mirror-quality stainless steel plate and a cheap galvanized steel sheet both offer perfect reflection. *Third*, we found no visible difference between the results collected in the two rooms.

We also verify 3D beamforming's ability of bypassing obstacles in the 2D plane by placing absorbers along the line connecting the two radios. The height of the absorber is similar to that of the horn antenna. We observe no difference in RSS even for the longest link distance achievable in the two rooms (30m).



(a) RSS (b) Link Data Rate (c) Interference Map

Figure 5: Measured propagation characteristics of 3D beamforming, transmit power at 0dBm. (a) Ceiling reflection introduces no energy loss; (b) Despite having a longer propagation path, 3D beamforming can still maintain similar data rate as 2D beamforming without blockage; (c) 3D beamforming can significantly reduce the interference footprint. The sender and receiver are at (0, 0), (2.4, 0) respectively.

We also examine the link throughput of the two beamforming methods. Because the HXI radios transmit at a single data rate (1Gbps), we examine a wider range of data rates using the measured RSS and the 802.11ad's receiver sensitivity table. Figure 5(b) shows the resulting link throughput as a function of the link distance L. Because the room where we performed experiments was of a limited length (30m), we derived the data rates for longer links using the RSS values generated by the propagation model. We see that even at a very low transmit power (0dBm), 3D beamforming can reach 6+Gbps when two endpoints are separated by 10m or less. At a link distance of 50m, it still offers nearly 2Gbps of throughput. If we set the transmit power to the standard level of 10dBm, link throughput, shown as the dotted line in the same figure, increases to 6.76Gbps at 30m and 4.5Gbps at 50m. Furthermore, compared to 2D beamforming, 3D achieves nearly the same data rate despite having a longer propagation path (see Eq. (1)).

Property 2: Reduced Radio Interference. Our second experiment examines the interference footprint of both 2D and 3D beamforming. For both methods, we first set up a target transmission link X, then keep the transmitter intact and move the receiver around to measure link X's power emission map. We divide the measurement space into 0.3m×0.15m grids. In each grid, we rotate the receiver antenna to locate the direction with the maximum signal strength, subtract this strength by the receiver antenna gain, and use the result as the maximum interference that link X produces to this location.

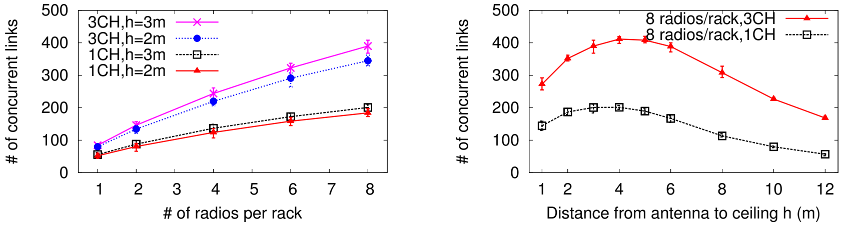
Figure 5(c) shows the measured interference footprint for both 2D (w/o blockage) and 3D beamforming, when the ceiling height h=2m. The sender and the receiver of the target link X are placed at position (0m, 0m) and (2.4m, 0m) on the map, respectively. For 2D beamforming, the directional wave still propagates freely in its beam direction, affecting other receivers along the path. The signal leakage also contributes to the level of interference. In contrast, 3D beamforming bounds the interference region to a much smaller area, and limits the impact of signal leakage. We also verified that the measured interference footprint aligns with the propagation model and the antenna pattern of the 10^o horn antenna [28, 29]. We omit those results due to space limitations.

3.2 Multiple Radios per Rack

In practice, a single rack can host many servers (e.g., 20-80), and is likely to request multiple simultaneous data connections with other racks. With a single 60 GHz radio, these transmissions will be performed in order, with the antenna re-orienting between transmissions. A much more desirable scenario is to put multiple radios on each rack to support parallel transmissions and reduce head-of-line blocking. For today's standard racks with size (4ft×2ft) and 60GHz radio size (1ft×1ft), we can place up to 8 radios per rack.

We quantify these benefits by using simulations to compute the number of concurrent wireless links supported for two data center configurations. The first configuration has size 15m×42m and contains 250 racks, similar to the layout used in [23]. Racks are grouped into 5×5 clusters, and each cluster is a row of 10 racks with no inter-spacing. Aisles separating the clusters are 3m (between columns) and 2.4m (between rows). The second deployment uses shipping containers [10]. It consists of 2×2 container clusters. Each cluster has 8 containers in a row with inter-spacing of 0.61m. Overall, the data center has size 15m×50m, and contains 256 racks.

