

FluidNet: A Flexible Cloud-based Radio Access Network for Small Cells

Karthikeyan Sundaresan
NEC Labs America, Inc.
karthiks@nec-labs.com

Mustafa Y. Arslan
NEC Labs America, Inc.
marslan@nec-labs.com

Shailendra Singh
UC Riverside
singhs@cs.ucr.edu

Sampath Rangarajan
NEC Labs America, Inc.
sampath@nec-labs.com

Srikanth V.
Krishnamurthy
UC Riverside
krish@cs.ucr.edu

ABSTRACT

Cloud-based radio access networks (C-RAN) have been proposed as a cost-efficient way of deploying small cells. Unlike conventional RANs, a C-RAN decouples the baseband processing unit (BBU) from the remote radio head (RRH), allowing for centralized operation of BBUs and scalable deployment of light-weight RRHs as small cells. In this work, we argue that the intelligent configuration of the front-haul network between the BBUs and RRHs, is essential in delivering the performance and energy benefits to the RAN and the BBU pool, respectively.

We then propose *FluidNet* - a scalable, light-weight framework for realizing the full potential of C-RAN. *FluidNet* deploys a logically re-configurable front-haul to apply appropriate transmission strategies in different parts of the network and hence cater effectively to both heterogeneous user profiles and dynamic traffic load patterns. *FluidNet*'s algorithms determine configurations that maximize the traffic demand satisfied on the RAN, while simultaneously optimizing the compute resource usage in the BBU pool. We prototype *FluidNet* on a 6 BBU, 6 RRH WiMAX C-RAN testbed. Prototype evaluations and large-scale simulations reveal that *FluidNet*'s ability to re-configure its front-haul and tailor transmission strategies provides a 50% improvement in satisfying traffic demands, while reducing the compute resource usage in the BBU pool by 50% compared to baseline transmission schemes.

Categories and Subject Descriptors

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

Keywords

Cellular, Cloud RAN

1. INTRODUCTION

Mobile network operators are facing the pressure to increase the capacity and coverage of their radio access networks to meet the ex-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MobiCom'13, September 30-October 4, Miami, FL, USA.
Copyright 2013 ACM 978-1-4503-1999-7/13/09
DOI string here from ACM form confirmation ...\$15.00.

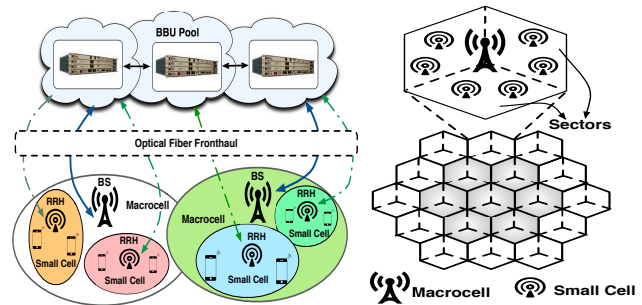


Figure 1: C-RAN Architecture. Figure 2: Network Deployment.

ponential growth in data traffic demand [8]. While leveraging the increased spatial reuse from smaller cells is a promising direction, every new cell adds to the capital and operational expenses borne by the operators. To address this problem, cloud-based radio access network (C-RAN) architectures have been considered by several operators [17] and service providers [15] as a cost-efficient way of realizing small cells. Unlike typical RANs where the baseband units (BBUs) and the radio units are situated together, the C-RAN concept (depicted in Fig. 1) migrates the BBUs to a datacenter (i.e., the BBU pool) hosting high performance general purpose and DSP processors, while providing high-bandwidth optical transport to the remote antennas called remote radio heads (RRHs). We define the high-bandwidth optical transport that carries the cellular signals between the BBUs and the RRHs to be the *front-haul* part of the network, whose bandwidth requirements could be significantly higher (tens of Gbps) than that of the backhaul depending on the nature of the signals (digital/analog, layer 1/2) carried [17]. The decoupling of the BBUs and radio units in a C-RAN allows for sophisticated centralized techniques for interference management, where the BBUs in the pool can seamlessly cooperate to improve the RAN capacity. In addition, the deployment of radio units is made light-weight and can be realized in a fast and scalable manner for small cells (other benefits of C-RAN are detailed in [17]).

In this work, we argue that the front-haul that is unique to a C-RAN has a critical role in delivering its performance and cost benefits. We note that although the BBUs are decoupled from the RRHs in terms of physical placement, there exists a one-to-one logical mapping between BBUs and RRHs in that one BBU is assigned to generate (receive) a signal (e.g., LTE or WiMAX frame) to (from) an RRH (although the mapping can change over time). This one-to-one mapping allows for generating a distinct frame for each small cell (deployed in the form of a RRH), which is key for enhancing

the network capacity via techniques such as dynamic fractional frequency reuse (dynamic FFR [4]) or coordinated multi-point transmissions (e.g., LTE CoMP [22]). We contend that this notion of a fixed, one-to-one mapping is not optimal in a practical cellular network deployment for two reasons.

RAN Performance: First, these techniques primarily apply to static users. The mobile users will have to bear frequent handoffs (exacerbated by smaller cells) and the associated performance penalties. In addition, tracking a mobile user’s location and channel may be difficult for such techniques. In fact for mobile clients, a traditional distributed antenna system (DAS [11]) is arguably better suited. In a DAS setting, the same signal (carrying the user’s data) is transmitted simultaneously by multiple small cells to provide coverage benefits (which in turn reduces handoffs) and diversity gain. DAS can be realized by changing the one-to-one to a one-to-many logical mapping in the C-RAN front-haul.

BBU Energy Consumption: Second, the one-to-one mapping requires several BBUs to be active and generating frames, which consumes energy in the BBU pool. However, the enhanced capacity of techniques such as [4, 22] may not be needed in all parts of the network or at all times (e.g., 50% of cells carry 5% of net traffic [5]). When the traffic load is low in a region (e.g., coverage area of multiple small cell RRHs), a single BBU may suffice to serve the offered load (via a DAS mapping). This in turn reduces the number of BBUs and hence the compute resources (e.g., CPU cores, DSPs), thereby allowing energy savings in the BBU pool.

Given the above observations, we envision a C-RAN architecture with a novel, flexible front-haul that supports one-to-one as well as one-to-many logical mappings between BBUs and RRHs. Our vision is to utilize this architecture *to address the traffic needs of users (static and mobile) while leveraging the energy savings made possible by the traffic load heterogeneity (i.e., temporal and spatial load variations in the network)*.

Towards realizing this vision, we present *FluidNet* - a flexible C-RAN system for small cells that houses an intelligent controller in the BBU pool, which dynamically re-configures the front-haul (at coarse time scales) based on network feedback to cater effectively to both heterogeneous user and traffic profiles. This allows *FluidNet* to maximize the amount of traffic demand satisfied on the RAN for both static and mobile users, while at the same time optimizing the compute resource usage in the BBU pool. Briefly, *FluidNet* adopts a two-step, scalable approach: based on spatial traffic distribution and demand from users, *FluidNet* first determines the optimal combination of configurations (one-to-one and one-to-many, i.e., DAS and FFR strategies) needed to support the traffic demand from a set (termed *sector*) of small cells. Then, it employs a novel and efficient algorithm (with an approximation factor of $\frac{3}{2}$) to consolidate (cluster) the configurations of multiple sectors in the network to further reduce the compute resource usage without compromising on the traffic demand satisfied. *FluidNet* is both standards and technology agnostic. It allows for desirable features such as co-existence of multiple mobile operators and technologies (LTE, WiMAX, WiFi) in the same C-RAN, while employing different front-haul configurations tailored to each of their respective traffic.

We prototype *FluidNet* on a small-scale WiMAX C-RAN testbed with 6 BBUs and 6 RRHs, employing radio-over-fiber (RoF) as the front-haul. With *FluidNet*’s algorithms, the logical BBU-RRH configurations are determined and executed on the fly. Real-world experiments with COTS WiMAX clients show that featuring flexible front-haul configurations and hence strategies, allows *FluidNet* to provide a 50% improvement in traffic demand satisfaction, while also reducing the compute resource usage in the BBU pool by 50%

compared to baseline DAS and FFR strategies. Complementary, standards-calibrated (3GPP) simulations for large networks show that the clustering component in *FluidNet* helps further reduce the compute resource usage by 50% during low traffic load periods. Our contributions are as follows:

- We propose *FluidNet* - a light-weight, scalable framework to determine the optimal use of strategies (DAS, FFR) to cater to dynamic user and traffic profiles, while realizing them through appropriate configurations that help minimize compute resource usage in the BBU pool.
- We design efficient algorithms with performance guarantees in determining the appropriate configurations.
- We build a small-scale C-RAN system with 6 BBUs-RRHs; prototype *FluidNet* on it; and conduct over-the-air experiments, complemented by standards-calibrated large-scale simulations to demonstrate its feasibility and benefits.

2. BACKGROUND

2.1 C-RAN Primer and Related Work

The C-RAN architecture, depicted in Fig. 1, includes three components: (i) remote radio heads (RRH), (ii) pool of baseband units (BBUs), and (iii) the front-haul (optical fiber based transport network).

RRHs: These are simple, light-weight radio units with antennas. Several proposals have focused on making RRHs power-efficient and scalable (e.g., [15, 9]) to support multiple bands and technologies (e.g., 3G, 4G).

