Information Raining and Optimal Link-Layer Design for Mobile Hotspots

Daniel Ho and *Shahrokh Valaee*

Department of Electrical and Computer Engineering,
University of Toronto,
10 King's College Road,
Toronto, ON, M5S 3G4

Phone: (416) 946-8032,
Fax: (416) 978-4425
email: valaee@comm.utoronto.ca

December 21, 2004

Abstract

In this paper, we propose a link layer design for mobile hotspots. We design a novel system architecture that enables high-speed Internet access in railway systems. The proposed design uses a number of repeaters placed along the track, and multiple antennas installed on roof of vehicle. Each packet is decomposed into smaller fragments and relayed to the vehicle via adjacent repeaters. We also use erasure coding to add parity fragments to original data. This approach is called *information raining* since fragments are rained upon the vehicle from adjacent repeaters. We investigate two instances of information raining. In *blind* information raining, all repeaters awaken when they sense the presence of the vehicle. The fragments are then blindly transmitted via awakened repeaters. A *vehicle station* installed inside the train is responsible to aggregate enough number of fragments. In the *throughput-optimized* information raining, the vehicle station selects a bipartite matching between repeaters and roof-top antennas, and activates only a subset of the repeaters. It also dictates the amount of transmission power of each activated repeater. Both the bipartite matching and power allocations are individually shown to be NP-complete. Matching heuristics based on Hungarian algorithm and Gale-Shapley algorithm are proposed. A simplex-type algorithm is proposed as the power allocation heuristics.

Index Terms

Emerging technologies, Network architecture and design, Wireless communication, Network protocols, Mobile communication systems, Mobile environments, Medium access control, Mobile hotspots, Graphs and networks, Linear programming, Constrained optimization, Graph theory, Combinatorial algorithms.

I. Introduction

Hotspot technology has become very popular recently. We witness the proliferation of hotspots in hotels, airports, coffee shops, shopping malls, etc. An apparent question is whether it would be possible to extend hotspots to mobile vehicles. This paper proposes a novel architecture to realize mobile hotspots for mass transportation vehicles such as trains, subways, and busses. If the hotspot technology is extended to this truly mobile environment, business travelers may connect to their corporate offices through Virtual Private Networks (VPNs) to check their email, do video-conferencing and finish their work. Leisure travelers may also send instant messages, surf the web, and use multimedia applications for entertainment. Although the focus of this paper is on the application of hotspot technology in trains and subways, similar approaches can be used for busses and automobiles.

Using conventional approaches to extend Internet access to vehicles is challenging. The direct application of access points (AP) of wireless local area networks (WLANs) along the track is highly unscalable. For instance, a vehicle travelling at 72km/h demands handoff every 10 seconds with AP coverage of 100m radius. These handoff rates are infeasible with the current Mobile IP architecture [1], [2]. Overall, the WLAN technology is designed for users with low mobility, thus it is not suitable to be directly used for mobile hotspots in mass transportation systems.

The challenges with cellular systems are similar to WLANs. For instance, there is an issue with cellular planning. Terrain obstacles such as hills, buildings and tunnels may cause shadowing and large delay spread of several microseconds [3] to certain sections of these routes, which impair transmission quality in terms of bit-error-rate (BER) and achievable bandwidth. Therefore, the use of microcells along the transportation route is justified. However, these microcells result in frequent handoffs due to high mobility, and may generate interference to existing macrocells in the vicinity. Very high velocity movements may also induce Doppler effects that are unanticipated by the system. For example, Maglev trains are intended to reach velocities up to 430km/h [4], whereas GSM is designed to handle up to 250km/h at 900MHz. Indeed, the International Telecommunication Union (ITU) International Mobile Telecommunications-2000 (IMT-2000) only expects 3G to provide a data rate of 144 kbps or higher in high mobility traffic, compared with the expected data rate of 2 Mbps or higher in indoor traffic. High Doppler rates are also known to create “floor phenomena” to BER curves, such that an increase in received signal level does not improve transmission quality [5], [6].

For rural environments, Low Earth Orbit (LEO) satellite services may be used to facilitate Internet access to trains and ships. A large constellation of LEO satellites orbit around 500km to 2000km above the Earth, providing global coverage to rural areas. The primary advantage of LEO system is the satellites' proximity to the ground, thus there is less propagation delay (about 10ms), and less transmission power is needed, compared to the traditional Geostationary Earth Orbit (GEO) satellite system. However, the service cost of LEO satellites are prohibitive for ordinary civilian uses. In current LEO satellite services, any bandwidth demand greater than voice-like transmissions require mobile equipment sizes that are not easily portable.

Mass transportation system operators see additional benefits with reliable high-speed access within their vehicles. The service may replace their Private Mobile Radio (PMR) system to provide voice communication among drivers, central operators and maintenance staff. Additional functionalities such as signaling control, scheduling and logistic support may be integrated into such a network to assist their existing infrastructure. Multimedia entertainment features such as movies and TV on demand may also be made available to passengers if broadband access is provisioned. Indeed, in 1993, Commission 7 of International Union of Railways (UIC) decided to adopt GSM as its basis to standardize pan-European railway communication, such that trains in Europe may communicate with rail stations across European countries. The decision sparked an interest to integrate PMR with cellular systems. In response, European Telecommunications Standards Institute (ETSI) elaborated a GSM specification for railway uses named GSM-R. Completed in 2000, GSM-R is not an extension version of GSM, but rather an integral part of the GSM standard. It provides important services including voice broadcast and voice group call, call priority and fast call setup that satisfy most needs of railway operators. With the ever-growing demand for mobile connectivity, it is reasonable to expect integration of PMR and public Internet access in mass transportation systems to flourish.

A. Related Work

The research result on high-speed access for railway systems is very limited. To our best knowledge, the only investigations that relate high-speed access on railway systems are briefly described in [7] and [8], where microcells are positioned at 1.0km to 1.1km intervals along railroad, with one mobile station antenna mounted at each end of the train, capable of short-distance communications over a length of 800 to 900m with the closest base stations. Thus, two separate channels may be established with negligible interference on each other. Our system architecture is more general and takes full advantage of spatial diversity based on the vehicle's size and surrounding environment.

At the time of this writing, there are several industrial efforts to implement mobile hotspot in railways systems [9][10] [11] [12] [13]. All of these solutions rely on existing networking infrastructure such as cellular and satellite systems to provide Internet connectivity. As such, coverage of the hotspot service is bounded by the limitations of the underlying technologies. For instance, coverage ceases at tunnels if there exists no relay equipment at the entrances. In general, these solutions provide a smart “hack” to existing systems, and are only satisfactory on an interim basis.

In the research community, a multihop wireless system is proposed, wherein intermediate mobile terminals may relay information of other terminals when they are neither the initial transmitter nor the final receiver [14][15] [16] [17]. Among other benefits, wireless multihop routing expands existing coverage area with low deployment cost in cellular networks. Integrated with mesh connectivity and load balancing schemes [18], [19], multihop networks may provide an adaptive solution to mobile hotspots in mass transportation systems. Indeed, wireless mesh networks allow every mobile user to act as a co-operative forwarding node and can be used as a backbone network that can transmit data to/from mobile vehicles [20].

However, there are several major shortcomings in the application of multihop wireless system to mobile hotspot. The first weakness is the assumption of cooperation. Mobile users usually turn off their handhelds and laptops when the device is not in use; even if they are turned on, devices have no incentive to use their limited battery power to relay information of other devices. The second weakness is the security issues. Because information is relayed by untrusted parties, security schemes must be applied to provide assured communication. Processing overhead and network overhead associated with security coding, connection management and key management must be discounted in multihop networks. Third, as previously discussed, mobile users in the same vehicle share the same large-scale path loss and shadowing to base stations, rendering multihop with other nodes in the vehicle ineffective. For instance, when a train travels across an underground section, all mobile users in the vehicle lose their connectivity, and thus hopping to their peers is futile. Overall, multihop wireless system can be an inefficient, overcomplicating solution when applied to mobile hotspot scenario in mass transportation systems.

B. Contributions and Paper Organization

In this paper, we propose a novel system architecture to facilitate mobile hotspots that is applicable to both WLAN and cellular systems. This is coherent with the recent development in the convergence of these two technologies. Although some concepts are applicable to other vehicular systems, we shall focus on long-haul railway and metropolitan subway systems. We then consider design issues when implementing our architecture at link layer, followed by system modeling, optimization analysis and simulation results. For the rest of this paper, we shall illustrate our discussion on downlink traffic forwarding due to the emergence of asymmetric data applications in mobile devices, although our architecture allows for traffic flow in both directions.

Our research intent is to investigate link layer design and optimization techniques of our proposed architecture specific to railway systems. Nonetheless, the scope of the problem is vast; this paper subsequently focuses on only the most noteworthy section (in the authors' opinion) of the architecture, which is the wireless environment between *multiple repeaters* and *vehicle antennas* that is described in the next section. Furthermore, we restrict our optimization objective to *throughput maximization*. However, we should mention that the definition of throughput in this paper does not match the conventional definition of throughput, which is the average number of successfully received packets at the destination. Instead, we think of the throughput as the average number of packets transmitted to the vehicle, while noting that, in certain conditions, some of these packets might be repeated and hence redundant.