We configure wireless links as follows. We assign m radio transceivers per rack and allow each transceiver to associate with one link. Given the size of our deployments, we use 60GHz radios with 10dBm transmit power and standard 10° horn antenna, so that every rack pair connects in 1-hop at 5+Gbps in both directions. We build bi-directional links by randomly selecting rack pairs, forming arbitrary rack to rack communication. We determine the number of concurrent links as follows. We admit links one by one in a random order, compute their cumulative interference to each other, and only admit a link if all links after admission achieve their stand-alone data rates as if there were no interference. In other words, these concurrent links do not interfere with each other. We consider two cases: when all the links operate on a single 2.16GHz channel, and when three 2.16GHz channels (for the US 60GHz band) are available. Our simulator uses the free-space propagation model (defined by Eq.(1)), which we verified via experiments in Section 3.1. We compute interference as the total energy accumulated from all concurrent transmissions, accounting for the impact of both antenna orientation and radiation pattern [28, 29].



(a) Impact of # of Radios per Rack

(b) Impact of Ceiling Height h

Figure 6: The number of concurrent links using 3D beamforming, varying the number of radios per rack and ceiling height h, using one (1CH) or three channels (3CH).

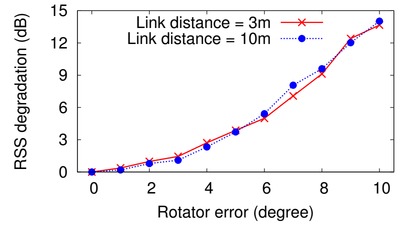


Figure 7: RSS degradation caused by antenna rotator error (measured by antenna mis-alignment angle), using 3D beamforming.

Figure 6(a) plots the number of concurrent links supported as a function of the number of radios per rack. The two topologies lead to similar results and thus we only show the result for the first. We make two key observations. *First*, with a single radio, an average of 55 randomly formed links can operate simultaneously on a single channel. When using three channels, about 88 links (70% of the total links) can operate simultaneously[3]. This result shows that we can simultaneously connect the majority of rack pairs wirelessly using a wire-like connection with 5+Gbps of bandwidth.

Second, the number of concurrent links grows linearly with the number of radios per rack. With eight radios and three channels, an average of 390 randomly formed bi-directional links can operate simultaneously, a 440% improvement over the single radio scenario. This also means that on average, each rack can communicate with four other racks simultaneously, while each bi-directional link achieves at least 5Gbps. This type of flexible and extended connectivity is particularly useful for popular data center jobs such as the "shuffle phase" of MapReduce and local multicast.

Impact of Ceiling Height h. From Figure 6(a), we also observe that increasing the ceiling height h from $2m$ to $3m$ leads to more concurrent links. This is because a larger h makes each beam arrive at its receiver at a larger elevation angle, effectively reducing the interference region. Increasing h beyond $4m$, however, leads to performance loss (Figure 6(b)). This is because increasing h also lengthens the signal propagation path and hence degrades the received signal strength. This loss starts to dominate when h exceeds $4m$, creating a sweet spot of h between 3-4m.

3.3 Sensitivity to Hardware

Finally, we examine the sensitivity of performance to different types of hardware and materials.

Sensitivity to Rotator Accuracy. The first question is whether the performance of 3D beamforming will degrade significantly if antenna directions are not calibrated accurately, *e.g.* due to rotator error. We verify this sensitivity using testbed experiments. To produce rotation errors, we first set up a link with accurately calibrated antennas and measure its received signal strength. We then rotate the receive antenna at 1° intervals while recording signal strengths. We repeat this experiment at various link distances and for the case of rotating the transmit antenna. Results in Figure 7 show that a misalignment within 1° leads to negligible impact on the signal strength, and an error of 5° only leads to a RSS degradation of 3.6dB. These results also closely match prior work on Kelleher's universal horn pattern [28]. Thus many existing rotators [7, 4] are sufficiently precise for our needs. Finally, we also verified that 2D and 3D beamforming have the same sensitivity to rotator errors, *i.e.* they receive the same level of performance degradation from rotator misalignments.

Impact of Reflector Materials. Throughout our experiments, our results show that both the cheap, lightweight steel plate and the mirror-quality stainless steel plate offer perfect reflection. This means that 3D beamforming does not require specialized polished metal surfaces, and can be deployed using low-cost metal sheets. Finally, we also tested the suitability of other building materials as reflectors, including standard smooth concrete and plaster walls. The results confirm results from prior studies [30]. For concrete walls, we observe a small signal strength degradation compared to the metal reflectors (roughly 3dB). For plaster walls, the degradation increases to 5dB. While more detailed study is necessary, these initial findings raise the possibility that we may be able to deploy 3D beamforming links without modifying data center ceiling materials.

4. SCHEDULING 60 GHZ LINKS

By forming high-throughput wireless interconnects on-demand, 3D beamforming can deliver additional burst bandwidth to data center applications without pre-provisioning wired capacity among all rack pairs. Prior 60 GHz data center proposals constrained links to neighboring racks, greatly limiting the distance and number of wireless links [23]. In contrast, 3D beamforming connects pairs of racks in large data centers in a single hop using indirect line-of-sight paths.

But to fully utilize the benefits of 3D beamforming, we must carefully schedule transmission links to maximize efficiency and minimize wireless interference. In this section, we identify the key challenges of scheduling 3D beamforming links in data centers, and present a centralized link scheduler to support flexible bandwidth allocation.