BBU Pool: This helps migrate bulk of the base station (BS) processing of a large set of cells to a datacenter [17], allowing for easier realization of interference (e.g., CoMP [22, 3], HetNet [13]) and mobility management solutions.

On the energy front, [7, 21, 16] have looked at the benefits of switching off entire macrocell BSs based on prevailing traffic conditions. Moving the processing to a central entity in C-RAN allows for fine-grained use of resources in the pool and hence better energy savings (evaluated in Section 7). Further, these savings can be obtained without having to switch off an entire BS (allowing RRHs to be ON) and hence sacrificing performance or coverage.

For the BBU pool, there are several proposals for the use of heterogeneous platforms consisting of general-purpose processors as well as DSPs for compute-intensive baseband functions [10, 12]. Recently, [6] focused on assigning processor cores in a homogeneous platform to different BBUs in the pool, to meet latency requirements. Being complementary to [6], we focus on optimizing the use of BBUs themselves, which has an impact not only on compute resource usage in the BBU pool (especially in a heterogeneous platform) but also on RAN performance.

Front-haul: Optical fiber with wavelength multiplexing serves as the front-haul and distributes signals from the BBU pool to the RRHs either as (i) digitized radio signals over CPRI (common public radio interface) [1], or (ii) analog radio signals via radio-over-fiber (RoF) [19]. While CPRI is more robust than RoF over long distances, it requires more transport bandwidth. Optical front-haul is already used in several DAS deployments [11]. Recently, [14] articulated the need for a re-configurable front-haul in a C-RAN, but did not offer a solution. Our focus is to design and build a dynamically re-configurable front-haul along with the intelligence to adaptively determine the appropriate configurations.

2.2 Overview of Strategies

Fractional Frequency Reuse (FFR): FFR is the mechanism for radio resource management (RRM) in cellular networks, whereby inter-cell interference is addressed. Unlike WiFi, the synchronous operation of downlink (BS-MS) and uplink (MS-BS) transmissions across cells requires transmissions to be intelligently scheduled to manage interference. In the popular 1-3 FFR scheme for macrocell networks, the spectrum is divided into four fixed-size bands. One band is used by all the cell-interior clients (in each cell), who do not see interference due to the close proximity to their BS, while the other three bands are used (by cell-exterior clients) in an orthogonal manner between the three sectors (Fig. 2) of a cell to mitigate interference with sectors of adjacent cells. Thus, while the band used by cell-interior clients is reused in each cell, the reuse of the other three bands are subject to the spatial reuse possible. Recently, dynamic FFR approaches [4] have been proposed specifically for small cells, and determine the number and size of bands to be used by each small cell only based on the aggregate traffic demand from its cell-interior and cell-exterior clients; they allow for better spectral utilization and do not rely on planned sectorization (unlike macrocells). Note that the FFR schemes only determine the set of spectral resources assigned to cells - scheduling of clients within those resources is done by each cell locally (based on per-client feedback) to leverage multi-user diversity.

We adopt [4] for FFR in *FluidNet*, although other FFR schemes can also be easily used. While point-to-point MIMO is automatically incorporated in FFR, other cooperative techniques such as multi-user MIMO and co-ordinated multi-point transmissions (CoMP) can also be applied under FFR.

Distributed Antenna Systems (DAS): In DAS, a common signal from a single source is delivered to multiple RRHs and transmitted simultaneously. This provides larger coverage and has been adopted by operators for both indoors and outdoors [11]. Unlike FFR that is focused on capacity, increasing the foot-print of the signal across multiple transmit points (small cells) under-utilizes the spectrum in DAS without scope for any spectral reuse.

Relation between Strategies and Configurations: Since interfering cells will be operating on potentially different spectral bands in FFR, different frames (with specific preamble, control, etc.) have to be generated for each cell, thereby requiring a one-to-one logical mapping (configuration) between a BBU and an RRH. This is the conventional mapping considered in C-RAN currently. However, in DAS, a single frame is transmitted by multiple RRHs, which in turn can be accomplished using a single BBU, thereby requiring a one-to-many mapping.

3. MOTIVATION AND MODEL

3.1 Motivation for a Reconfigurable Front-haul

With the help of a simple experiment conducted on a WiMAX C-RAN testbed (details in Section 6), we now motivate why a one-to-one signal mapping between BBUs and RRHs is highly sub-optimal. Consider a system with 3 BBUs and 3 RRHs, serving three clients as shown in Fig. 3, where each RRH interferes with its neighbor's client.

1. **Traffic Heterogeneity:** Consider a scenario, where the clients are static, but their data rate varies (see Fig. 4(b)). When the total rate (e.g., 8 Mbps per client = 24 Mbps) exceeds the max. data rate supported by all the sub-channels in one frame (≈ 16 Mbps in our testbed), the increased capacity with FFR (by reusing orthogonal half of sub-channels as in Fig. 3) is essential to meet the traffic demand, while DAS is limited to

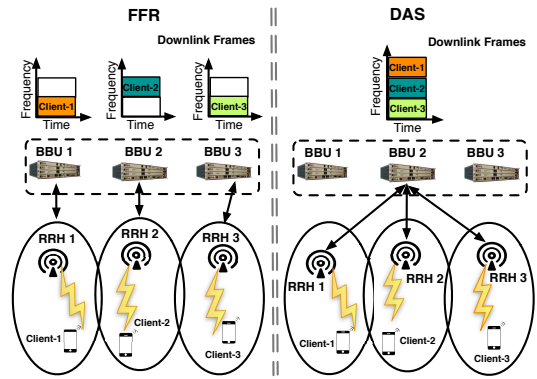


Figure 3: 3 BBU-RRH setup for DAS vs FFR.

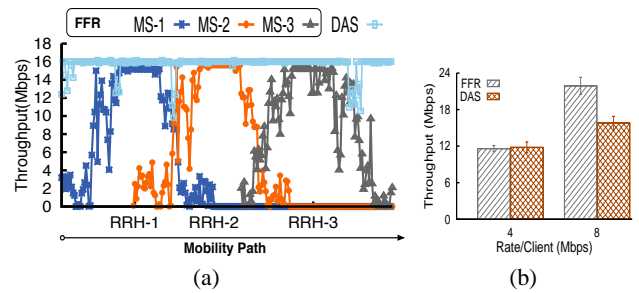


Figure 4: (a) Effect of mobility. (b) Effect of traffic load.

one frame's capacity of 16 Mbps. On the other hand, at low load (e.g., 4 Mbps per client), DAS's capacity is sufficient to serve the clients with just one BBU, allowing the other two BBUs to be off. This is unlike in FFR, where all the BBUs have to be active to generate different frames to the RRHs; it lowers compute resource usage and thus enables significant energy savings in the BBU pool.

2. **User Heterogeneity:** Now, let all the three clients be mobile, moving between the 3 RRHs. Catering to the mobile clients through dynamic FFR from individual RRHs is very challenging for multiple reasons - (a) with small cells, there are frequent handoffs, whose associated latency has an adverse impact on throughput, (b) there is increased signaling load on the front-haul and the mobile core network due to frequent handoffs, (c) it may be hard to track the mobile user to specific small cells to efficiently apply dynamic FFR and leverage reuse. Even notwithstanding such drawbacks and assuming ideal handoffs for FFR, Fig. 4(a) clearly shows DAS' ability to deliver consistent coverage and performance for mobile users. Note that while relegating the mobile user to the macrocell is one option, DAS is ideal for such mobile users, since it achieves a coverage similar to a macrocell, while also increasing the link capacity (through shorter links and diversity gain, see Fig. 3).

Thus, while dynamic FFR is best suited for static users in high traffic load conditions, employing DAS also has benefits both from RAN performance (for mobile traffic) as well as BBU resource usage (for low traffic) perspectives. Given this, it is imperative for the front-haul to be re-configurable to realize flexible combinations of one-to-one and one-to-many BBU-RRH mappings.

3.2 Problem Definition

3.2.1 Network Model

Given that small cells have to co-exist with macrocells, we consider a large number of small cells to be deployed as an under-layer to an operator's macrocell network (Fig. 2). Since a macrocell will interfere with the small cells, thereby limiting their spatial reuse ability, two kinds of RRM solutions have been considered in literature: (i) macrocells and small cells operate on different carrier frequencies, and (ii) both use the same frequency but orthogonalize their radio resources at coarse time scales. We consider the former model and focus entirely on the downlink operation of small cells for ease of exposition. However, our solutions are equally applicable to the latter model and to uplink as well. Further, while small cells themselves could be deployed in an un-planned manner, we leverage the over-layer of macrocells and borrow the notion of logical sectors (from macrocells) to refer to the location of small cells.

3.2.2 Objective

Recall from Section 3.1 that while FFR (one-to-one configuration) supports the maximum amount of traffic through reuse, it does not save on computing resources in the BBU pool. On the other hand, while DAS (one-to-many configuration) minimizes the resource usage and caters to mobile traffic, it under-utilizes the spectrum. By appropriately employing FFR and DAS in combination in different parts of the network, *FluidNet*'s goal is to strike a fine balance between them. Specifically, subject to the primary requirement of supporting as much traffic (D) as the optimal configuration (D_{OPT}), *FluidNet* strives to minimize the corresponding amount of compute resources needed in the BBU pool (resource usage RU , defined in Sec. 4) for the purpose.

$$\min_{\Gamma} RU_{\Gamma}, \quad \text{subject to } D \geq \lambda \cdot D_{OPT} \quad (1)$$

where Γ represents a possible configuration, and λ is the fraction of (optimum) traffic demand that must be satisfied (e.g., $\lambda = 0.99$). The optimal configuration would depend on the relative composition of mobile and static traffic and their priorities ($D_{OPT} = D_{FFR}$ when there is only static traffic demand). We assume mobile traffic to be prioritized over static traffic, albeit other models are also possible. Also note that minimization of compute resource consumption is only subject to satisfying as much of the traffic demand as possible and does not come at the expense of the latter.