In the following section, we propose an open architecture that employs repeaters located along the trackside and multiple antennas mounted on the exterior of the vehicle. We then suggest two general methods of transmission between each repeater-antenna pair. First, we introduce and analyze the *blind information raining* method in Section III. In blind information raining, downlink data is equally distributed among multiple repeaters to be blindly transmitted to the air interface with equal power and data rate. Multiple vehicle antennas act as receivers where each of them independently tunes to one of the repeaters to recover information. Section IV presents the interference model used between the set of active repeaters and antennas. Then, in Section V, we propose a throughput-optimized method based on resource allocation. In addition to traditional resource considerations such as power, rate and signal-to-noise ratios, we consider the freedom of *matchings* between repeaters and antennas as another significant resource allocation problem. Unfortunately, the mathematical complexity of the problem is shown to be NP-complete. Several heuristic algorithms are proposed as a consequence in Section VI and Section VII.

Lastly, we compare techniques presented in Section III and Section V through simulation analysis in Section VIII. We investigate how some of the parameters of the modeled railway environment affect performance in terms of system throughput. We also provide system enhancements based on these insights. From these simulations, we are able to obtain interesting insights about our setting, and relate them to other works. Finally, conclusions are presented at the end of the paper.

II. System Architecture

All handheld devices have critical constraints in hardware component size, computation power and battery power. Consequently, they can only process relatively simple procedures, and only one small antenna can be installed in most devices. Conversely, one common feature among trains is their large physical size. More powerful networking equipment can be installed inside vehicles without practical space and battery power limitations. It is also ideal to install multiple, more powerful antennas around vehicles, connecting them to the networking equipment.

Furthermore, unlike generic mobile users, trains have a network of defined paths to travel. By installing repeaters at close vicinity along the network of paths, line-of-sight (LOS) can be guaranteed between repeaters and vehicles that move along the network.

We propose a system architecture for mobile hotspots in railway system, as shown in Figure 1. Similar to backbone networks and mobile switching centers of cellular systems, the mass transportation system communications network is a cloud of networking equipment that is responsible for routing traffic between the Internet and local information distribution centers that we refer to as *zone controllers* (ZC). Zone controllers are responsible for traffic dissemination within their local region, such as a railway section of several kilometers. They are also responsible for detecting the presence of vehicles and their mobile users. The ZC need not be a standalone equipment, but a functional procedure inside the “transportation communications network”, and thus the implementation may leverage resources of existing network. Stationary repeaters are positioned along the responsible path. Repeaters and ZC may be connected via fiber cables, or via daisy chaining of wireless links with intermediate repeaters. These repeaters then relay

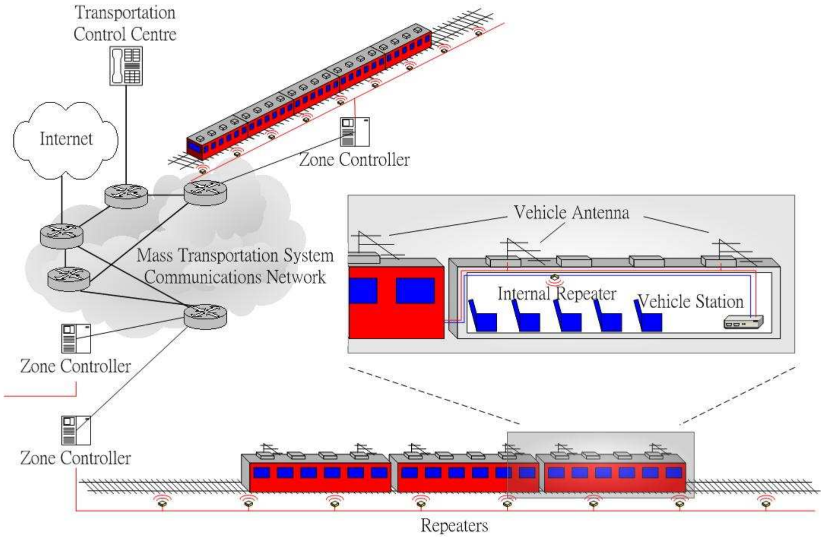

Fig. 1. Proposed architecture for mobile high-speed access in railway system

the traffic of ZC to multiple antennas that are installed on top of moving vehicles. Inside each vehicle locates a *vehicle station* (VS) that gathers traffic from vehicle antennas, and relays them to access points. Thus, passengers enjoy seamless mobile hotspot service with no adjustment at the mobile terminal.

We categorize system implementation alternatives of our paradigm into three approaches, distinguished by different networking layers. At the physical layer viewpoint, the described paradigm can be modelled as a MIMO channel, where channel gain and coding gain may be achieved with space-time codes [21], [22]. At the network layer viewpoint, the described paradigm can be modelled as a multipath network, where benefits in throughput and fault-tolerance may be achieved with multipath routing [18], [23]. These two models are well understood in general; as an alternative, in this paper we are interested in designing our architecture at link layer.

III. Blind Information Raining

In link layer design, the ZC receives downlink data from system network and disseminates data to multiple repeaters at the vicinity of the vehicle. The overall system diagram for downlink traffic is shown in Figure 2. Each repeater forms a one-to-one wireless channel to each vehicle antenna and transmits such data via the channel. The VS at the vehicle receives the packets and forwards them accordingly. (For brevity, we shall refer to vehicle antennas simply as "antennas" henceforth.) Thus, handoff of wireless links within the zone is implicitly handled

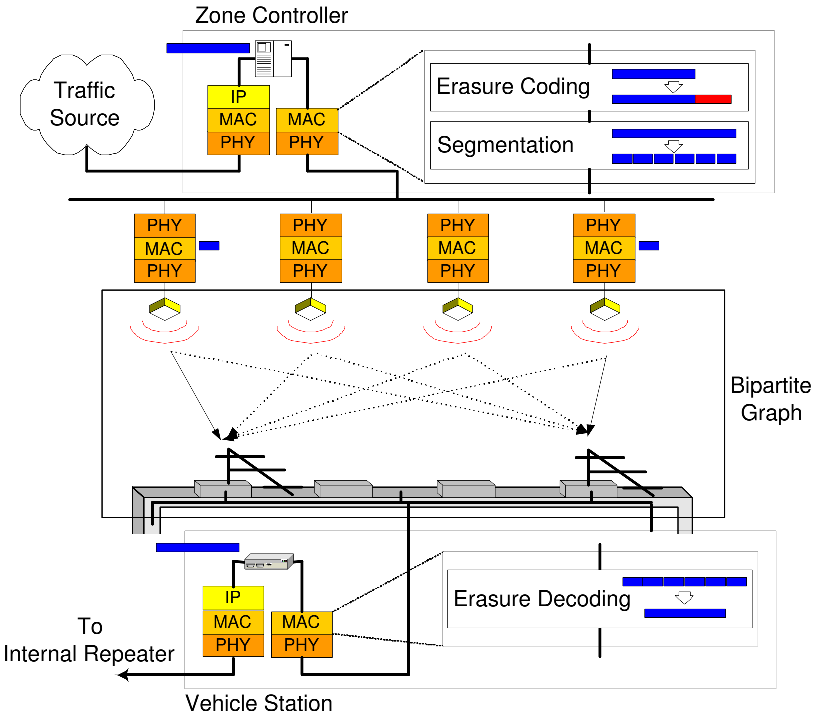

Fig. 2. System block diagram for downlink traffic with information raining

by link acquirement renewals.

Similar to multipath routing techniques, it is possible to improve system reliability with erasure coding on packets [24], [25], [19]. Instead of error detection and correction, erasure codes are added to provide fault tolerance if a segment of data is lost, or “erased”, during transmission. By segmenting the encoded data into segments of equal length, these segments can be disseminated to the repeaters, which act as dumb terminals that repeat segments to the air interface. Since each wireless link may temporarily lose link due to fading and interference, not every segment may be received by the corresponding antenna. However, the decoder at VS can reconstruct the original data if a certain number of unique segments arrive successfully, regardless of the specific subset of segment arrivals.[1] Effectively, erasure coding enhances robustness to the inherently unreliable wireless channels. From a network layer standpoint, the traffic flow between ZC and the vehicle is viewed as a single transparent link. We metaphorically describe this approach as *information raining*, where segments of information are rained upon the vehicle by repeaters, and antennas resemble buckets that retrieve as much information as possible.

We propose a general MAC layer structure, as shown in Figure 3. One *master antenna* is employed at the vehicle to broadcast a beacon signal to repeaters in the vicinity. Repeaters

[1]Any repeated segment is only used once in the received set of segments.