4.1 Challenges

Our link scheduler must address three key challenges:

First, designing our scheduler requires an accurate interference model for 3D beamforming links. Given the reflective nature of our beamforming links, the interference a receiver

[3]The number of concurrent links does not grow proportionally with the number of channels because interference patterns are not uniform across links. Such negative effects should gradually diminish in larger data center topologies.

experiences is no longer dominated by energy leakage from the nearest transmitter. Instead, because of the dense deployment of these links in data centers, the main source of interference is the accumulation of signals from the many transmitting neighbors. This *accumulative interference* effect is significant and must be accounted for. Our initial simulations show that if we use conventional pairwise interference models that ignore accumulative interference [34], up to 30% of our scheduled links will fail.

Second, our scheduler needs to handle short-lived traffic bursts [15, 25], and thus must be online. To maximize the number of concurrent transmissions and minimize job execution time, our scheduler must be efficient and lightweight, *i.e.* introduce minimal overhead in control traffic and scheduling delay.

Third, scheduling account for antenna rotation delay. This is particularly important when using horn antennas. Using today's rotators [4, 7], rotation delay ranges between 0.01 and 1 second, which is likely in the range of (and longer than) the full transmission times of some links. Furthermore, since the amount of rotation carried by a mechanical rotator directly affects its lifetime and reliability, we need to minimize such overhead.

Assumptions. We further assume that the centralized scheduler has full knowledge of the rack traffic demands to be carried by the 60GHz network. It receives link requests, generates link schedules periodically, and notifies the scheduled racks with the channel, the radio, and the beam direction they should use. We assume that control messages are sent via a separate control channel independent of the 60GHz wireless network. In practice, the scheduler can either use a provisioned wired network, or use a dedicated local WiFi network for control signaling. We leave the detailed design of control channels as future work.

4.2 Scheduler Design

We propose a greedy scheduling algorithm that addresses these challenges. Our scheduler's primary goal is to schedule as many concurrent links as possible, thus maximizing channel usage and minimizing transmission time. To do so, it derives a "conflict degree" for each link from its accumulative interference, described below. Finally, within the framework of this scheduler, we seek to reduce rotational delay by considering radio orientation in assigning links to radios. We will now describe the components of the scheduler in detail, starting with the estimation of link conflicts.

Conflict Estimation. At the core of the algorithm, we must estimate conflicts to not only calculate the conflict degrees of each link, but also to determine whether a specific radio link can be added to the existing active links without affecting their current link rates. In a nutshell, conflict estimation is based on the calculation of Signal-to-Interference-Noise-Ratio (SINR) for each link.

The scheduler derives the signal and interference using model prediction since the 3D beamforming's propagation environment in data center is very predictable. This avoids additional measurement overhead. The calculation is based on the free-space propagation model (Eq. (1)) verified by our experiments, and also the antenna orientations. We apply Kelleher's universal horn pattern [28, 29] to model the radiation pattern of the horn antenna.

To compute conflict degrees, the scheduler considers each possible pair of link requests assuming they operate on the same channel. Let D_i, D_j denote two link requests, we then calculate D_i's SINR value SINR_{ij} in the presence of D_j as:

$$SINR_{ij} = \frac{S_i}{N + I_{ji}}, \tag{2}$$

where S_i is the signal strength received at D_i's receiver, N is the noise level, and I_{ji} is the interference D_j's transmitter creates on D_i's receiver. D_i and D_j conflict if either has a SINR below the threshold of its required data rate. Thus, the conflict degree of D_i is the number of link requests that conflict with D_i when considered in this way.

To determine if a candidate radio link L_i can be admitted given the presence of scheduled links $\mathbb{L}$ on the same channel, the scheduler calculates $\mathrm{SINR}_{k,\mathbb{L}'}$ for each $L_k \in \mathbb{L}' = \mathbb{L} \cup \{L_i\}$ as the following:

$$SINR_{k,\mathbb{L}'} = \frac{S_k}{N + \sum_{j:L_j \in \mathbb{L}' \setminus \{L_k\}} I_{jk}}, \tag{3}$$

where the notation is identical to those in Eq. (2). If each $\mathrm{SINR}_{k,\mathbb{L}'}$ satisfies the corresponding data rate requirement, then L_i can be scheduled on this channel. This accounts for the interference accumulated by multiple links, and ensures that scheduled links can be active simultaneously without conflict.

Conflict-Degree based Greedy Scheduling. With the goal of minimizing the job completion time, the scheduling problem can be mapped to a traditional graph coloring problem that aims to use the minimal number of colors to color all nodes. In our case, the colors map to 60 GHz frequency channels and time slots. We employ techniques from the graph coloring literature [19] in a greedy fashion, where we schedule rack-level requests in an order based on their *conflict degrees*. The conflict degree d_i of an unscheduled request D_i is defined as the number of other unscheduled requests denoted by set D^C, such that if D_i and any request in D^C are on the same channel, at least one cannot achieve the required data rate.