BBU Usage as a resource metric: The main components of energy consumption in a traditional base station (BS) are those of air conditioning (≈ 2 KW) and the BS equipment itself (≈ 0.7 KW) [17]. A C-RAN system helps towards both these components by not only simplifying the cell site to a RRH (eliminating the need for air conditioning), but also consolidating the BS processing in the BBU pool. With respect to the latter component, reducing the number of BBU units and hence the frames that need to be processed, has a direct impact on energy consumption for two reasons. (1) BBU processing involves layer 1 (framing, FFT/IFFT, decoding, etc.), layer 2 (HARQ, resource/QoS scheduling, etc.) and layer 3 (connection management) functions. While layer 3 and part of layer 2 can be handled by generic processors, some of the time-sensitive layer 2 (resource scheduling) and layer 1 (framing, FFT/IFFT, decoding) functions are typically handled by dedicated DSPs for each BBU. (2) When DAS is employed, the traffic demand of multiple cells is handled without any spectral reuse. Hence, while the (traffic) load-dependent processing component is limited to that needed to handle the total number of slots (e.g., resource blocks in LTE) in a single frame, the basic processing component (FFT/IFFT) scales with the number of cells (frames) and

soon dominates the former (see [6] for realistic values). Note that optimizing the BBU usage is complementary to assigning compute resources (e.g., GPPs) to the BBUs themselves, for which solutions such as [6] can be leveraged.

4. DESIGN ELEMENTS IN FLUIDNET

We motivate *FluidNet*'s design by addressing key aspects relevant to the operation of transmission strategies and its impact on the compute resource usage in the BBU pool.

4.1 Granularity and Choice of Configurations

A strategy (configuration) is applied to a set of small cells. In macrocells, each sector has its own cell ID and is the smallest granularity for RRM operations. Given this, *FluidNet* adopts *sector* (referring to set of small cells located within the logical sector) to be the minimum granularity for configurations.

Depending on the user and traffic profiles in a sector, one has to determine the appropriate transmission strategy: DAS or FFR. However, picking either DAS or FFR in *isolation* often results in insufficient or spare spectral resources respectively, in handling the offered traffic load. Hence, *FluidNet* employs a flexible *combination* of DAS and FFR (called *hybrid* configurations) in each sector. It devotes the right fraction of spectral resources between the two configurations, thereby supporting the offered traffic load with the least possible use of BBU resources.

4.2 Realization of Hybrid Configurations

Since two configurations cannot co-exist in the same time-frequency resource, hybrid configurations have to be multiplexed either in time or frequency. If multiplexed in time, a hybrid configuration can be realized at the granularity of an epoch spanning several super-frames (10 ms each in LTE), where a contiguous subset of the sub-frames (1 ms each) operate in a DAS configuration, while the rest operate in FFR. If multiplexed in frequency, the operator's spectrum can be divided into coarse spectral blocks (separate carriers in a multi-carrier scenario such as LTE-advanced; e.g., similar to orthogonal channels in WiFi), which are then split between the two configurations (see Fig.5). The fraction of carriers allocated to the configurations is such that the traffic load is satisfied with the least possible use of BBU resources. Since a DAS configuration minimizes the use of BBU resources but supports the least amount of traffic, this is equivalent to finding the largest allocation to the DAS configuration that is capable of sustaining the offered load.

Note that, frequency-multiplexing allows appropriate number of BBU resources to be assigned to each carrier (based on the configuration using it), which do not have to be changed unless the hybrid configuration itself is updated (which happens at coarse time scales; order of minutes). This is unlike time-multiplexing, where the assignment of BBU resources has to be re-mapped even within a hybrid configuration, i.e. switches between DAS and FFR (granularity of super-frames - tens of ms). Although feasible, the time scales of the latter may limit the potential for resource and energy savings in the BBU pool. Hence, *FluidNet* adopts multiplexing configurations in the frequency domain.

4.3 Clustering for Reduced Resource Usage

In regions of the network with low traffic load, it is possible to support the traffic demand from multiple sectors jointly with a single DAS configuration. While aggregating such sectors reduces the compute resource usage in the BBU pool, it must be done in a scalable manner. *FluidNet* proposes a novel clustering mechanism for this purpose.

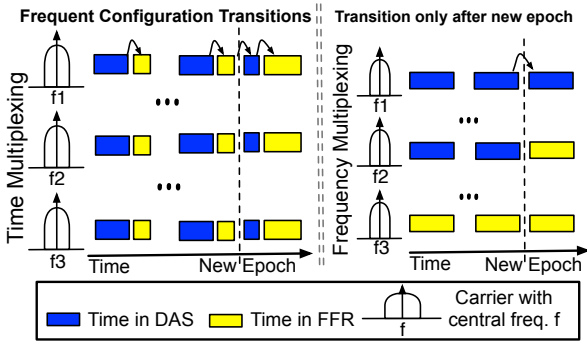


Figure 5: Realizing Hybrid Configurations.

To capture the BBU resource usage for a hybrid configuration in a sector, we define the resource usage metric, RU:

$$RU(b_i, n_i) = b_i \cdot 1 + (B - b_i) \cdot n_i \quad (2)$$

where, n_i is the number of small cells in sector i and b_i , the number of carriers (out of B total) allocated to its DAS configuration. In every carrier, the number of BBU units needed for DAS is one, while it is equal to the number of small cells (n) for FFR. Thus, RU captures the effective number of BBU units needed to support the offered load on the given spectral resources (OFDMA resources in B carriers).

Using the RU metric, *FluidNet* employs a scalable algorithm (details in Section 5) that clusters two neighboring sectors (i and j) at a time, until either their net offered load cannot be supported or the RU of the resulting cluster ($i \cup j$) cannot be improved, i.e.,

$$RU(b_{i \cup j}, n_i + n_j) > RU(b_i, n_i) + RU(b_j, n_j) \quad (3)$$

where $b_{i \cup j}$ captures the new split of carriers between DAS and FFR in the cluster). While applying DAS to serve user traffic on $b_{i \cup j}$ resources is straight-forward (shared between users without any reuse), dynamic FFR now has to be applied on $B - b_{i \cup j}$ carriers for a larger number of cells ($n_i + n_j$). The latter, being a non-trivial RRM process, could become computationally intensive as the size of the cluster increases. Hence, for large clusters, *FluidNet* can run its FFR solution separately in each cluster's constituent sectors (for scalability), albeit on the same set of $B - b_{i \cup j}$ carriers.

4.4 Handling User Mobility

So far we had assumed that the offered traffic load in a sector or cluster can be scheduled on any of the carriers operating on either DAS or FFR. Recall that for mobile (mainly vehicular) users, a DAS configuration is essential not just for reducing compute resource usage but even for performance. Identifying such mobile users can be done in many ways (e.g., mobile operator maintains user's mobility state). Then the offered traffic load from vehicular users can be isolated from the rest of the traffic and scheduled on resources supporting the DAS configuration. Hence, the net traffic load from mobile users in a sector or cluster would place a constraint on the minimum number of carriers that need to be allocated to its DAS configuration. Subject to this constraint, the rest of the operations (resource allocation, multiplexing, clustering, etc.) are performed as mentioned above.

4.5 Handling Interference across Sectors

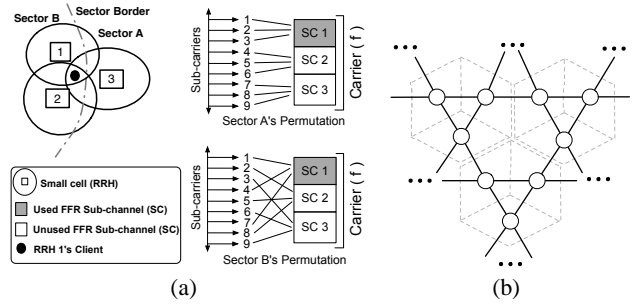


Figure 6: (a) Handling Inter-sector Interference. (b) Sector graph.

Since FFR is executed at the sector granularity for scalability, interference is managed only between cells within a sector. The conventional (simple) solution to handle interference across sectors (or clusters) is to consider all external interference as noise. A more sophisticated approach is to make implicit provisions in the transmission strategy of a sector for alleviating interference across sectors (and hence clusters) without any coordination. Recall that, in a carrier allocated to FFR, only a subset of the sub-channels (called resource blocks in LTE) are used by any of the cells in the sector to account for intra-sector interference (e.g., cells 1 and 2 in Fig. 6(a)). When coordination across sectors is allowed, these sub-channels would be further chosen so as to avoid interference between sectors. However, in the absence of any coordination (for scalability), the sub-carriers constituting the sub-channels in the carrier can be *permuted differently* across sectors. While this does not provide the same performance as performing FFR over the interfering sectors jointly, it does provide an interference averaging (alleviating) effect (cells 1 and 3 in Fig. 6(a)). Note that, this is not possible when operating in DAS, where all sub-channels in the carrier are used in every interfering sector.