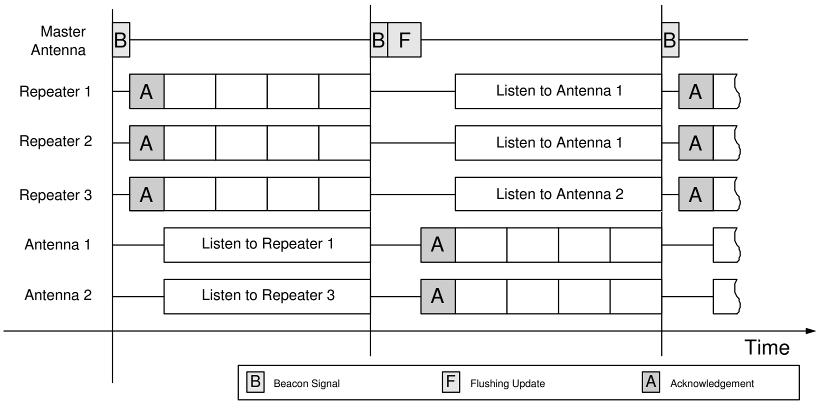

Fig. 3. MAC layer frame format for blind information raining

that detect the presence of this signal "awaken", and broadcast their unique identification signals as acknowledgment. Each antenna detects their existence, and performs subchannel estimation on each awakened repeater. One simple scheme is to allow each antenna to tune independently to the repeater that yields the strongest link gain; we label this scheme as *blind information raining*. All repeaters then broadcast segments during the frame payload, of constant length in time. A similar procedure occurs in uplink frames, but with repeaters and vehicle antennas in reversed-role.

At the end of each frame, some packets are successfully recovered. It would be redundant for repeaters to transmit segments of packets that are successfully recovered in previous frames. Consequently, we recommend a redundant-segment flushing process. The master antenna broadcasts a *packet recovery update record* (Flushing Update in Figure 3) after the beacon signal. Repeaters then discard all segments of the corresponding packets that remain in their buffer. This scheme effectively increases throughput of the system, and is particularly important when the system employs erasure coding with high protection ratio. The redundant-segment flushing update occurs only after a downlink frame, because in uplink, ZC is unlikely able to quickly feedback packet recovery updates to repeaters at the next frame header, and there is no equivalency of the master antenna to broadcast to the vehicle even if such information is available. Regardless of traffic direction, a complimentary segment-timeout mechanism must be implemented at both VS and repeaters to discard lingering segments.

Information raining allows minimal intelligence from the repeaters, and the possibility to build them with relatively cheap, standards-ready components. This is important for deployment cost because numerous repeaters must be installed along the transportation system.

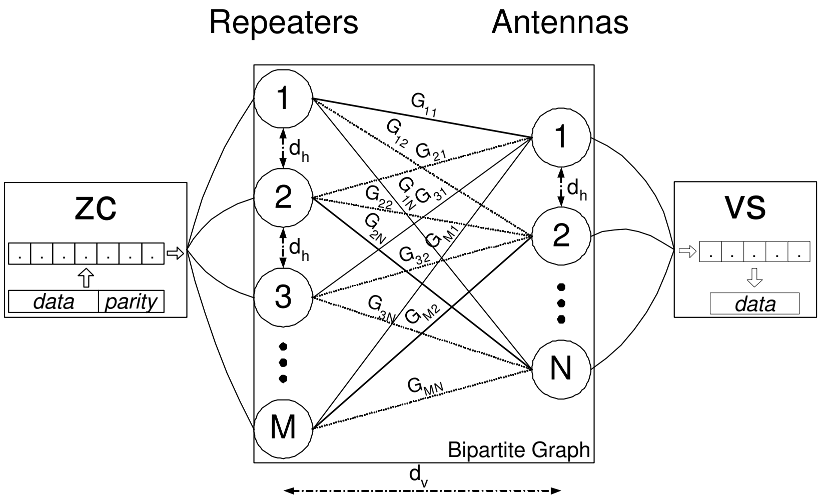

Fig. 4. System model from Zone Controller to Vehicle Station

Clearly, blind information raining is not performance-oriented; for instance, multiple antennas may tune to the same repeater, resulting in retrieval of redundant segments. However, the simple scheme has several implementation advantages. First, link establishment decisions are decentralized to the antennas. With off-the-shelf wireless components, it may be infeasible to pass channel estimation data to VS in real-time (i.e. with respect to channel coherence time) to generate a centralized decision. Second, there is no need for explicit MAC layer control. The repeaters and antennas are not required to "communicate" with each other in terms of per-link synchronization and addressing; repeaters only need to blindly transmit, and antennas only need to listen.

IV. Interference Model

Let us model our link layer architecture with downlink information travelling from ZC to VS, as shown in Figure 4. M repeaters are assumed to position in the vicinity of the vehicle, with distance d_r apart from each other. We assume that d_r, the *horizontal distance*, is larger than the coherence distance. These repeaters are assumed to receive segments from ZC via high quality links. At the vehicle side, N vehicle antennas are connected to the VS with high quality links, with distances d_a apart from each other. Let the shortest possible distance between a repeater and a vehicle antenna be d_v, the *vertical distance*.

At the physical layer, the air interface is shared by M transmitters and N receivers. We assume the employment of omni-directional antennas. A simple fading model is considered; all $M \times N$ possible subchannels exhibit flat-fading and slow-fading, such that fading coefficients of all subchannels remain constant over one frame. It is assumed that the noise is additive white-Gaussian with constant noise power N_0. Fading is considered to be contributed by two factors, one from the small-scale variations due to random multipath components, and the other from large-scale path loss which is distance dependent. Thus, link (power) gain from repeater i to antenna j is modelled as

$$G_{ij} = \alpha_{ij}^2 \mathcal{L}_0 \left(\frac{d_{ij}}{d_0} \right)^{-\kappa}, \tag{1}$$

where α_{ij} is a Rician distributed random variable which represents the envelope of a non-zero mean complex Gaussian random variable with Rician factor K, d_{ij} is the distance separating repeater i and antenna j, κ is the path loss exponent, and $\mathcal{L}_0$ is the signal attenuation level at the close-in reference distance d_0. The second term follows the popular log-distance path loss model. The model is appropriate because a direct line-of-sight can be observed in all subchannels. For simplicity, we assume that each antenna j can estimate G_{ij} for each repeater i with no error.

Because d_r and d_a are farther than coherence distance, α's of all subchannels are assumed to be mutually independent. Therefore, the elements in *link gain matrix* $G = (G_{ij})$ are also mutually independent. We further assume that G updates independently for every frame. Because the signals are Rician and multiple interference signals are also Rician regardless of matching decision, the model is referred to as Rician/Rician fading environment in the literature.

We model the environment of M repeaters and N antennas as a *complete weighted bipartite graph*[2], as shown in Figure 4. The antennas and repeaters are represented by two separate sets of vertices. The wireless link between every possible repeater-antenna pair is represented by an edge of the graph, with each end incident with a vertex of a different set. The link gain G_{ij} is associated as the weight of each edge (i, j). The choice of transmission with multiple wireless links among all repeater-antenna pairs, with the restriction of one-to-one connection at each repeater and antenna, corresponds to a *matching*[3] of the bipartite graph. For notation convenience, we define a matching X of the bipartite graph in two equivalent ways:

- In the form of a set of edges $X = \{(i, j)\}_{1 \le i \le M, 1 \le j \le N}$, where all edges $(i, j) \in X$ may be incident on repeater i or antenna j only once.

[2]A bipartite graph is any graph $G = (V, E)$ that can be partitioned into two mutually disjoint sets of vertices $V = X \dot{\cup} Y$ for which every edge in E has one vertex in X and the other in Y. A complete weighted bipartite graph is a bipartite graph where every pair of vertices with one vertex in X and the other in Y is an edge with an associated real number, or 'weight', in the graph.

[3]A matching in any graph G is defined as a set of edges, no two of which have a common end-vertex.

- In the form of a matrix $X = (x_{ij})_{1\le i\le M, 1\le j\le N}$, where $x_{ij} = 1$ if the link between repeater i and antenna j is chosen towards the matching, and $x_{ij} = 0$ if the link is not chosen. Clearly, the sum of every row or column of X must be 0 or 1 to qualify as a matching matrix.

Any one of the two notations is sufficient to define a bipartite matching. The use of either notation will be implicit henceforth. Moreover, a matching X has a *cardinality* of k when it contains k edges; we write $|X| = k$ in this case.

We further model each repeater i to transmit with signal power P_i and link rate r_i. Clearly, $P_i = 0$ and $r_i = 0$ if the repeater is inactive. We respectively define *power allocation vector* and *link rate allocation vector* as $\boldsymbol{P} = [P_i]_{1\le i\le M}$ and $\boldsymbol{r} = [r_i]_{1\le i\le M}$.

We further assume all wireless links communicate by the DS/SSMA technique. Given a matching X, the popular $E_b/(N_0 + I_0)$ metric on matched link (i, j) is

$$\gamma_{ij} = \left(\frac{W}{r_i}\right) \frac{G_{ij}P_i}{N_0 + \sum_{(k,l)\in X, k\neq i} G_{kj}P_k} \tag{2}$$

where W is the system bandwidth. The summation in the denominator is the interference experienced by link (i, j). Notice that $\frac{W}{r_i} > 1$ is the processing gain of the link.