There is an issue of link preemption in the scheduler. Given long-lived links that provide less than ideal link usage, should the scheduler preempt them, *i.e.* pause them, in order to schedule competing links? In a non-preemptive model, the scheduler keeps the unfinished links untouched, and checks which new ones can be added. This policy ensures that scheduled links will not be disturbed until they complete, thus minimizing the antenna rotation delay and control overhead by interrupting an ongoing link transmission. In a preemptive model, the scheduler pauses ongoing links, treats them as new requests with the remaining unsent traffic, and schedules them together with new link requests. Since previous link requests must compete with new requests, this policy could lead to interruptions to ongoing active links. The benefit, however, is that such a policy could increase the number of concurrent, active links. While we evaluate both policies in Section 5, our default scheduler is non-preemptive.

We note that more complex policies can be added to our scheduler, such as alternative ranking metrics that prioritize job by their deadlines in deadline-driven data centers [43]. In such priority-based scheduling policies, care must be taken to avoid link starvation by gradually increasing the priority of jobs as their waiting time increases. We leave the design of those metric-based schedulers as future work.

Assigning Radios to Links. To minimize the antenna rotation overhead, the scheduler assigns scheduled links to unassigned radios during scheduling. It applies a simple policy. First, if an idle radio on the rack is already pointing to the desired destination rack, the scheduler assigns the link to this radio. Second, if multiple candidate/idle radios exist, the scheduler checks the existing orientations of these antennas and selects the one that is closest in angle to the desired angle for the new link's destination rack.

5. ADDRESSING TRAFFIC HOTSPOTS

In this section, we use network simulations to quantify 3D beamforming's ability to deliver additional bandwidth to data center environments, and its advantages over its 2D counterpart. We consider the case of using wireless links to cover traffic hotspots on top of an existing wired network in data centers. Specifically, we seek to answer three key questions:

1) *Does adding 3D beamforming links to existing wired networks significantly increase available bandwidth for hotspots*?

2) *How significant are the benefits of 3D beamforming over 2D beamforming, and where are they most visible*?

3) *Will antenna rotation delay of today's rotators be a performance bottleneck for 3D beamforming*?

While we answer these questions using results from synthetic traffic traces, we hope trends identified by our study will serve as useful guidelines for practical deployment of 3D beamforming systems.

5.1 Simulation Setup

For our traffic hotspot simulations, we use the data center and radio configurations described in Section 3.2. It consists of 250 racks and a total of 5000 servers. The distance between the antenna and the ceiling (h) is 2m. We use 60GHz radios with 10dBm transmit power and 10° horn antennas. There are three channels of width 2.16GHz, and each radio can operate on one channel at a time. We derive data rates following the specifications of the IEEE 802.11ad standard. For this data center size, every possible pair of racks is able to form a 1-hop link at 5+Gbps using 3D beamforming.

There are 8 radios positioned at the top of each rack. To account for the rotation delay, we assume each antenna uses a rotator from FLIR [4]. The pan speed is 300°/second and the tilt speed is 60°/second. We also examine the case where the rotation is instantaneous, which represents the best possible (ideal) performance for 3D beamforming links.

Traffic Generation. Existing traces [15, 16, 23, 25] do not map data sources to rack locations. So to produce hotspots, we used synthetic traffic generated based on popular workloads [18]. We simulate a simple scenario, where 400 of the 5000 total servers in the data center each sends a fixed data payload to 200 other servers. The 400 servers are chosen randomly from any rack in the data center. Each data payload is 128MBytes, and each of the 400 transmitters chooses a set of 200 servers to receive its data. We refer to a complete cycle where each server sends a single payload to each of its 200 destinations as a single round.

To produce controlled traffic hotspots, we introduce a slight bias in server selection. This might emulate a slight preference for certain machines based on their properties such as compute power, uptime, memory size, or network proximity to storage servers. We identify 50 random servers in the data center as "preferred" servers. In each round, each of the 400 transmitters chooses 200 receivers to receive their data, randomly, but with a small bias. As we choose each of the 200, there is a 10% chance that the receiver server is one of the 50 preferred servers. Once we have chosen 200 unique receivers for each of the 400 transmitters, we aggregate these server-pair traffic loads based on their server locations to produce traffic demands at the rack level. We repeat the above procedure to generate 10 rounds of rack-level workloads.