FluidNet determines the sector-exterior traffic that is prone to interference from neighboring sectors and operates it in an FFR configuration to alleviate interference. Hence, similar to the minimum set of carriers needed for DAS (for mobile traffic), *FluidNet* reserves a minimum set of carriers for FFR to accommodate sector-exterior traffic.

5. ALGORITHMS IN FLUIDNET

5.1 Overview of Solution

The sequence of operations in *FluidNet* for every epoch (spanning several minutes) is as follows.

Step 1: For every sector, obtain the aggregate traffic demand (over the previous epoch) from each of its small cells. Determine the minimum set of carriers needed for the DAS and FFR configurations based on traffic demand from mobile and sector-exterior traffic respectively.

Step 2: Determine the optimal multiplexing (in frequency) of DAS and FFR configurations for each sector. This would automatically classify the appropriate traffic that needs to be scheduled on a particular configuration. Based on the resulting allocation of carriers to the configurations, determine the RU metric for the sector.

Step 3: Cluster sectors two at a time based on their RU metric until either their net offered load cannot be supported or the RU of the resulting cluster cannot be improved.

Step 4: For each cell in the cluster, apply the configurations on their allocated carriers as determined by the cluster's RU metric

and assign respective traffic to carriers allocated to their appropriate configurations.

We now describe each of the steps in detail.

5.2 Estimation of Radio Resource Demand

Each small cell maintains an estimate of the aggregate traffic demand from its users in the current epoch (of length T s). Given a traffic demand ($d_{c,u}$ in bits) from a user u in cell c , this is translated to the corresponding radio resource demand per sub-frame (i.e. OFDMA resource slots/ms). For this, the average MCS (modulation and coding rate, $\ell_{c,u}$) used to serve the user in the epoch is kept track of and used to obtain the radio resource demand per sub-frame as $r_{c,u} = \frac{d_{c,u}}{T \cdot 1000 \cdot \ell_{c,u}}$ slots. Each cell (c) classifies its net user traffic demand d_c as either mobile or non-mobile. The non-mobile category is further classified as cell-exterior or cell-interior traffic (for FFR purposes) based on presence or absence of interference respectively from neighboring small cells. At the end of the epoch, every cell (c) then provides 3 parameters as input to the central controller: aggregate radio resource demand from mobile ($d_{c,mob} = \sum_{i \in \text{mob}} r_{c,i}$), cell-interior ($d_{c,int} = \sum_{i \in \text{int}} r_{c,i}$) and cell-exterior ($d_{c,ext} = \sum_{i \in \text{ext}} r_{c,i}$) traffic. Note that with centralized processing in C-RAN, there is no associated feedback overhead in providing this information.

Each sector (j) then further aggregates the radio resource demands from mobile traffic in each of its small cells ($D_{j,mob} = \sum_{c \in \text{cell } j} d_{c,mob}$). The minimum radio resource demand needed for its DAS configuration is then the smallest number of carriers needed to satisfy the net mobile traffic demand, i.e. $b_{DAS} = \min_{b \cdot M \geq D_{j,mob}} b$, where M is the number of OFDMA resource slots on each carrier. Similarly, to determine the minimum radio resource demand for FFR, it aggregates the cell-exterior traffic from all its small cells that are on the edge of the sector ($D_{j,ext} = \sum_{c \in \text{edge}(j)} d_{c,ext}$), scales them by $\alpha = 0.25$, and obtains $b_{FFR} = \min_{b \cdot M \geq \alpha D_{j,ext}} b$. Note that approximately only half of the cell-exterior traffic of the sector-edge cells will be vulnerable to other small cells from neighboring sectors. Further, every alternate sector-edge cell would be able to reuse the radio resources in the sector. Accounting for both these aspects, reduces the radio resource demand approximately by a quarter that is captured by α .

Remarks: (1) Since *aggregate* traffic demand from a sector of small cells changes slowly at coarse time scales, determining configurations for every epoch (order of minutes) based on the aggregate demand is both appropriate and robust. For the same reason, it also suffices to estimate the *approximate* radio resource demand from sector exterior traffic. (2) *FluidNet* requires only one parameter (mobile traffic demand) from each small cell in addition to those already required by FFR schemes (i.e. cell-interior and cell-exterior traffic demands). However, one can eliminate the former and simplify *FluidNet*'s design by not catering to mobile and sector-exterior traffic separately (i.e. $b_{DAS} = b_{FFR} = 0$).

5.3 Optimal Sector Configuration

With the estimates of aggregate radio resource demands, *FluidNet* determines the optimal split of carriers between DAS and FFR configurations in a sector (j) as follows. With b_{DAS} and b_{FFR} serving as the minimum number of carriers needed for the DAS and FFR configurations, *FluidNet* uses an iterative approach (Algorithm 1) to determine the optimal split ($b_j, B - b_j$) by starting with b_{FFR} as the minimum set of carriers needed for FFR and allowing it to expand till the radio resource demand can be satisfied or if the limit of $B - b_{DAS}$ carriers is reached. Since mobile and sector-exterior traffic demands are already accounted for, to check if net radio resource demand can be met, *FluidNet* essentially needs to

check only if the remaining resource demand ($\sum_{c \notin \text{edge}(j)} d_{c,ext} + \sum_{c \in \text{cell } j} d_{c,int}$) can be accommodated by the current split (say $b, B - b$) in the iteration, with $b - b_{DAS}$ and $B - b - b_{FFR}$ carriers in DAS and FFR configurations respectively. Note that this would involve running an FFR scheme on $B - b - b_{FFR}$ carriers first (step 3), wherein to maximize the amount of traffic demand satisfied through FFR, the cell-interior traffic ($\sum_{c \in \text{cell } j} d_{c,int}$) that provides maximum spatial reuse is assigned to FFR prior to the cell-exterior traffic ($\sum_{c \notin \text{edge}(j)} d_{c,ext}$). The remaining traffic demand ($D - D_{FFR}$) is then scheduled through DAS on the $b - b_{DAS}$ carriers (step 4).

If the total number of carriers B is small, then a simple, sequential iteration (with increments of one carrier) would suffice. However, if B is large, then the FFR operation in each iteration could be computationally expensive. In this case, *FluidNet* employs binary search, where the split is moved to the left if the traffic demand cannot be met (steps 8-9), and moved to the right if spare resource slots (f_{DAS} , normalized to total # resource slots in a carrier M) remain in DAS configuration after demand satisfaction (steps 5-6). It converges at the split (say b_j), where the number of carriers allocated to FFR cannot be further reduced, while still satisfying the demand. Binary search reduces the number of iterations and hence FFR operations from linear ($O(B)$) to logarithmic ($O(\log_2(B))$). After convergence, the RU of the sector is computed using Eqn. 2 as $RU(b_j, n_j)$.

Algorithm 1 Optimal Configuration for Sector j

- 1: Initialize $b_{low} = b_{DAS}$, $b_{high} = B - b_{FFR}$, $b = b_{high}$,
 $D = \sum_{c \notin \text{edge}(j)} d_{c,ext} + \sum_{c \in \text{cell } j} d_{c,int}$
 - 2: **while** $b_{high} \neq b_{low}$ **do**
 - 3: (f_{FFR}, D_{FFR}) = Schedule_FFR($B - b_{FFR} - b, D$)
 - 4: (f_{DAS}, D_{DAS}) = Schedule_DAS($b - b_{DAS}, D - D_{FFR}$)
 - 5: **if** $f_{DAS} > 0$ **then**
 - 6: $b_{low} \leftarrow b$; $b \leftarrow \frac{b + b_{high}}{2}$; $b_{cur} \leftarrow b$
 - 7: **else**
 - 8: **if** $D - D_{FFR} - D_{DAS} > 0$ **then**
 - 9: $b_{high} \leftarrow b$; $b \leftarrow \frac{b + b_{low}}{2}$
 - 10: **end if**
 - 11: **end if**
 - 12: **end while**
 - 13: $b_j \leftarrow b_{cur}$
-

In addition to RU, every sector keeps track of two metrics: spare radio resources (β_j) and reuse factor (r_j) in the sector (for use in clustering). Note that since minimum set of carriers are determined for FFR configuration, spare resource slots, if any, will appear only in the DAS configuration. This is normalized to the total number of slots (M) in each carrier to yield β_j . Similarly, reuse factor determines the number of actual resource slots needed to support the traffic demand in the sector (and captures the average reuse resulting from FFR): $r_j = \frac{\sum_{c \in \text{cell } j} d_{c,mob} + d_{c,ext} + d_{c,int}}{(B - \beta_j)M}$.

THEOREM 5.1. *FluidNet's iterative scheme converges to the optimal split of carriers between FFR and DAS configurations in each sector w.r.t. the objective in Eqn. 1.*

The proof is deferred to [20] in the interest of space.

5.4 Properties of RU Metric

We present properties of the RU metric that are relevant for clustering (proofs are deferred to [20]). For ease of exposition, we do not consider mobile traffic in the discussions.

PROPERTY 5.1. When two sectors i, j are clustered, the split of carriers in the resulting cluster has to be the minimum of those in the constituent sectors ($b_{i \cup j} = \min\{b_i, b_j\}$) to maximize RU.