Similar to the outage probability definition, we say that each link must exceed a required threshold γ_{th} to maintain link fidelity. For brevity, we shall henceforth refer to this criterion as the *SINR criterion*. If this condition is not satisfied, no segments are successfully received from the link. Therefore, ignoring the effects of erasure coding, we define *system throughput* as the normalized aggregate link rate from multiple links,

$$R_{system} = \frac{\gamma_{th}}{W} \sum_{\substack{(i,j)\in X \\ \gamma_{ij}\ge\gamma_{th}}} r_i \tag{3}$$

V. Throughput Optimization

In this section, we maximize the system throughput. We now consider VS with the freedom to decide upon matching X, power $\boldsymbol{P}$ and link rate allocation vector $\boldsymbol{r}$, given link gain matrix $\boldsymbol{G} = [G_{ij}]$ among all repeaters and antennas.

As defined previously, any link (i, j) must satisfy the SINR criterion to successfully receive segments. For any given X and $\boldsymbol{P}$, we propose to set each link rate r_i such that γ_{ij} is equal to (or barely exceeds) γ_{th}. The strategy is to utilize every excessive SINR to maximize system throughput. Since the minimum processing gain is 1, the link rate is upper bounded by W. For simplicity, we further assume that $r_i \in \mathbb{R}$.

Consequently, neglecting the W bound, we set each r_i as

$$r_i = \left(\frac{W}{\gamma_{th}}\right) \frac{G_{ij}P_i}{N_0 + \sum_{(k,l)\in X, k\neq i} G_{kj}P_k} \tag{4}$$

and thus the system throughput defined in (3) becomes

$$\begin{aligned} R_{system} &= \frac{\gamma_{th}}{W} \sum_{(i,j)\in X} r_i \\ &= \sum_{(i,j)\in X} \frac{G_{ij}P_i}{N_0 + \sum_{(k,l)\in X, k\neq i} G_{kj}P_k} \end{aligned} \tag{5}$$

which is our maximization objective function. We are now ready to pose our optimization problem:

$$\begin{array}{lll} maximize & R_{system} & \\ subject\ to & \sum_{j=1}^{N} x_{ij} = 0 \text{ or } 1 & (1\leq i\leq M) \\ & \sum_{i=1}^{M} x_{ij} = 0 \text{ or } 1 & (1\leq j\leq N) \\ & x_{ij} \in \{0,1\} & (1\leq i\leq M, 1\leq j\leq N) \\ & \frac{G_{ij}P_i}{N_0+\sum_{(k,l)\in X, k\neq i} G_{kj}P_k} \leq \gamma_{th} & (\forall\ x_{ij}=1) \\ & 0 \leq P_i \leq P_{max} & (1\leq i\leq M) \end{array} \tag{6}$$

The first three constraints arise from the definition of bipartite matching. If repeater i or antenna j is active, the row sum or column sum equals to 1, respectively. If it is inactive, the sum is 0. The fourth constraint stems from the processing gain restriction of link rates. Because a processing gain cannot be less than 1, $r_i > W$ is unrealizable. Bounding (4) with $r_i \leq W$ yields the fourth constraint. Henceforth, we shall refer to this constraint as the "W upper bound". Intuitively, there is no incentive for the system to provide a γ_{ij} so high that it breaks the fourth constraint; P_i can be decreased to reduce interference of other links without shrinking its link rate $r_i = W$, for instance. Finally, the last constraint is the regulatory or system limitations on transmitting power.

The MAC layer framer of throughput-optimized information raining is illustrated in Figure 5. Notice that some repeaters and antennas may be inactive, which is governed by the matching decision X. Individual link rates and power allocation may also differ from one another. Therefore, each active antenna must engage its matched repeater based on repeater identification signals, to perform link rate and power allocation. The link engagement phase (Signal E in Figure 5) may also include PN sequence synchronization, symbols training and MAC addressing.

Before a formal attempt to solve (6), we would like to provide some intuition to the problem. The freedom of "matching", to the best of our knowledge, has not been considered in previous optimization literature on wireless systems. In a general wireless network such as

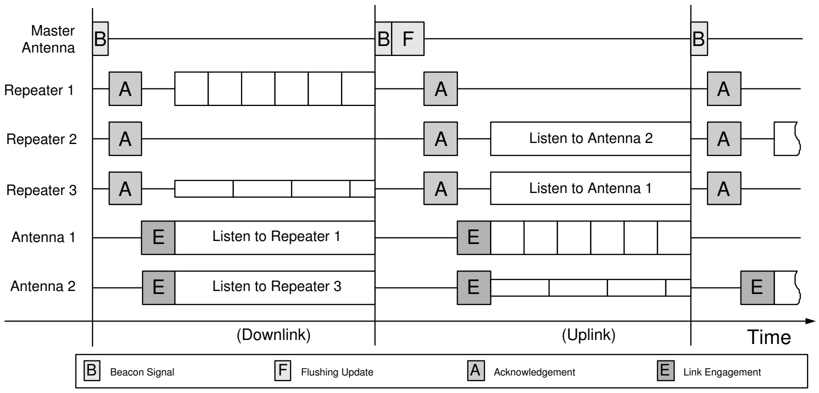

Fig. 5. MAC layer frame format for throughput-optimized information raining

cellular systems or ad-hoc wireless systems, pairs of transmitters and receivers communicate with each other with a fixed partner. Conversely in our setting, all matched links serve to transfer information between ZC and VS, and thus any link may be "sacrificed" in the optimization problem for the sake of interference reduction to increase system throughput. As such, the cardinality of X does not need to be equal to the number of antennas N; that is, not all antennas must be active to achieve the maximum system throughput. Consider an example where there exist only two antennas and two repeaters, with link gain $G_{11} = G_{12} = G_{21} = G_{22} = 1$ and $N_0 = 0.5mW$, with power $P_1 = P_2 = 1mW$. Then, activating one link achieves a better system throughput than activating both links.

Often, one's immediate instinct to the matching problem is that we shall match each antenna to its nearest repeater. This may be true in some cases, but may not hold in other instances. In general, the matching decision and cardinality of X depends on the background noise power N_0 and the distribution of the link gain matrix G, which in turn depends on other physical parameters such as repeater-antenna distances and fading effects.

In addition to bipartite matching, we need to find the optimal power allocation vector. The optimization problem of interest can be viewed as an integration of two problems of different qualities; a bipartite matching problem that is combinatorial in nature, and a power allocation problem that is algebraic in nature. Despite much effort, we are unable to produce an integrated algorithm that generates a joint solution for the power allocation vector and matching. As such, we shall consider our problem as two separate optimization problems: one problem that concerns with optimal matching X, and the other concerns with optimal power allocation $\boldsymbol{P}$. We propose to first solve the matching problem, and then use the resulting matching X to solve the corresponding power allocation problem. Clearly, our process is heuristic because the resulting power allocation and matching cannot guarantee a global maximum in system throughput among all combinations of power allocations and matchings. However, one can think of iterating the two problems to arrive at a globally optimum solution. We do not study possible convergence of such procedure in this paper and defer it to a later work. Nonetheless, as we shall see, both the power allocation problem and the matching problem are individually “hard” problems to solve.

VI. BIPARTITE MATCHING

Our first objective is to develop matching algorithms such that a high R_{system} may be reached when power allocation algorithms are processed with these “good” matchings. However, it is analytically difficult to derive a clear definition of good matchings, because the SINR criteria is a function of both $\boldsymbol{P}$ and X.

In order to be fair among possible links, we set equal constant power $P' \leq P_{max}$ among all repeaters, so $P_i = P'\ \forall i$. It is clear that the resulting problem is an Integer Program, the class of which are known to be NP-complete in general [26].

A related classical problem on matchings in bipartite graphs is the *assignment problem* [26], which is the quest to find the optimal assignment of workers to jobs that maximizes the sum of ratings, given all non-negative ratings c_{ij} of each worker i to each job j. Posed as an optimization problem, the assignment problem is as follows,

$$\begin{array}{ll} maximize & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{ij} \\ subject\ to & \sum_{i=1}^{N} x_{ij} = 1 \quad {\scriptstyle (1\leq j\leq N)} \\ & \sum_{j=1}^{N} x_{ij} = 1 \quad {\scriptstyle (1\leq i\leq N)} \\ & x_{ij} \in \{0,1\} \quad {\scriptstyle (1\leq i\leq N, 1\leq j\leq N)} \end{array} \tag{7}$$

The last constraint $x_{ij} \in \{0,1\}$ in (7) suggests that this is an Integer Linear Program (ILP)—the class of programs where polynomial-time algorithms do not exist in general. However, if we change the last constraint to be real-valued such that $x_{ij} \geq 0$, the overall constraint set becomes the so-called *assignment polytope*. The extreme points in the assignment polytope are always integral in all x_{ij}. Furthermore, the set of extreme points are in one-to-one correspondence with the set of possible matchings. Consequently, when the constraint $x_{ij} \geq 0$ is added, the integral constraint becomes redundant. The optimization problem in (7) becomes a LP of N^2 variables, which can be solved rather efficiently.