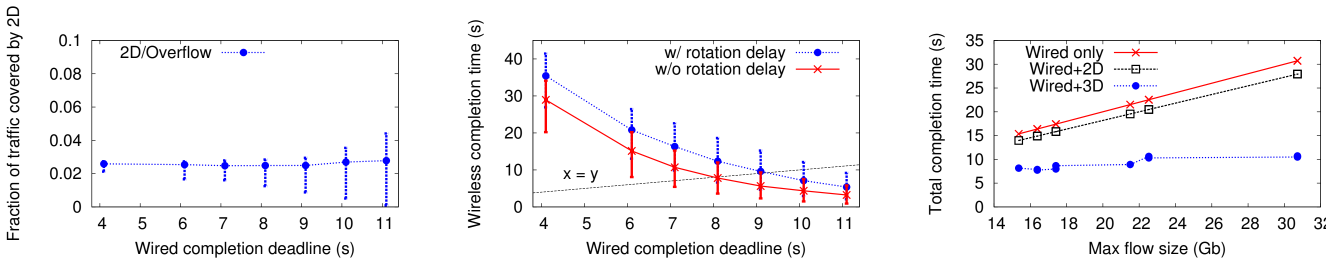
We assume an underlying wired network offering 1Gbps network bisection bandwidth. We also explore larger values ranging from 2 to 6Gbps, which are typical given the over-subscription of today's data centers [14]. Since our goal is to understand how 3D beamforming addresses traffic hotspots, one issue that arises is how traffic demands are split across wired and wireless network links. Without advocating any particular allocation policy, we assume that the traffic is split between the two networks by setting a fixed "deadline" for wired links to finish their portion of the transmission. This allows us to define the portion of traffic sent over the wired network *a priori*, thus deriving the amount of "overflow" traffic allocated to the 60 GHz network. To compute job completion time, we assume that wired and wireless links send data in parallel.

5.2 Impact of Adding 2D/3D Beamforming

Coverage of 2D Links. One of the primary limitations with 2D beamforming is potential blockage issues that result in lower signal strength and loss in data throughput. Prior proposals limited 2D beamforming links to connecting neighboring racks that have no potential blockage issues. Our first experiment looks considers portion of the total overflow traffic can be sent over 2D links, where overflow traffic is the traffic that cannot be sent by the wired network by the specified deadline. Figure 8(a) plots this as a ratio of total traffic across neighboring links, *i.e.* traffic that can be sent across 2D links, over total overflow traffic. In all of our graphs, we plot error bars covering the 90% confidence interval.

Less than 3% of overflow traffic can be addressed using 2D links, regardless of how much traffic is sent across the wired network. We note that this figure might increase, depending on how well the data center managers scheduled jobs to increase rack affinity and limit hotspots to neighboring racks, but this would introduce an additional constraint in job scheduling. Note that we do not plot similar values for 3D links, because 3D links can connect all possible rack pairs in our scenario with a single hop, thus coverage is 100%.

Impact of Antenna Rotation Delay. Next, we look at 3D performance, and try to understand the impact on end-to-end latency by antenna rotation delays. In Figure 8(b), we plot the wireless completion time, *i.e.* time required to send overflow traffic over the 3D links, against the wired completion deadline. A longer wired deadline means more traffic will go over the wired network and less overflow traffic will be left for wireless links. We also draw the line "$y = x$" to show the minimum time to complete transmissions if we used the ideal traffic allocation between wired and wireless. By plotting 3D links with and without rotational delay, we see that all transmissions are completed in 9 seconds in a realistic system, but can be improved to 8 seconds if we



(a) Portion of Traffic Covered by 2D Links

(b) Impact of Antenna Rotation Delay on 3D Performance

(c) Total Completion Time

Figure 8: Performance of 2D and 3D beamforming links in conjunction with a wired network with 1Gbps bisection bandwidth. (a) Less than 3% of overflow traffic can be addressed using 2D links, while 3D links can cover 100%. (b) The time required for 3D to send all the overflow traffic vs. the wired completion deadline, using today's rotators or ideal rotators with zero delay. (c) Adding 3D beamforming cuts the total job completion time by half.

completely eliminate antenna rotation delays. This is the effective upper bound on how well we can perform given perfect radio assignment, and shows that antenna rotation delay is only a small component in end-to-end performance.

Total Completion Time. Next, in Figure 8(c), we plot the effective completion time (if both wired and wireless networks transmitted simultaneously with optimal traffic split) for different rounds of our simulation, assuming a 1Gbps bisection on the wired network. Regardless of the size of the biggest flow, 3D wireless is generally able to reduce total transmission time by half. Since 2D links can only address a small portion of the overflow traffic, its impact on completion time is limited.

Next, we ask the question, "will 3D beamforming links become less useful for wired networks with higher bisection bandwidth?" We vary the underlying bisection bandwidth of the wired network between 1 Gbps and 6 Gbps, and in each case, compute the minimal completion time if we performed the ideal traffic split between the wired and wireless networks. Figure 9 shows that even as the wired network grows in bisection bandwidth, adding 3D beamforming can still reduce total transmission time significantly (ranging from more than 50% to slightly less than 40% as wired bandwidth ranges from 1 Gbps to 6 Gbps). This is not surprising. Since the demand in our hotspot scenario is fixed, a wired network with larger bisection bandwidth consumes more traffic, leaving less overflow traffic for the 60 GHz links.

5.3 Impact of Scheduler Policies

We now evaluate the net impact of different design choices in our link scheduler.

Preemption vs. No Preemption We discussed the issue of preempting existing links in Section 4. To understand which policy is more preferable, we compare the job completion time when applying these two policies with the same workload as above. Non-preemptive is the default policy used in prior experiments, where once a link is scheduled, it utilizes the radio until it finishes. The preemptive policy considers all radios on a rack when scheduling requests, and where deemed appropriate by the scheduler, will pause an existing link to give its radio to a higher priority link request.