PROPERTY 5.2. RU metric does not satisfy the “local” property, i.e. if clustering sectors i, j, k improves the RU, then this does not mean that clustering a subset of its constituent sectors also improves RU.

$$RU(b_{i \cup j \cup k}, n_i + n_j + n_k) \leq \sum_{\ell=\{i,j,k\}} RU(b_\ell, n_\ell)$$

$$\Rightarrow RU(b_{i \cup j}, n_i + n_j) \leq RU(b_i, n_i) + RU(b_j, n_j)$$

PROPERTY 5.3. To cluster sectors i and j (with say $b_i \leq b_j$), we need all of the following to be satisfied.

1. Both sectors must have spare radio resources in the DAS configuration, i.e. $\beta_i < b_i$ and $\beta_j < b_j$.
2. The aggregate traffic from the DAS and FFR configurations of the two sectors must be satisfied by the new split of carriers in the cluster. Equivalently, $b_j - r_j(b_j - b_i) \leq \beta_i + \beta_j$.
3. The RU of the resulting cluster must be improved. Equivalently, $b_j \leq \frac{n_j}{n_j - 1} b_i$.

5.5 Clustering of Sectors

Based on the above established properties, *FluidNet* designs a light-weight clustering algorithm (Algorithm 2) to improve the RU of configurations applied in the network. Representing as a graph $G = (V, E)$, each sector forms a vertex in the graph, while an edge $e = (u, v)$ exists between two vertices (u and v) if the corresponding sectors are adjacent (Step 1). Each edge e carries a weight (w_e), which evaluates property 5.3 in identifying if the corresponding sectors u and v can be clustered, and if so assigns the resulting RU of the cluster as its weight ($w_e = RU(\min\{b_u, b_v\}, n_u + n_v)$). If however, clustering is not feasible, then this is denoted by $w_e = \infty$ (Step 2).

Algorithm 2 Clustering of Sectors

- 1: Construct Sector Graph: $G = (V, E)$, $V = \{\text{sectors}\}$, $E = \{e = (u, v) : v = N(u)\}$
 - 2: $w_e = RU(\min\{b_u, b_v\}, n_u + n_v)$ if Property 5.3 is satisfied; and $w_e = \infty$ otherwise
 - 3: Let $G' = (V', E')$; initialize $V' = V$, $E' = E$
 - 4: **while** (1) **do**
 - 5: Pick $u = \text{Rand}(V')$
 - 6: Select $v^* = \arg \min_{v:e=(u,v) \in E'} w_e$
 - 7: **if** $v^* \neq \emptyset$ **then**
 - 8: Contract (u, v^*) in V' , i.e. $(u, v^*) \rightarrow u'$
 - 9: Add edges in E' , $(u', v) : (u, v) \in E$ or $(v^*, v) \in E$
 - 10: Update edge weights in E' $w_{e'}$, $\forall e' = (u', v) : v \in N(u') \ \& \ v \in V'$
 - 11: **else**
 - 12: Exit
 - 13: **end if**
 - 14: **end while**
 - 15: Output clustered graph $G' = (V', E')$
-

With the above weighted graph, *FluidNet* clusters sectors through a graph coarsening approach. At each step, it picks a random vertex u (Step 5), then selects the neighboring vertex v (Step 6) that when clustered together minimizes the resulting RU ($v = \arg \min_{e \in E'} w_e$,

where $e = (u, v)$). It then contracts u and v , along with edges between them to a new clustered node u' (Steps 7-9). Weights of edges incident on u and v are updated after the contraction (Step 10). The process is repeated until no more clustering is possible. Each vertex in the final graph ($v \in V'$) represents the clustering of sectors in the network for improved RU (Step 15). Further, the RU of each clustered node, represents the common split of carriers between the DAS and FFR configurations for all sectors in that cluster.

Recall that RU does not satisfy the local property (property 5.2). Hence, while local clustering schemes are light-weight and scalable, they might miss out on potential clusters that improve the RU. To reduce the impact of such sub-optimality, *FluidNet* leverages the structure of the sector graph as follows. The logical 3-sector operation of macrocell networks results in a graph that has only cliques of size 3 and cycles of size 6 (see Fig.6(b)). This special form of G is called a “sector graph”. Hence, *FluidNet* includes the following optimization, where in addition to computing the weight of each edge, it also computes the weight of each clique ($w_{(u,v,w)} = RU(\min\{b_u, b_v, b_w\}, n_u + n_v + n_w)$). Hence, it first starts contracting (clustering) all possible cliques ($\frac{|V|}{3}$ in number) before moving to the contraction of edges. This would help improve RU from potential 3-sector (clique) clusters, which would not otherwise result from their constituent 2-sector (edge) clusters.

As with most clustering problems, it can be shown that the problem of finding the network-wide configuration with the smallest RU is NP-hard. We have the following performance guarantee (proof deferred to [20]) for *FluidNet*.

THEOREM 5.2. *FluidNet’s algorithms yield network-wide transmission configurations with a RU that is within a factor of $\frac{3}{2}$ and 2 from the optimal for sector and general graphs respectively.*

5.6 Scalable Realization

While carriers assigned to DAS and FFR (say $(b', B - b')$) in a cluster are fixed for an epoch and determined by the cluster’s resulting RU (computed based on aggregate radio resource demands from previous epoch), DAS and FFR strategies are applied to appropriate incoming traffic demand at finer time scales (order of seconds) during the epoch. Further, the DAS traffic of all the constituent sectors simply share the radio resources through a common DAS configuration on b' carriers. However, the FFR for the constituent sectors is executed individually within each sector (and not jointly), albeit on the same set of $B - b'$ carriers. This keeps the complexity of running FFR schemes low (restricted to cells in a sector). Not running FFR jointly across all sectors in the cluster will result in inter-sector interference. However, this does not hurt the estimated RU of the cluster since it is implicitly incorporated in the RU of the constituent sectors prior to clustering. Further, adopting a two-step approach - first determining the RU-optimal DAS-FFR configuration in each sector, then improving RU of the network by clustering sectors through a light-weight process, forms the key in ensuring scalability of operations in *FluidNet*.

6. PROTOTYPE OF FLUIDNET

6.1 Architecture

The core intelligence of *FluidNet* resides in the central processing entity managing the BBU pool, which consists of two key components.

1. **Resource Manager:** The resource manager is responsible for two key functionalities: (i) determining the appropriate

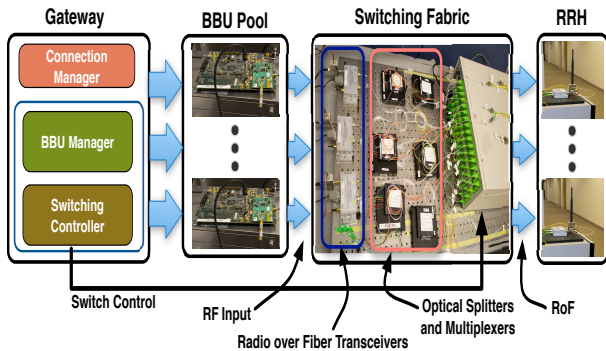


Figure 7: Testbed components of *FluidNet*.

number of BBU units (using *FluidNet*'s algorithms) needed to generate distinct frames and how these frames from BBUs are mapped to specific RRHs, and (ii) assigning compute resources (DSPs, cores, etc.) to each BBU unit. *FluidNet* focuses on the former functionality and is complementary to the processor scheduling problem addressed by studies with the latter functionality [6].

2. **Switching Element:** While the resource manager determines the logical mapping of BBU signals to RRHs, the switching element is responsible for realizing these mappings. Since some BBU frames are sent to multiple RRHs (as in DAS), while other frames are sent individually to specific cells (as in dynamic FFR), the switching element allows for both unicast and multicast switching. Based on the configuration determined by the resource manager on a given carrier, the switch module activates the appropriate set of output ports for an incoming BBU signal depending on the intended set of recipient RRHs. Since a BBU pool may potentially serve tens to hundreds of small cell RRHs, to ensure scalability, the switching fabric may be composed of multiple smaller-size switches (as opposed to one big switch). The size of the switches may be chosen to tradeoff the level of multicasting capability (e.g., for DAS) with cost.

6.2 Implementation

We have built a full-fledged, small-scale C-RAN testbed, capable of over-the-air transmissions. Given that LTE requires licensed spectrum, our set-up is currently based on WiMAX (with an experimental license). However, both LTE and WiMAX being OFDMA-based, our testbed suffices to demonstrate the proposed concepts in *FluidNet* that are equally applicable to LTE as well. Our testbed is depicted in Fig. 7.

BBU Pool, clients and gateway: Since our focus is on the front-haul configuration, we consider six WiMAX BSs (from PicoChip [18]) directly as our BBUs. We use netbooks with USB WiMAX dongles as the clients. *FluidNet*'s algorithms to determine configurations, are implemented in the WiMAX gateway, whose primary role is to manage the traffic flows from/to the clients. In our set-up, a single gateway is instrumented to manage all the 6 BBUs and their clients. The gateway also hosts the controller to instruct the switch for mapping of BBU signals to RRHs. We implement the controller using LabVIEW and communicate the desired configurations to the switch via serial port (RS232).