A. Hungarian Algorithm

The first polynomial-time solution of the assignment problem came from a combinatorial technique known as the *Hungarian method* by Kuhn [27] in 1955, due to the implicit work of Hungarian mathematicians Egeváry and König on the algorithm. The complexity of the algorithm is $O(N^3)$ [28, p.142]. The classic problem has been studied thoroughly since; more efficient algorithms have been found, where the best to date is the Hopcroft-Karp algorithm [29], which is a modification of the Hungarian algorithm that runs in $O(N^{2.5})$ time. However, results from Darby-Dowman [30] show that the Hopcroft-Karp approach is inferior to the Hungarian approach in practice due to the modified subproblem being larger than the amount of several single augmentations, by running tests on a significant set of problems.

Returning to our matching problem, one simplification is to ignore the W upper bound, such that our matching problem has the potential to relax the integrality condition. This approach is employed in all matching algorithms described in this paper. We argue that this simplification is tolerable because our power allocation algorithms can handle the W upper bound. Consequently, we consider our bipartite matching optimization problem as follows,

$$\begin{array}{lll} maximize & \sum_{i=1}^{M}\sum_{j=1}^{N}\left[\frac{G_{ij}x_{ij}}{N_0/P'+\sum_{\substack{k=1\\k\neq i}}^{M}\sum_{l=1}^{N}G_{kj}x_{kl}}\right] & \\ subject\ to & \sum_{j=1}^{N}x_{ij}=0 \text{ or } 1 & {\scriptstyle (1\leq i\leq M)} \\ & \sum_{i=1}^{M}x_{ij}=0 \text{ or } 1 & {\scriptstyle (1\leq j\leq N)} \\ & x_{ij}\in\{0,1\} & {\scriptstyle (1\leq i\leq M, 1\leq j\leq N)} \end{array} \tag{8}$$

Due to these simplifications, we now resolve to creating heuristics that yield “good” matchings. An obvious approach is a direct application of the Hungarian method; by assuming equal and constant interference among matchings, i.e. $N_0/P'+\sum_{\substack{k=1\\k\neq i}}^{M}\sum_{l=1}^{N}G_{kj}x_{kl}=C$, where C is a constant, (8) takes the form of the assignment problem, where we directly apply the Hungarian method on link gain matrix G to obtain matching.

The obvious weakness in this direct methodology is that all interference terms are neglected, generating a set of links that are “inconsiderate” to other links. As such, one may view the Hungarian method as a greedy approach to our matching problem. From this standpoint, it is also intuitive that the Hungarian method always yields a *maximal matching*[4]. We then depend on our power allocation algorithm to inactivate some of the links to reduce interferences in order

[4] A maximal matching is a matching that is not a proper subset of any other matchings in a graph. In our case, any matching X is a maximal matching if and only if $|X|=N$ in our complete bipartite graph.

Algorithm 1 The Hungarian Method

Let $\forall i \ a_i = \max_j\{c_{ij}\}, \forall j \ b_j = \max_i\{c_{ij}\}$; let $a = \sum_i a_i, b = \sum_j b_j$
if $a \leq b$ **then**
  Define $\forall i \ u_i = a_i, \forall j \ v_j = 0$
else if $a > b$ **then**
  Define $\forall i \ u_i = 0, \forall j \ v_j = b_j$
end if
From now on, associate an incidence matrix $Q = (q_{ij})$ with the rating matrix (c_{ij}) and cover $\{u_i, v_j\}$,
  $q_{ij} = \begin{cases} 1, & \text{if } u_i + v_j = c_{ij} \\ 0, & \text{otherwise} \end{cases}$
Initialize X: $\forall i, j \ x_{ij} = 0$
Initialize *row cover*: $\forall i \ s_i = 0$
while true **do**
  Call Routine *MaximumMatching* (Algorithm 2), which produces the matching X of the largest cardinality based on the incidence matrix Q
  if $|X| = N$ **then**
    Stop; we have found the solution X.
  end if
  Construct *column cover*: $t_j = \begin{cases} 1, & \text{if for any } i, \ x_{ij} = 1 \text{ and } s_i = 0 \\ 0, & \text{otherwise} \end{cases}$
  Let $d = \min\{u_i + v_j - c_{ij} | s_i = 0, t_j = 0\}$
  if $u_i > 0 \ \forall i | s_i = 0$ **then**
    Let $m = \min_i\{d, u_i\}$
    Update $u_i = u_i - m, \ \forall i | s_i = 0$
    Update $v_j = v_j + m, \ \forall j | t_j = 1$
  else
    Let $m = \min_j\{d, v_j\}$
    Update $u_i = u_i + m, \ \forall i | s_i = 1$
    Update $v_j = v_j - m, \ \forall j | t_j = 0$
  end if
end while

to maximize system throughput. We summarize the Hungarian algorithm in Algorithm 1 and Algorithm 2.

B. Hungarian Algorithm with Effective Weights

An alternative approach considers the case where we fix a subset, denoted $S \subseteq \{1, 2, \ldots, M\}$, of M repeaters to be active, and others inactive. Then, by setting all active repeaters to operate at P_{max}, the interference experienced by each antenna is fixed. We denote G'_{ij} as the *effective weight* experienced by repeater $i \in S$ and antenna j,

$$G'_{ij} = \frac{G_{ij}}{N_0/P_{max} + \sum_{\substack{k \in S \\ k \neq i}} G_{kj}} \tag{9}$$

Algorithm 2 Routine *MaximumMatching*

for all column j such that $\sum_i x_{ij} = 0$ **do**
  Attempt to construct an augmenting path starting at column j
    (through depth-first search):
  Initialize path $P = \{\}$
  Call Routine A
end for

Routine A
Search in column j of Q for i such that $q_{ij} = 1$ and i is not in any edges of P
if No such $q_{ij} = 1$ exists **then**
  if $P = \{\}$ **then**
    No augmenting path is found in our search; X is the maximum matching given Q
    Exit routine MaximumMatching
  else
    An alternating path is found;
    Set $s_i = 1, \forall(i, j') \in P$
    remove the last two edges from P
  end if
else
  for all i such that $q_{ij} = 1$ **do**
    Call Routine B for row i
  end for
end if

Routine B
Search in row i of X for j' such that $x_{ij'} = 1$
if No such $x_{ij'} = 1$ exists **then**
  An augmenting path is found; append $P = P \cup (i, j)$, and augment X by the following:
    $\forall(i, j) \in P \ x_{ij} = 1$, and $\forall(i, j') \in P \ x_{ij'} = 0$
    Clear row cover: $s_i = 0, \forall i$
  Restart Routine MaximumMatching
else
  append $P = P \cup (i, j) \cup (i, j')$, and then set $j = j'$
  Call Routine A for column j
end if

For a fixed S, we see that G'_{ij} is constant. Furthermore, our matching problem in (8) becomes

$$\begin{array}{lll} maximize & \sum_{i \in S} \sum_{j=1}^{N} G'_{ij} x_{ij} & \\ subject\ to & \sum_{j=1}^{N} x_{ij} = 0 \text{ or } 1 & {\scriptstyle (i \in S)} \\ & \sum_{i \in S} x_{ij} = 0 \text{ or } 1 & {\scriptstyle (1 \le j \le N)} \\ & x_{ij} = 0 & {\scriptstyle (i \notin S, 1 \le j \le N)} \\ & x_{ij} \in \{0, 1\} & {\scriptstyle (i \in S, 1 \le j \le N)} \end{array} \tag{10}$$

which is an assignment problem corresponding to the active repeaters $i \in S$, and can be solved by the Hungarian method. Given such a fixed instance of S, the solution yields the optimal system

1	2	1	4	3		1	2	1	4	3
2	3	4	1	2		2	4	3	1	2
3	1	3	2	4		3	3	2	1	4
4	4	1	2	3		4	2	1	3	4

Men's Preference Women's Preference

$$X_0 = \{(1,2),(2,3),(3,1),(4,4)\}$$
$$X_1 = \{(1,2),(2,4),(3,3),(4,1)\}$$
$$X_2 = \{(1,1),(2,4),(3,3),(4,2)\}$$

TABLE I

STABLE MARRIAGE INSTANCE OF SIZE 4 AND ITS SET OF SOLUTIONS

throughput with $P_{i \in S} = P_{max}$ and the omission of the W upper bound. Clearly, the effective-weight approach describes the system more accurately than the straight-forward approach given a subset S, but only with the application of such an S.

Ideally, we generate matchings from all 2^M subsets of repeaters, which is exponential to our problem size. Clearly that is not acceptable, and thus we propose to solve our matching problem with all repeaters being active. We label this approach as the *Hungarian algorithm with effective-weight*.