We plot the completion time of the wireless traffic load in Figure 10(a). The results match our expectations. A non-preemptive policy reduces the completion time by up to 25% compared to the preemptive policy. This is because rescheduling the paused link introduces significant additional overhead (and more antenna rotation), thus reducing overall efficiency.

Wired completion time (s)	4	6	7	8
Rotator usage reduction via intelligent radio assignment	27%	22%	22%	21%

Table 1: Intelligent radio assignment leads to moderate reduction in rotator usage.

Intelligent Radio Assignment. Finally, we evaluate the impact of using intelligent radio assignment during link scheduling. Figure 10(b) plots the completion time for the wireless traffic load for our default scheduler (including intelligent radio assignment) and a basic scheduler, which uses random choice to choose between available radios. Neither scheduler uses link preemption.

The results are varied, depending on the traffic load on the wireless network. When the large majority of traffic is sent through the wired network, only the strongest of the hotspots remain for the wireless network. In this case, there will be high contention for radios at a small number of racks, and very little choice in terms of radio assignment. Thus the optimization shows small benefit. When a larger portion of the traffic is delegated to the wireless network, more racks carry moderate traffic that leaves some number of radios free for assignment. In this case, optimizing radio assignment provides some moderate benefit, which is ultimately bounded by the overhead of antenna rotation delay, shown in Figure 8(b).

We also compare the amount of rotations performed by each rotator using the two schedulers. Results in Table 1 show that depending on the traffic load given to the wireless network, intelligent radio assignment can reduce the rotator usage by up to 27%.

Summary. Our findings in this section can be summarized as follows. *First*, we find that in data centers with random traffic patterns, 2D beamforming restricted to neighboring racks can only address a very limited ($\sim$ 3%) portion of traffic hotspots, compared to 100% for single hop 3D beamforming links. *Second*, we find that for many scenarios involving bursty traffic hotspots, using 3D beamforming links in conjunction with the existing wired network can generally reduce completion time by half or more. *Finally*, we

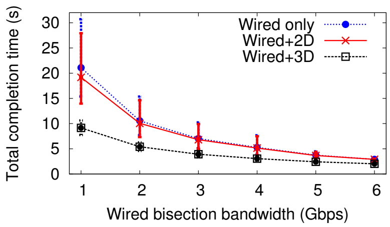
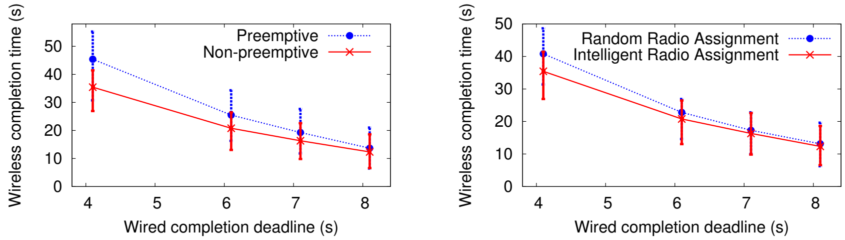


Figure 9: Overall impact of adding beamforming links.



(a) Preemption vs. Non-preemption (b) Intelligent Radio Assignment

Figure 10: Impact of different scheduling policies.

find that when sizable payloads are involved, *e.g.* 128MB, antenna rotation delays only contribute a small portion of the overall completion time, and much of that can be recovered using simple heuristics such as choosing radios that are closer to the desired transmission angle.

6. DEPLOYMENT CHALLENGES

In this section, we briefly discuss some key challenges of deploying 3D beamforming in data centers.

Physical Rack/Reflector Placement. 3D beamforming performs the best when there are no obstacles between the top of rack/container and the ceiling. When physically arranging racks/containers as well as ceiling reflectors, data center managers should avoid obstacles such as cables and cooling pipes. This is not an issue for container-based data centers [10]. For other types of data centers, raised floors can be used to house cables and pipes in the ground. Any suspended cable trays can be concealed within aluminum-plated ducts, essentially lowering the reflection point from the ceiling. When unavoidable, one can also plan multi-hop transmissions or reflect off walls to route around obstacles. An open question is whether physical rack and reflector placement can be jointly optimized with network communication patterns.

Reflector Curvature. We observe in our experiments that the performance of 3D beamforming is sensitive to the curvature of the reflector. Reflected by a flat surface, the beam propagates following the free-space model. When the surface becomes slightly concave, we observe a visible increase in the measured RSS value. Similarly, we observe a drop when the surface becomes convex. This could be the result of reflection creating multipath signals to the receiver, which could degrade the link performance. Ideally, the reflector should be kept as flat as possible. Yet an open question is whether one can manipulate reflector curvature to further improve 3D beamforming performance.