Radio-over-Fiber: Ideally, baseband signals should be transported in the digital domain between BBU pool and RRHs to allow for scalable, low-latency switching between configurations. How-

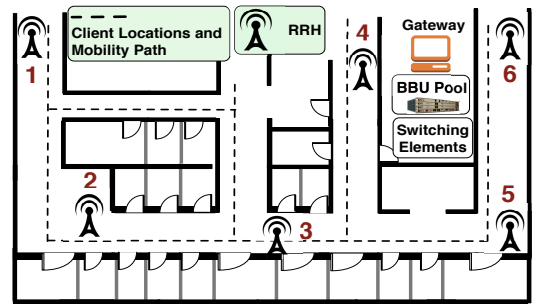


Figure 8: Testbed deployment

ever, the lack of commercially available products to manipulate the baseband signals between BBU pool and RRHs in the digital domain (over CPRI), has prompted us to pick an alternate design, wherein we employ analog RF signal transmission based on radio over fiber (RoF) techniques. With latencies of about $5 \mu\text{s}/\text{Km}$ over the fiber, we have verified that RoF can retain the signal synchronization between RRHs as well as the timing constraint between downlink and uplink signals for reasonable distances of around 10 Km between the BBU pool and RRHs. With RoF, the modulated RF analog signal from a BBU is converted into an optical carrier using a COTS optical transceiver, and delivered to RRHs on a single mode optical fiber.

RRHs: Since all the signal processing (even modulation and RF up/down-conversion) is done at the BBU pool, our RRH design is simple and consists of an optical transceiver attached to an antenna. The optical wavelengths (carrying multiple RF signals) are photo-detected and converted back to the RF domain (for over-the-air transmission) by the optical transceiver. On the uplink (from RRHs to BBUs), the operations are similar but in reverse order.

Switching Element: Since BBU signals are carried as analog RoF, to realize various configurations, we enable switching in the optical domain, which is controlled from the gateway. Since our optical switch supports only one-to-one switching, we enable flexible switching (one-to-one and one-to-many) indirectly by using optical splitters and multiplexers with CWDM. While the latency in switching between configurations is negligible if implemented in the digital domain, it could be appreciable in the optical domain depending on the sophistication of the switch. With our inexpensive optical switch that reconfigures individual port switches, this could amount to 1 s. This is still acceptable if hybrid (DAS, FFR) configurations are realized in the frequency domain (across spectral carriers), where they need to be changed only with appreciable load changes at the granularity of several seconds or minutes.

The gateway controls the optical switch to turn on or off each independent path from each BBU to any RRH to create various configurations. Since each switch in our testbed is limited to supporting all configurations in a set-up with at most 4 BBUs and 4 RRHs, we employ two such switches jointly to serve our 6 BBU-RRH system.

7. PERFORMANCE EVALUATION

7.1 Prototype Evaluation

7.1.1 Set-up

Testbed: Our testbed consists of six small cell RRHs deployed in an indoor office setting, driven by six physical BBUs co-located in a single room through optical fiber (see Fig. 8). There are six

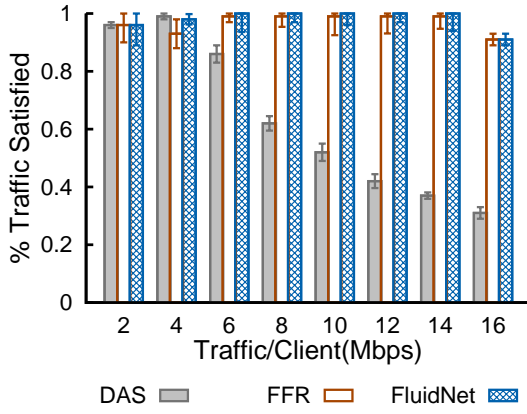


Figure 9: Traffic satisfaction with variable traffic demand.

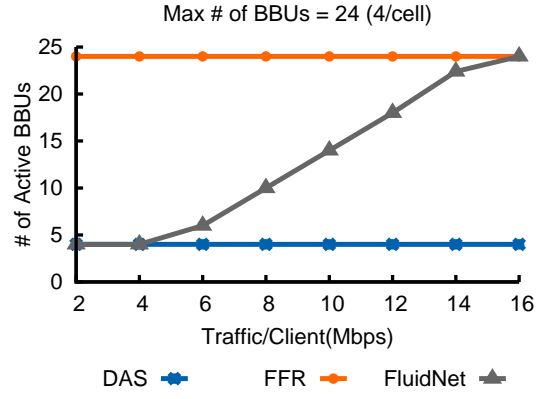


Figure 10: Energy efficiency with variable traffic demand.

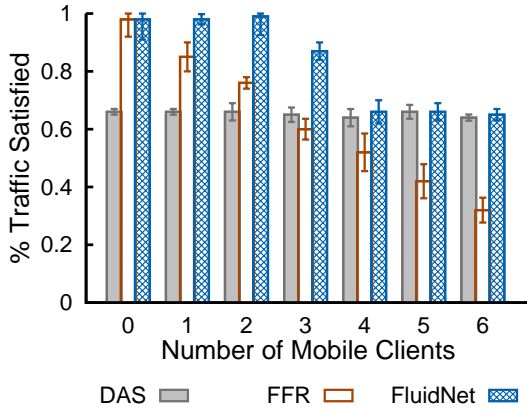


Figure 11: Mobile: Traffic satisfaction.

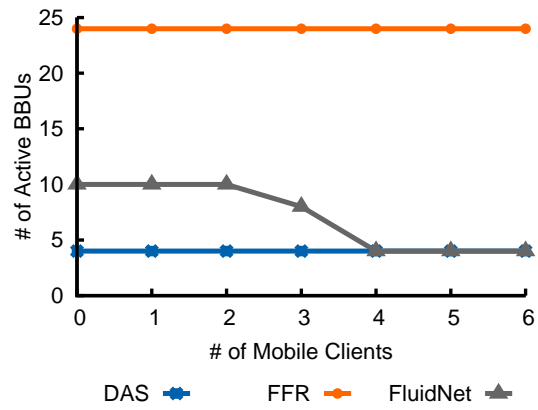


Figure 12: Mobile: Energy efficiency.

clients, each associated to one of the six cells. All the small cells are assumed to be in a single sector of a macrocell. The BBUs can generate WiMAX RF signals over two 10 MHz bands: at 2.59 GHz and 2.61 GHz, for which an experimental FCC license has been acquired to conduct over-the-air transmissions. Hence, we consider four spectral blocks (i.e., carriers), each with 5 MHz bandwidth to realize hybrid configurations. Since our BBUs are BSs themselves, we can operate a BS and hence an RRH on only one carrier at any given time. Due to this technical difficulty, we run DAS and FFR configurations sequentially on the appropriate blocks to realize the hybrid configuration for the sector. This would equivalently amount to 4 logical BBUs (one per carrier) per small cell and hence a maximum of 24 logical BBUs in the system.

Strategies and Metrics: We evaluate *FluidNet* against both the DAS scheme (labeled “DAS”) and an FFR scheme (labeled “FFR”) for baseline comparison (we consider other baselines in simulations). For FFR, our topology allows each small cell to operate on half the set of sub-channels, while being orthogonal to those of its neighbors. In DAS, a single BBU frame serves all the RRHs and clients. Traffic loads (2 - 16 Mbps) and profile (static, mobile) of clients are the parameters studied. The maximum net throughput that can be delivered in a WiMAX frame (at 64 QAM) in our set-up is around 16 Mbps for 10 MHz bandwidth. Each experiment takes 180 seconds and is repeated multiple times with varying client locations. Impact of rate adaptation is isolated by picking the MCS that delivers maximum throughput for a client (we try all MCSs).

The fraction of the offered load supported and the effective number of BBU units consumed in the process are the metrics of evaluation.

7.1.2 Impact of Traffic Heterogeneity

With six static clients, we study the percentage of average traffic satisfied and the number of BBUs required by each scheme with varying per-client traffic demand in Figs. 7 and 7, respectively. With high load, FFR is essential to support the traffic demand, while DAS can support only a third of the demand (Fig. 7). When the load is low, DAS is sufficient and activates only a sixth of the BBUs required by FFR (Fig. 7). While *FluidNet* blends the best of DAS and FFR under extreme load conditions, its benefits are more pronounced in the intermediate regime (e.g., 10 Mbps demand per-client), where it outperforms both DAS and FFR. By employing hybrid configurations and adapting them to traffic profiles, *FluidNet* sustains twice as much traffic as DAS and requires only half the BBUs activated by FFR.

7.1.3 Impact of User Heterogeneity

We vary the number of mobile clients in a six client scenario, with each client’s traffic fixed at 8 Mbps. To eliminate the adverse impact of handoffs in FFR (triggers, delays, etc.), we move a mobile client at pedestrian speed only in the vicinity of its RRH (sample path in Fig. 8). In contrast, seamless coverage and lack of handovers, allow a client to be moved in all deployment areas with DAS and *FluidNet*. Hence, the results in Figs. 7 and 7 are optimistic for FFR. We see that with increasing fraction of mo-

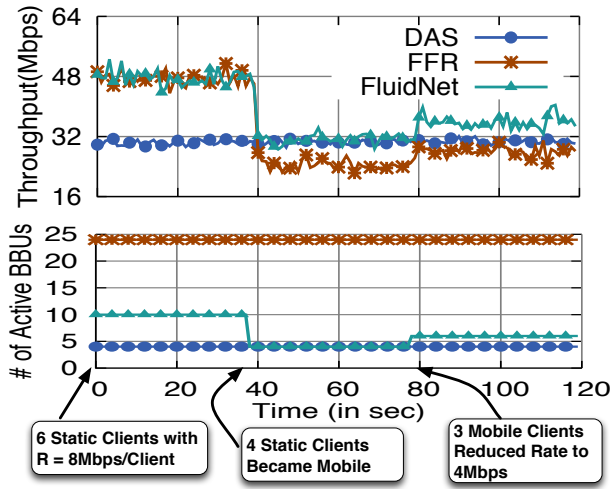


Figure 13: Network Dynamics.

mobile traffic, FFR’s performance degrades and ends up being much worse than that of DAS (Fig. 7). We observed that, even without handovers, when a client moves away from its RRH, its link deteriorates and faces high interference from the control region of frames of neighboring RRHs in FFR (only data part of the frame is protected in FFR). While DAS’s coverage provides consistent link quality, it under-utilizes the spectrum when mobile traffic is low. *FluidNet* strikes a fine balance between the two configurations to support as much as 50% more traffic, while incurring a BBU energy consumption that is only slightly more than that of DAS.