C. Stable Matching

As the third heuristics, we describe the *stable marriage problem* [31], [32], [33], which is the quest of finding *stable matchings* between N men and N women. Each person ranks all members of the opposite sex in strict order of preference. Given a maximal matching X that "marries" off each woman to each man, we denote $p_X(m)$ to be the woman that is married to man m, and $p_X(w)$ to be the man that is married to woman w. A man m and a woman w are said to *block* the matching X if m prefers w over $p_X(m)$ and w prefers m over $p_X(w)$. The existence of a *blocking pair* (m, w) represents a situation in real life in which the pair would run off with each other, breaking matching X. Thus, a *stable matching* is a matching without such a blocking pair, or otherwise it is an *unstable matching*. We illustrate an instance of the stable marriage problem and its set of solutions in Table I. For instance, man 1 prefers woman 2 over woman 1, over woman 4, and over woman 3. This particular example has 3 stable matchings, denoted as X_0, X_1 and X_2.

One of the first astonishing facts about the stable marriage problem is that every instance of the problem always admits at least one stable matching. The Gale-Shapley (GS) algorithm [31] generates one such stable matching in $O(N^2)$ time; a run-time that is surprisingly fast. Since

Gale and Shapley's fundamental result, the stable marriage problem has been explored in detail and efficient representations and algorithms to the problem have been discovered. For instance, the set of all stable matchings $\mathcal{X}$ can be well-represented by the so-called *rotation poset*, which can be constructed in $O(N^2)$ time, even though $|\mathcal{X}|$ may be exponential in N. More importantly, the rotation poset may be utilized to enumerate *all* stable matchings in $O(N^2 + N|\mathcal{X}|)$ time.

We cast our matching problem to the stable marriage problem, as follows. The "men's preference lists" are constructed by ranking the link gain of all antennas experienced by each repeater i. This is equivalent to sorting each row i of the link gain matrix $G = (G_{ij})$ in decreasing order and marking their original indices. Similarly, we construct the "women's preference lists" by ranking gain of all repeaters experienced by each antenna j, by sorting each column j of G. These constructions can be done in $O(N^2)$ time[5].

It has been shown in [34] that Exactly one stable matching can be found when solving for the corresponding stable marriage problem of (8). The proof concerns the underlying mathematical structure of the stable marriage problem, which is out of the intended scope of this paper. Nonetheless, because only one stable matching can be found, the use of GS algorithm is sufficient to find this one matching; it is unnecessary to construct the rotation poset and perform enumeration of "all" stable matchings. We summarize our stable matching algorithm in Algorithm 3; the GS algorithm is also presented as a routine.

VII. Power Allocation

We develop power allocation algorithms based on the above matching results. Let us consider a simpler problem of (6), where a fixed matching X has been given. Without loss of generality, let the matching X map each repeater i to antenna i, for all $i = 1, 2, \dots, |X| \leq N$. This can be achieved by appropriately re-ordering the indices of repeaters and antennas. Then, the optimization problem becomes

$$\begin{array}{lll} maximize & \sum_{i=1}^{|X|} \frac{G_{ii}P_i}{N_0 + \sum_{j=1, j\neq i}^{|X|} G_{ji}P_j} & \\ subject\ to & \frac{G_{ii}P_i}{N_0 + \sum_{j=1, j\neq i}^{|X|} G_{ji}P_j} \leq \gamma_{th} & (1 \leq i \leq |X|) \\ & 0 \leq P_i \leq P_{max} & (1 \leq i \leq |X|) \end{array} \tag{11}$$

[5]To deal with unequal number of repeaters and antennas when mapping to the stable marriage problem, we setup "dummy antennas" such that the number of repeaters and antennas become the same. These dummy antennas have arbitrarily small link gain to repeaters, and preference lists of the stable marriage problem are generated as usual. For any stable matching X found in the stable marriage problem, the links that cover these dummy antennas are removed.

Algorithm 3 Stable Matching Algorithm

Setup "dummy anntennas" j, $N < j \leq M$, by assigning link gain matrix G_{ij} with very small arbitrary values to all repeaters i;
for all $1 \leq i \leq M$ **do**
  Construct man i's preference list by ranking row i of G in decreasing order;
  Construct woman i's preference list by ranking column i of G in decreasing order;
end for
Solve for stable matching X^{stable} using GS algorithm;
Remove links of X^{stable} that cover dummy antennas;
Output X^{stable}.

Routine: (Man-oriented) Extended Gale-Shapley Algorithm
Assign each person to be free;
while some man m is free **do**
  Let w be the first woman on m's list;
  if w is already engaged with some man m' **then**
    Assign m' to be free;
  end if
  Assign the pair (m, w) to be engaged;
  for all man p that is behind m on w's list **do**
    remove p from w's list, and remove w from p's list;
  end for
end while

Before we continue with our discussion, we require several background definitions on convexity and linear fractional programming [35, Ch.2,3]. The notation $\boldsymbol{x}^k$ below serves as an index to the vector variable.

Definition 1: A function f is *pseudoconvex* on a non-empty open convex set $X \subseteq \mathbb{R}^n$ if f is differentiable, and if $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$, $(\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \nabla f(\boldsymbol{x}^2) \geq 0 \implies f(\boldsymbol{x}^1) \geq f(\boldsymbol{x}^2)$, or equivalently

$$f(\boldsymbol{x}^1) < f(\boldsymbol{x}^2) \implies (\boldsymbol{x}^1 - \boldsymbol{x}^2)^T \nabla f(\boldsymbol{x}^2) < 0$$

Definition 2: A function f is *explicit quasiconvex* on non-empty convex set $X \subseteq \mathbb{R}^n$ if $\forall\ \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$, $f(\boldsymbol{x}^1) \neq f(\boldsymbol{x}^2)$, and $\forall\ t \in (0,1)$, $f[t\boldsymbol{x}^1 + (1-t)\boldsymbol{x}^2] < \max[f(\boldsymbol{x}^1), f(\boldsymbol{x}^2)]$.

Definition 3: A *linear fractional function* $f : \mathbb{R}^n \to \mathbb{R}$ is of the form $f(\boldsymbol{x}) = (\boldsymbol{c}^T\boldsymbol{x} + c_0)/(\boldsymbol{d}^T\boldsymbol{x} + d_0)$. A *linear fractional program* is to maximize $f(\boldsymbol{x})$ subject to linear constraints, with the denominator $\boldsymbol{d}^T\boldsymbol{x} + d_0$ maintaining a constant sign (say positive) throughout the domain of feasible solutions (i.e. the set of points that satisfy all of the constraints).

Although a linear fraction is both explicit quasiconvex and explicit quasiconcave, nothing (either quasiconvex or quasiconcave) can be said about sums of linear fractions in general, of which R_{system} takes the form. However, derived in Appendix I, we show the below lemma holds true.

Lemma 7.1: The objective function in (11) is pseudoconvex on $\boldsymbol{P}$.

Since f is pseudoconvex, f is also explicit quasiconvex [35, p.46]. Because f is explicit quasiconvex, we also know that f reaches its global maximum in one or more *extreme points* in its feasible domain [35, p.49]. In conclusion, we know that the solution of (11) exists in one of the extreme points confined by its constraints. Therefore, similar to linear programming problems, we need to enumerate extreme points of our feasible set to find our solution. Unfortunately, as shown in Appendix II, this is an *NP-complete* problem.

Theorem 7.2: The optimization problem (11) is NP-complete.

Proof: See Appendix II. ■

In addition to the intractability of this problem, the proof also demonstrates that if instead of real-valued power control we only turn on or off the links, then the optimization problem of R_{system} given a fixed matching X will also be NP-complete. Because of this intractability, heuristics is proposed for power allocation.

We develop an algorithm that is similar to the revised simplex method for solving a LP. Although the objective function is not linear in $\boldsymbol{P}$, we remind again that all constraints of (11) are linear. Indeed, the defined feasible region is a *polytope* — a bounded intersection of a finite set of half-spaces. Similar to the simplex method, which belongs to the class of adjacent vertex methods, we enumerate through neighbouring extreme points of the polytope. At each iteration, we move towards the adjacent extreme point at which the objective function value experiences the largest increase. The process continues until an extreme point is reached such that one cannot find another adjacent extreme point to increase the objective function.

Similar to the conventional LP, the process is finite since the number of extreme points in a polytope is finite, while the objective function value increases per iteration. Because our objective function is explicit quasiconvex but not explicit quasiconcave nor concave, a point of local maximum does not guarantee a point of global maximum. Consequently, our simplex-type method cannot guarantee global maximum upon termination, unfortunately.

We now describe our simplex-type method for solving the power allocation problem, rearranged as standard equality form in

$$\begin{array}{ll} maximize & \sum_{i=1}^{|X|} \frac{G_{ii}P_i}{N_0+\sum_{j=1,j\neq i}^{|X|} G_{ji}P_j} \\ subject\ to & \frac{G_{ii}}{\gamma_{th}}P_i + \sum_{j=1,j\neq i}^{|X|} -G_{ji}P_j + s_i = N_0 \\ & P_i + t_i = P_{max} \\ & P_i, s_i, t_i \geq 0 \\ & {\scriptstyle (1\leq i\leq |X|)} \end{array} \quad (12)$$

Slack variables s_i, t_i, which are essential to the simplex method in general, are introduced in (12). The power allocation procedure is described in Algorithm 4.