Rotator Reliability. To communicate with different racks, each transceiver must adjust its beam direction in both azimuth and elevation. For radios with horn antennas, this requires a mechanical rotator to rotate the antenna, leading to extra rotation delay and the issue of rotator reliability. Today's off-the-shelf pan-tilt rotators can provide roughly 3-5 million cycle durability [4] or 15,000-hour lifetime [9]. Clearly, the failure rates of these components must be weighed against their performance when choosing between horn antennas and antenna arrays.

Antenna Arrays. The above concerns on the rotator delay and reliability can be addressed by replacing horn antennas with switched beam smart antennas or antenna arrays. Antenna arrays use electronic beam rotation, with delay as low as 50ns in existing 16-element arrays [40]. However, they still do not eliminate issues of link blockage or interference. In addition, antenna arrays produce more signal leakage than horn antennas, leading to weaker link signal strength and stronger interference to neighboring flows [46]. One potential solution is to configure beam patterns to steer away from neighboring links or to nullify interference [33]. We leave exploration of these issues for future study.

7. RELATED WORK

Data Center Networks. Most prior work addresses traffic congestion through network architecture design and traffic scheduling [11, 13, 14, 18, 21, 22], or modeling network traffic characteristics [15, 16, 25, 17]. 60 GHz wireless was first proposed to data center networking in [35] as a solution to reduce the cabling complexity. Recent proposals use wireless links to augment [23, 26] or replace [38] wired links. In this work, we identify two practical issues of existing approaches, and propose a new beamforming paradigm to effectively address these issues.

While prior work has applied the principle of radio signal reflection to connect non line-of-sight links [35] or to reduce WiFi interference [31], we generalize it to 60GHz links. Our key contribution is to use ceiling reflection in the data center to extend connectivity and suppress interference. We also used detailed hardware experiments to validate our design.

Optical circuit switching [20, 39, 42] is an alternative for adding burst bandwidth to data centers. Optical circuit switching promises tremendous bandwidth but the technology incurs relatively substantial cost and does not offer some of the benefits of wireless augmentation of data center bandwidth, namely reduced cost and deployment complexity.

60GHz Wireless Technology. Prior work mainly focuses on radio and antenna design issues [8, 40], and signal propagation and reflection modeling [37, 45, 12, 30, 36]. Today, there is a wide selection of commercial 60GHz radio products [5]. One recent effort [40] developed a 16-element on-chip antenna array, allowing each radio to switch a beam to any of the 32 predefined directions within 50ns. Our work leverages readily available hardware, and focuses on designing new wireless interconnects explicitly for data centers.

8. CONCLUSION

Traffic in today's data centers is unpredictable, often producing traffic hotspots that result in congestion and delay. Instead of overprovisioning the wired network for bursts by rewiring a data center network at scale, we advocate the use of 60 GHz wireless beamforming links to alleviate traffic hotspots as they occur.

Our work addresses limitations of 60 GHz beamforming that arise from signal blockage and interference caused by signal leakage. Our insight is that by aiming 60 GHz beamforming links at a reflective ceiling, we can achieve indirect line-of-sight between most or all rack pairs in a data center, while minimizing interference. The net effect is that 3D beamforming greatly expands the reach and capacity of 60 GHz links, making them feasible as flexible and reconfigurable alternatives to wired cabling. Our testbed measurements confirm that 3D beamforming links suffer zero energy loss from reflection, and effectively avoid blocking obstacle and reduce interference footprint.

While wired networks will continue to serve high-end data center needs, we believe that efforts such as 3D beamforming can provide significant benefits to a broad range of data center deployments, by potentially reducing deployment complexity and reducing cost compared to a fully provisioned wired network.

9. ACKNOWLEDGMENTS

The authors thank Romit Roy Choudhury and the reviewers for their feedback, and Daniel Halperin and Lei Yang for their insights on 60 GHz hardware. This work is supported in part by NSF grant CNS-0905667.

10. REFERENCES

[1] Data center managers not fixed on wireless. `http://searchdatacenter.techtarget.com/tip/Data-center-managers-not-fixed-on-wireless`.

[2] ETS-Lindgren. `http://ets-lindgren.com/Absorbers`.

[3] Flann Microwave. `http://www.flann.com/`.

[4] FLIR. `http://www.flir.com/mcs/products/`.

[5] HXI Milimeter Wave Products. `http://www.hxi.com/`.

[6] IEEE P802.11ad/D0.1: Enhancements for very high throughput in the 60 GHz band. Draft 0.1, June 2010.

[7] NSI. `http://www.nearfield.com/`.

[8] SiBeam. `http://sibeam.com/whitepapers/`.

[9] Sidus. `http://www.sidus-solutions.com/`.

[10] Sun modular datacenter s20/d20 overview. `http://download.oracle.com/docs/cd/E19115-01/mod.dc.s20/index.html`.

[11] Abu-Libdeh, H., et al. Symbiotic routing in future data centers. In *Proc. of SIGCOMM* (2010).

[12] Ahmadi-Shokouh, J., et al. Reflection coefficient measurement for house flooring materials at 57-64 GHz. In *Proc. of GLOBECOM* (2009).