7.1.4 Adaptation to Network Dynamics

We now evaluate *FluidNet*’s adaptability to network dynamics. We start with six static clients, each with a 8 Mbps traffic load. Two events are triggered, one at 40 seconds into the experiment and another at 80 seconds. In the first event, four of the clients become mobile. Then at the 80 second mark, one of the mobile clients becomes static again and the remaining mobile clients reduce their rate to 4 Mbps. From Fig. 13, we see that *FluidNet* tracks FFR performance initially (albeit at less number of BBUs activated), when there are more static clients inducing a high traffic load. When a majority of the traffic demand becomes mobile at the first event, unlike FFR that suffers in performance, *FluidNet* immediately (but for a short transition delay) adapts its configuration to track DAS performance that is optimal for the updated network conditions. Similarly, when the traffic load of static clients starts to dominate, while still involving mobile clients at the second event, *FluidNet* employs a hybrid configuration to sustain a higher traffic load compared to both DAS and FFR, while incurring a BBU usage comparable to DAS. This clearly indicates *FluidNet*’s ability to effectively adapt its configurations to varying network conditions.

7.1.5 Multi Operator/Technology Customization

One of *FluidNet*’s key features is its ability to allow for multiple operators to customize the configurations needed to serve their respective clients simultaneously. To illustrate this, we design an experiment with three BBUs and three RRHs. There are two operators, one operating at 2.59 GHz and the other at 2.61 GHz, each with 10 MHz bandwidth. Both operators share the same set of three RRHs to cater to three clients each simultaneously. While all clients for operator 1 are static and impose a net rate requirement of 21 Mbps, those for operator 2 are all mobile with a net

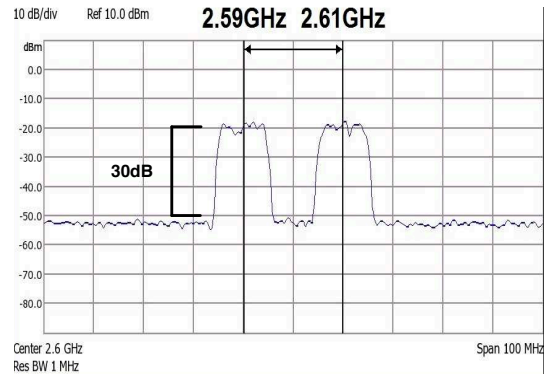


Figure 14: Two Operators: Signal spectrum.

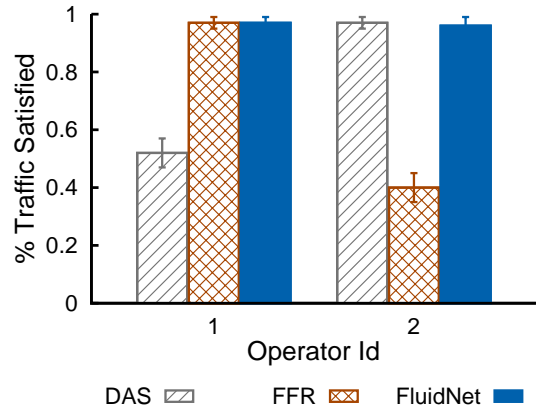


Figure 15: Two Operators: Traffic satisfaction.

rate requirement of 12 Mbps. To check transmission feasibility on our front-haul over longer distances, the fiber between BBUs and RRHs is made to be a 10 Km fiber spindle. Fig. 7.1.4 presents the spectral graph from one of the RRHs captured using a spectrum analyzer. It is clearly seen that both the operators are able to co-exist simultaneously on the same front-haul without any interference to each other’s RF signal. Furthermore, this is achieved over a large distance of 10 Km, which demonstrates feasibility for an outdoor cellular deployment. Also, Fig. 7.1.4 shows that *FluidNet* tailors the right configuration for each operator to provide maximum satisfaction of traffic demand.

This is also evident from Figs. 7.1.4 and 7.1.4, where a single operator uses two different access technologies (WiFi and WiMAX) to serve five clients (each with 10 Mbps traffic rate) through 3 RRHs. Two of the clients on WiFi (2.43 GHz) are static and associated to two of the RRHs, while the other three are on WiMAX (2.59 GHz) and mobile. It is interesting to see that *FluidNet* is capable of simultaneously supporting an asynchronous (WiFi; one-to-one for CSMA) and synchronous (WiMAX; one-to-many for DAS) access technology for the same operator. *FluidNet*’s support for multiple operators and technologies are very useful features in a C-RAN, given the growing popularity of RAN-sharing and dual carrier small cells (for WiFi offload).

7.2 Simulation

Set-up: We use a 3GPP-calibrated system simulator to create an outdoor heterogeneous cellular network, with 19 macrocell sites (each has three sectors) and ten small cells per sector. Thus, the

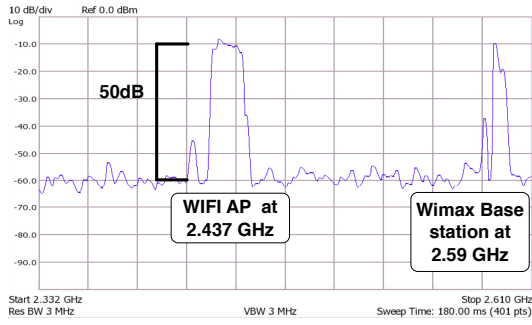


Figure 16: Signal Spectrum: WiFi + WiMAX.

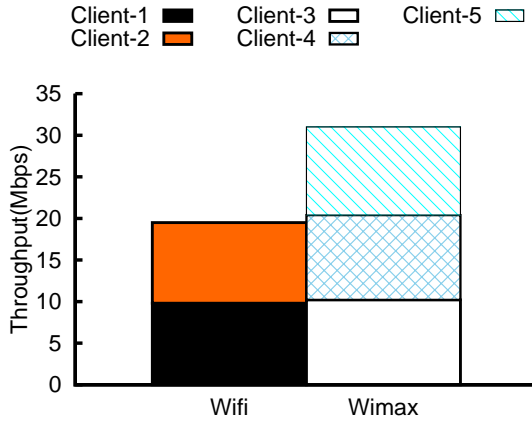


Figure 17: Traffic Satisfaction: WiFi + WiMAX.

network has a total of 627 cells (57 macro + 570 small) based on the scenarios defined in 3GPP 36.814 [2]. We distribute 3600 small cell clients according to the ‘4b’ distribution [2]. We assume that the macrocells and their clients use pre-determined spectral resources orthogonal to the ones used by the small cells and their clients, and thus ignore the interference from/to the macrocell network.

To generate traffic demands, we resort to emulating a typical operational day in outdoor cellular networks. Since we do not have access to such operator data (and public data does not exist to the best of our knowledge), we use the reported peak hour distribution from [7] as follows. We mark each sector (and the small cells in it) as either ‘business’ or ‘residential’. As seen in Fig. 2, we geographically determine that the central, shaded sectors are business sectors (there are a total of 21 such sectors with 210 small cells in them) and peripheral sectors are residential sectors (36 of them exist). The small cells in a business sector hit their peak loads between 10 a.m. and 4 p.m. and residential cells have peak hours between 4 p.m. and 8 p.m. The traffic outside the peak hours is chosen such that there is a gradual increase until the peak interval and a decrease after that.

We compare *FluidNet* against three other schemes. The first (labeled ‘FFR’) is a pure FFR solution running with a fixed cluster size corresponding to a macrocell (3 sectors = 30 small cells). The second (labeled ‘DAS’) is a pure DAS solution with opportunistic clustering. When the total load of neighboring sectors is less than a frame’s worth of resources (i.e., the max. capacity of DAS), they are merged in a DAS cluster and thus served by one BBU. The third (labeled ‘GRID’) is reported in [7] and addresses energy consumption by turning small cells off during non-peak periods.

Traffic Heterogeneity: We first simulate a network where no clients are mobile. Each result is the average of five different runs with randomly selected traffic demands from clients, subject to the spatio-temporal traffic distribution.

Figs. 18(a) and 18(b) plot the traffic satisfaction ratio and the energy consumption (*RU*), respectively. We first see that *FluidNet* has a competitive traffic satisfaction ratio with FFR (is only 3% worse on average). The slight reduction is because FFR explicitly accounts for inter-sector interference by considering a cluster size of three sectors. In contrast, *FluidNet* applies FFR at a granularity of one sector and resorts to resource permutations to address inter-sector interference in a scalable manner. We also see that while having a competitive traffic ratio, *FluidNet* is much more (3x on average) energy efficient than FFR. DAS, albeit the most energy efficient strategy, suffers from lack of spatial reuse and hence satisfies only 65% of the traffic on average.