Algorithm 4 Simplex-type Algorithm for Power Allocation

```
Form linear system Ax = b from constraints in (12),
   where x = [P^T, s^T, t^T]^T, and A, b are the appropriate coefficients;
Initialize starting feasible point x with P = 0, s = N_0, t = P_max;
Initialize the set of index positions of s, t at x as our initial feasible basis B;
Set R_system = R̂_system_max = 0;
loop
   for all entering index s ∉ B do
      Solve for A_B d = A_s, where A_s is the sth column of A, and A_B = [A_j : j ∈ B];
      for all leaving index r ∈ B do
         if d_r ≠ 0 and x_r/d_r ≥ 0 then
            Let t = x_r/d_r;
            Attempt pivot: Set x̂_B = x_B − td, x̂_s = t, B̂ = {B ∪ {s}} \ {r};
            if x̂ contains no negative elements, and yields R̂_system > R̂_system_max then
               Remember configuration: Set x̂_max = x̂, B̂_max = B̂, R̂_system_max = R̂_system;
            end if
         end if
      end for
   end for
   if R_system > R̂_system_max then
      Local maximum is reached; retrieve P_opt from x, and then exit.
   else
      Update configuration: Set x = x̂_max, B = B̂_max, R_system = R̂_system_max;
   end if
end loop
```

At each iteration of the simplex method in conventional LP, one can first find an *entering variable* for the pivot, regardless of the choice of *leaving variable*. The same does not hold true in our case; there is no definitive metric for the search of entering variables that is independent of the leaving variables. As such, we need to consider all combinations of entering and leaving variables at each iteration, as demonstrated by the two *for*-loops in Algorithm 4. The rules and procedures within the *for*-loops are similar to the revised simplex method; solve for linear system ($\boldsymbol{A_B d} = \boldsymbol{A_s}$), and then find the ratio t (unrelated to slack variables $\{t_i\}$). Because every

x_i in non-negative (restricted by $P_i, s_i, t_i \geq 0$ in (12)), we must ensure that every element of the new configuration $\hat{x}_i$ is also non-negative to act as a legal pivot. Therefore, $t = \hat{x}_s \geq 0$ is a requirement for a feasible solution $\hat{\boldsymbol{x}}$, and we may skip the pivot step when the requirement is not satisfied for a particular combination of entering index s and leaving index r.

VIII. Simulation Analysis

We simulate a typical scenario that reveals approximate values to some of the variables of the model discussed thus far. A fast-moving train travelling at 360km/h=100m/s that uses 2.4GHz carrier frequency has a maximum Doppler shift of $\Delta f = 800$Hz. As a rule of thumb, the corresponding channel coherence time may be approximated by $T_c = 0.423/\Delta f = 530\mu$s [36]. In order to qualify as a slow-fading channel, let the duration of frame-body be $0.1T_c = 53\mu$s. With a wireless link that operates at approximately 50Mbps, it can transmit 2650 bits per frame.

Let the length of the train be 150m, and antenna separation distance be $d_r = d_a = 15m$, that separates repeaters at $d_v = 3m$. We install $N = 10$ antennas on the train in this scenario. Notice that it takes 150ms for the train to pass one repeater, which is equivalent to 2800 frame periods. Thus, it is safe to assume that ZC can approximately track which repeaters are in the vicinity of the train, so that segments can be disseminated to them.

Lastly, let noise power $N_0 = 1$mW, system bandwidth $W = 100$MHz, and $E_b/(N_0 + I_0)$ metric threshold $\gamma_{th} = 10$dB. Let the path loss exponent be $\kappa = 2.7$, the close-in reference distance and attenuation be normalized to $d_0 = d_v = 3m$ and $\mathcal{L}_0 = 0$dB respectively. This is to fairly analyze SINR when comparing various system environments. Let the Rice factor be $K = 7$dB. In practice, the interference of adjacent transmitters using non-orthogonal DS/SSMA is usually represented by a constant factor smaller than 1. For simplicity, we have assumed that this interference factor is 1.

A. Link Rate Allocation in Blind Information Raining

In blind information raining, the lack of MAC layer control and channel information at the VS imply that repeaters are not individually managed by the vehicle. By symmetry, then all repeaters transmit with equal power and link rate. Obviously, each repeater should transmit at its individual maximum power P_{max} to mitigate the background noise. However, we need to choose an appropriate link rate, such that a decent aggregate data rate is achieved.

Intuitively, if link rate is set too low, then the aggregate rate, or system throughput, falls below its capability. If link rate is set too high, however, then many links lose their connectivity by failing to meet the SINR criteria, and system throughput again falls below its capability. Figure 6 plots a simulation of the average system throughput and outage probability against repeaters' link rate (normalized by γ_{th}/W), and with antennas perfectly aligned with the repeaters. The transmit power of each repeater i is set to $P_i = P_{max} = 1.0$mW. The outage probability in this case is defined as the probability that a segment transmitted from any repeater is not successfully received by any antenna. In blind information raining, there are two reasons of outage: 1) the transmitting repeater is not listened by any antennas during a frame, or 2) none of the listening antennas satisfy the SINR criteria. Furthermore, notice that a link may meet the criteria, but does not increase throughput because it tunes to a repeater from which another antenna successfully receives segments.

The plot agrees with our intuition on system throughput and outage probability; the system throughput initially increases with link rate, and then decreases due to more frequent link outages, as shown in the bottom subplot. The outage probability never falls below $3/13 \approx 0.23$ because there are more repeaters ($M = 13$) than antennas ($N = 10$), and thus at least 3 segments are not received per frame regardless of rate allocation. More importantly, it shows the existence of a global maximum system throughput around the normalized link rate of 7.0. Therefore, we can allocate an appropriate link rate to maximize system throughput.

B. Cyclicity and Anti-cycling

When the separation distance between adjacent repeaters, d_r, equals separation distance between adjacent antennas, d_a, a cyclical phenomenon in system throughput is observed as the train shifts forward, as shown in solid lines of Figure 7. The stand-alone solid line represents blind information raining with optimal normalized link rate, and the group of solid lines above it correspond to various throughput-optimized heuristics; it is not important to distinguish and differentiate these lines at the moment. The alignment position is the displacement, normalized by separation distance, of an antenna with respect to the nearest repeater behind it. We also remind that R_{system} is defined as the sum of (successful) link rates, normalized by γ_{th}/W. As such, both expressions are unitless.

The optimal system throughput reaches its maximum when repeaters and antennas are

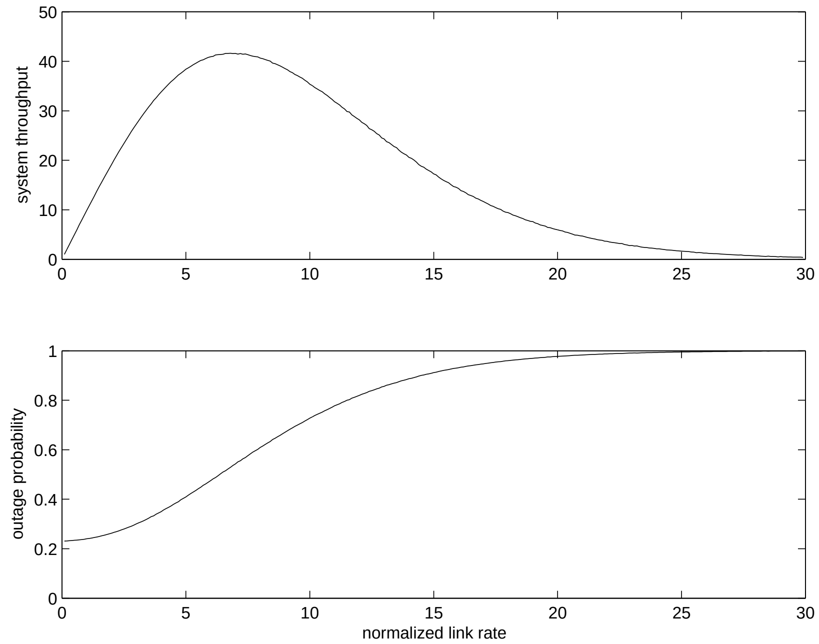

Fig. 6. System throughput and outage probability versus normalized link rate

perfectly aligned, and falls to its minimum when repeaters and antennas are half-way between each other. This is expected as link gain diminishes with distance, but not desired, as the communication service becomes fluctuational. Moreover, the optimal link rate also fluctuates with alignment position, which may become a problem to the system.

A simple approach to significantly reduce cyclicity is to vary separation distances. For instance, we may increase repeater separation distance by a small amount such that some repeaters are aligned with the nearest antennas, while others are mis-aligned, at any alignment position. The dashed lines of Figure 7 plot the scenario at which the repeater separation distance, denoted as d_r, is $N/(N-1)$ times the antenna separation distance, d_a, where N is the number of antennas on the train. Clearly, the fluctuation in system throughput is mostly eliminated, at the cost of a decrease in system throughput. This is due to greater repeater separation distances, resulting in less available repeaters in the vicinity of the train.