[13] Al-fares, M., et al. Hedera: Dynamic flow scheduling for data center networks. In *Proc. of NSDI* (2010).

[14] Al-Fares, M., Loukissas, A., and Vahdat, A. A scalable, commodity data center network architecture. In *Proc. of SIGCOMM* (2008).

[15] Benson, T., Akella, A., and Maltz, D. A. Network traffic characteristics of data centers in the wild. In *Proc. of IMC* (2010).

[16] Benson, T., et al. Understanding data center traffic characteristics. *ACM CCR 40* (2010), 92–99.

[17] Chen, Y., Ganapathi, A., Griffith, R., and Katz, R. The case for evaluating MapReduce performance using workload suites. In *Proc. of MASCOTS* (2011).

[18] Curtis, A., Kim, W., and Yalagandula, P. Mahout: Low-overhead datacenter traffic management using end-host-based elephant detection. In *INFOCOM* (2011).

[19] Diestel, R. *Graph Theory*, 4th ed. Springer-Verlag, Heidelberg, 2010.

[20] Farrington, N., et al. Helios: A hybrid electrical/optical switch architecture for modular data centers. In *Proc. of SIGCOMM* (2010).

[21] Greenberg, A., et al. VL2: A scalable and flexible data center network. In *Proc. of SIGCOMM* (2009).

[22] Guo, C., et al. BCube: a high performance, server-centric network architecture for modular data centers. In *Proc. of SIGCOMM* (2009).

[23] Halperin, D., et al. Augmenting data center networks with multi-gigabit wireless links. In *Proc. of SIGCOMM* (2011).

[24] Kamil, S., et al. Communication requirements and interconnect optimization for high-end scientific applications. *IEEE TPDS 21* (2009), 188–202.

[25] Kandula, S., et al. The nature of data center traffic: measurements & analysis. In *Proc. of IMC* (2009).

[26] Kandula, S., Padhye, J., and Bahl, V. Flyways to de-congest data center networks. In *Proc. of HotNets* (2009).

[27] Katayama, Y., et al. Wireless data center networking with steered-beam mmwave links. In *WCNC* (2011).

[28] Kelleher, K. *The Microwave Engineers' Handbook and Buyers' Guide*, 5th ed. New York: Horizon Press, 1964.

[29] Knop, C., and Wiesenfarth, H. On the radiation from an open-ended corrugated pipe carrying the HE11 mode. *IEEE Trans. on Antennas and Propagation* (1972), 644 – 648.

[30] Langen, B., Lober, G., and Herzig, W. Reflection and transmission behavior of building materials at 60 GHz. In *Proc. of PIMRC* (1994).

[31] Liu, X., et al. DIRC: Increasing indoor wireless capacity using directional antennas. In *Proc. of SIGCOMM* (2009).

[32] Manabe, T., Miura, Y., and Ihara, T. Effects of antenna directivity on indoor multipath propagation characteristics at 60GHz. In *Proc. of PIMRC* (1995).

[33] Nikolaidis, G., et al. Cone of silence: adaptively nulling interferers in wireless networks. *SIGCOMM CCR* (2010).

[34] Qiu, L., et al. A general model of wireless interference. In *Proc. of MobiCom* (2007).

[35] Ranachandran, K., et al. 60GHz data-center networking: wireless => worryless? *NEC Technical Report* (2008).

[36] Sato, K., et al. Measurements of reflection and transmission characteristics of interior structures of office building in the 60-GHz band. *IEEE Trans. on Antennas and Propagation 45*, 12 (1997), 1783 –1792.

[37] Schönthier, J. WP3-study "the 60 GHz channel and its modelling". *Tech. Report IST-2001-32686* (2003).

[38] Shin, J.-Y., et al. On the feasibility of completely wireless data centers. *Technical Reports, Cornell University* (2011).

[39] Singla, A., et al. Proteus: a topology malleable data center network. In *Proc. of HotNets* (2010).

[40] Valdes-garcia, A., et al. Single-element and phased-array transceiver chipsets for 60-GHz Gb/s communications. *IEEE Communications Magazine* (2011).

[41] Vardhan, H., et al. Wireless data center with millimeter wave network. In *Proc. of GLOBECOM* (2010).

[42] Wang, G., et al. c-through: part-time optics in data centers. In *Proc. of SIGCOMM* (2010).

[43] Wilson, C., Ballani, H., Karagiannis, T., and Rowstron, A. Better never than late: Meeting deadlines in datacenter networks. In *Proc. of SIGCOMM* (2011).

[44] Yang, F., Shanmugasundaram, J., and Yerneni, R. A scalable data platform for a large number of small applications. In *Proc. of CIDR* (2009).

[45] Yong, S.-K., Xia, P., and Valdes-Garcia, A. *60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice*. WILEY, 2011.

[46] Zhang, W., Zhou, X., Yang, L., Zhang, Z., Zhao, B. Y., and Zheng, H. 3D beamforming for wireless data centers. In *Proc. of HotNets* (2011).