When compared with GRID, while the fraction of traffic satisfied does not differ considerably, *FluidNet* activates 2.2x less BBUs than GRID. This is due to the fact that while energy savings from BS-switching approaches such as GRID are inherently limited based on physical proximity of cells, *FluidNet* can cluster arbitrarily large numbers of cells to yield more energy savings. This is exemplified in Fig. 18(e) where we plot the temporal progression of clusters in *FluidNet*; clusters (color-coded) are seen to shift spatially from residential areas in the morning to business areas in the evening. White (uncolored) sectors are clusters of size one (i.e., cannot be merged with other sectors due to high traffic load). To closely look at clustering in *FluidNet* during non-peak hours, we compare *FluidNet* with and without the clustering component (the latter called ‘*FluidNet-NC*’). As seen in Fig. 18(c), even without its clustering component *FluidNet* outperforms GRID. Further, while *FluidNet-NC* requires 80 BBUs on average, *FluidNet* requires only 43 BBUs, resulting in much lower energy consumption. This shows that clustering is critical in realizing high energy savings.

In summary, *FluidNet* effectively exploits the spatial and temporal load asymmetry in the network and yields more energy savings than state-of-the-art solutions while satisfying a high fraction of the traffic demand.

User Heterogeneity: We now evaluate *FluidNet* with vehicular mobility. Here, we take the peak traffic hour of the day (4 p.m.) and investigate the traffic satisfaction ratio (averaged over 5 runs) with varying percentage of mobile clients. Each client moves at 60 miles per hour, only within its sector. From Fig. 18(d) we see that DAS performance is not affected by mobility since it results in a uniform signal quality for mobile clients; the network capacity is unchanged. With FFR, performance degrades as we increase the percentage of mobile clients (due to handovers and degraded SNR). With *FluidNet*, increasing number of mobile clients results in more carriers being allocated for DAS. While associating mobile traffic with DAS is beneficial in most of the cases, it can lead to lower performance (compared to FFR) when *all* the traffic is mobile. Ideally, one would need to identify the tradeoff between DAS (uniform per-client SNR but no spatial reuse) and FFR (degraded client SNR but high spatial reuse) for mobile traffic, and make careful decisions.

8. DISCUSSIONS AND REMARKS

We presented *FluidNet* - a framework for dynamically re-configuring the front-haul of a C-RAN to meet the dual objective of improved RAN performance with reduced resource usage in the BBU pool. Our evaluations show promising benefits towards these goals. Going forward, we would like to consider the following.

Applicability to other C-RAN Models: Since *FluidNet* focuses on *logical* front-haul configurations, it can work with any front-

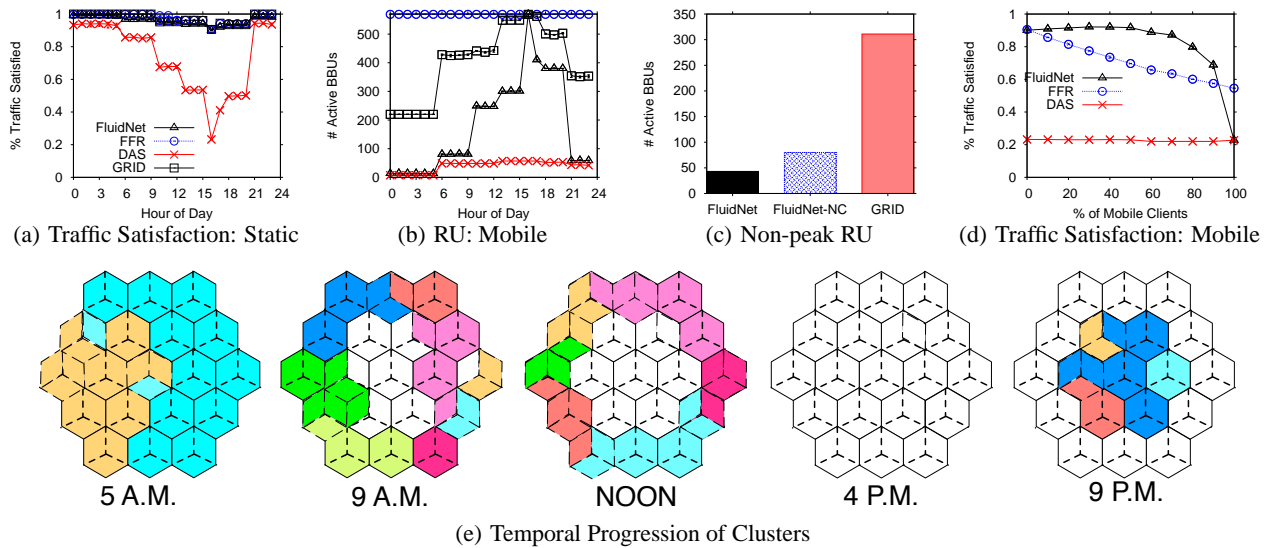


Figure 18: *FluidNet* has comparable traffic satisfaction ratio to FFR (a), and is 3x and 2.2x more energy efficient than FFR and GRID respectively.

haul (e.g., microwave wireless) as long as the latter can support the data rates needed for transport of BBU signals. Similarly, it also applies in a partially-centralized C-RAN model [17], where more processing is entrusted to the RRHs to reduce the load on the front-haul. However, the energy savings in this model needs to be investigated.

Co-existence with Carrier Aggregation: LTE-advanced systems will support multiple component carriers and carrier aggregation. Carrier split for configurations in *FluidNet* can be realized much more easily with multiple component carriers. However, the interaction of FFR and DAS with joint scheduling on multiple carriers needs further study.

Migrating to Digital Front-Haul Transmissions: Instead of using RF over Fiber, we would like to migrate our BBUs to those that provide access to digital I-Q streams that can be transported over CPRI. This would allow for scalable realization of our configurations in the digital domain.

9. REFERENCES

- [1] Common radio public interface. www.cpri.info.
- [2] G. T. 36.814. Evolved Universal Terrestrial Radio Access (E-UTRA): Further advancements for E-UTRA physical layer aspects. <http://bit.ly/Ysh6Cq>.
- [3] J. Andrews, W. Choi, and R. Heath. Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks. In *IEEE Wireless Communications*, volume 14, Dec 2007.
- [4] M. Y. Arslan, J. Yoon, K. Sundaresan, S. V. Krishnamurthy, and S. Banerjee. FERMI: A Femtocell Resource Management System for Interference Mitigation in OFDMA Networks. In *ACM MobiCom*, Sept 2011.
- [5] U. Barth. How to reduce green house emissions from ICT equipment: Wireless Networks, EARTH research project. In *ETSI Green Agenda*, Nov 2009.
- [6] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo. CloudIQ: A Framework for Processing Base Stations in a Data Center. In *ACM MobiCom*, Aug 2012.
- [7] P. Chunyi, L. Suk-Bok, L. Songwu, L. Haiyun, and L. Hewu. Traffic-Driven Power Saving in Operational 3G Cellular Networks. In *ACM MobiCom*, Sept 2011.
- [8] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017, Feb 2013.
- [9] Ericsson. AIR: Antenna Integrated Radio. <http://bit.ly/dLOIAu>.
- [10] T. Flanagan. Creating cloud base stations with TI's KeyStone multicore architecture. <http://bit.ly/ztLLcz>.
- [11] T. D. Forum. www.thedasforum.org/news/.
- [12] Intel. Heterogeneous Network Solution Brief. <http://intel.ly/ZIMRcI>.
- [13] H. Li, J. Hajipour, A. Attar, and V. Leung. Efficient HetNet Implementation Using Broadband Wireless Access With Fiber-Connected Massively Distributed Antennas Architecture. In *IEEE Wireless Communications Magazine*, Jun 2011.
- [14] C. Liu, K. Sundaresan, M. Jiang, S. Rangarajan, and G. Chang. The Case for Re-configurable Backhaul in Cloud-RAN based Small Cell Networks. In *IEEE Infocom*, 2013.
- [15] A. Lucent. lightRadio Network: A New Wireless Experience. <http://bit.ly/VpR4Cb>.
- [16] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo. Optimal energy savings in cellular access networks. In *IEEE GreenCom*, Jun 2009.
- [17] C. Mobile. C-RAN: The Road Towards Green RAN. <http://bit.ly/YalzuW>.
- [18] Picochip. <http://www.picochip.com>.
- [19] M. Sauer, A. Kobayakov, and A. Ng'oma. Radio Over Fiber for Picocellular Network Architectures. In *IEEE Journal on Lightwave Technology*, volume 25, 2007.
- [20] K. Sundaresan, M. Arslan, S. Singh, S. Rangarajan, and S. Krishnamurthy. Fluidnet: A flexible cloud-based radio access network for small cells. In *NEC Laboratories Technical Report*, Mar 2013.
- [21] D. Tipper, A. Rezgui, P. Krishnamurthy, and P. Pacharintanakul. Dimming Cellular Networks. In *IEEE Globecom*, Nov 2010.
- [22] Q. Wang, D. Jiang, J. Jin, G. Liu, Z. Yan, and D. Yang. Application of BBU+RRU Based Comp System to LTE-Advanced. In *IEEE ICC Workshops*, Jun 2009.