We refer to the process of avoiding cyclicity through specific setup of separation distances as *anti-cycling*. In anti-cycling, the choice of d_r is not only restricted to the value above. In Figure 8, we plot the maximum and minimum system throughput, which is achievable in all alignment positions, versus different repeater separation distances. We continue to fix $d_a = 15m$, and use stable matching with simplex-type power allocation in this plot. The difference between maximum and minimum system throughput is also plotted. As repeaters are separated further

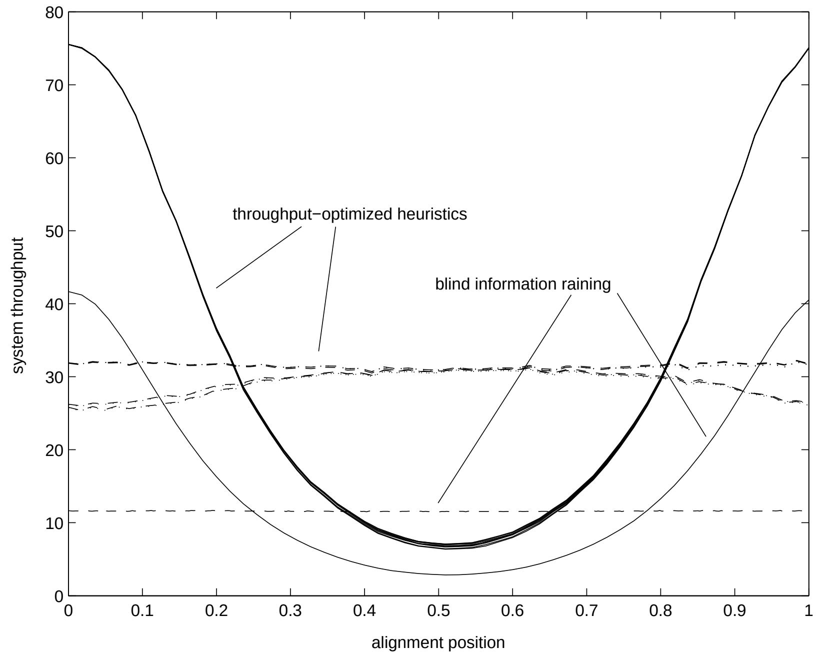

Fig. 7. System throughput versus alignment position in cycling and anti-cycling scenario

Interference-Limited	Noise-Limited
$d_r < d_v$	$d_r \gg d_v$
large number of active repeaters	small number of active repeaters
low noise power N_0	high noise power N_0
low path-loss exponent κ	high path-loss exponent κ

TABLE II

FACTORS OF INDUCING INTERFERENCE-LIMITED VERSUS NOISE-LIMITED ENVIRONMENTS

apart, the difference is decreased in general. However, at specific ranges of repeater separation distances, this difference can have local peaks and troughs. For instance, peaks can be observed when d_r/d_a reach integer values. This is expected because nearest pairs tend to meet and leave each other synchronously again.

C. The Role of Interference

We categorize some of these parameter variations with respect to interference power that are generated when all repeaters are active. We consider two extremes; when a change of some parameters results in high-level of interference experienced by antennas, we consider the environment as *interference-limited*. Conversely, when a change of some parameters results in low-level of interference with respect to background noise, we consider it as *noise-limited*. Table II lists the parameter changes that result in either of the extremes.

To illustrate the role of interference in the railway setting, let us consider a hypothetical

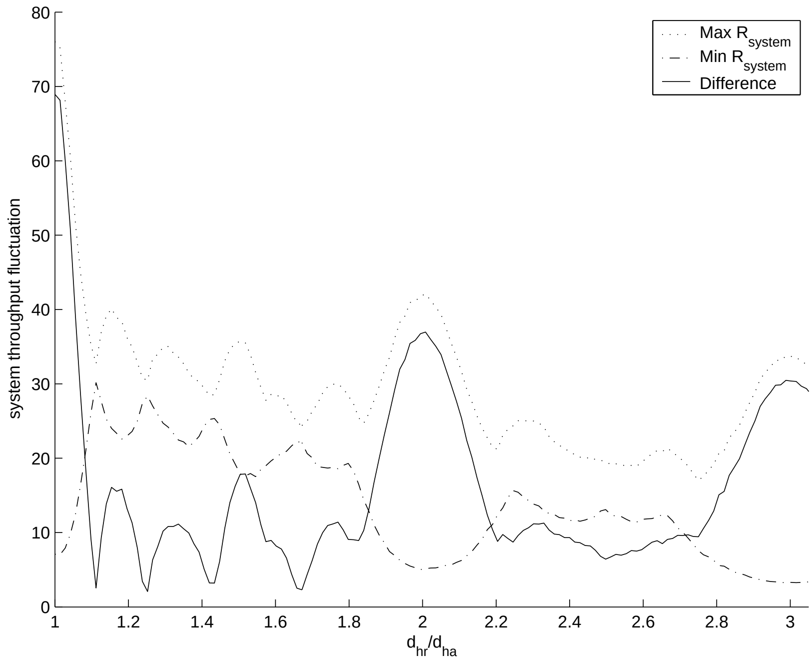

Fig. 8. System throughput fluctuation versus separation distance ratios

scenario, where a train with a fix length d_{train} is given. We then install N antennas on the train, spreading them apart with $d_a = d_{train}/N$. We set the relationship $d_r = \frac{N}{N-1}d_a$ for anti-cycling. We ask, how would an increase in the number of antennas installed enhance system throughput?

Figure 9 plots the system throughput and the number of successful links versus N under such a scenario with blind information raining (with optimal link rate chosen with each N), Hungarian algorithm, Hungarian algorithm with effective-weight, and stable matching algorithm. The simplex-type power allocation algorithm is employed in all heuristics. The number of successful links is defined as the average amount of links that count towards system throughput.

Notice that the system throughput of blind information raining is inferior to all heuristics, among which they are comparable with one another. Since the GS algorithm has the fastest run-time at $O(N^2)$, compared with $O(N^3)$ with Hungarian algorithm, we recommend the stable matching algorithm as the preferred matching algorithm in our proposed system.

When the number of antennas installed are few, they are distributed far apart from one another, which induces a noise-limited environment. Thus, it is possible to add an extra antenna to increase system throughput, as plotted by the figures. However, when a large number of antennas are installed, the environment becomes interference-limited. Adding an antenna in this crowded setup yields only a small benefit, as plots in this region reveal a “saturation” in system

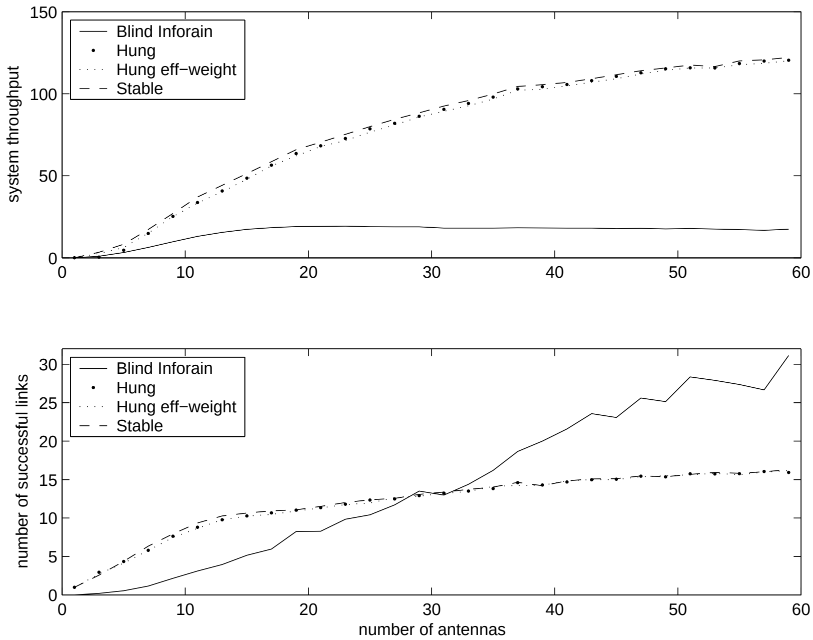

Fig. 9. System throughput and the number of successful links versus number of antennas, with fixed size of train

throughput.

The argument is further justified by observing the number of successful links in all heuristics with these two cases. With a small N, almost all antennas are actively receiving, revealing that it is noise-limited. As N becomes larger, the number of successful links begins to saturate towards a threshold, as the power allocation algorithm inactivates links to reduce interference.

In blind information raining, the number of successful links do not appear to saturate with large number of antennas installed; however, the resulting system throughput is much inferior to heuristics. Because all repeaters are transmitting in a dense area, the optimal link rate is chosen to be very low, as discussed in previous simulations, and thus a low system throughput is obtained. The system throughput is only comparable with heuristics when the number of antennas installed are very few ($N \leq 4$) and far apart, where the environment is strongly noise-limited. Although blind information raining cannot match heuristics in system throughput under most scenarios, it has several implementation advantages, as discussed previously.

The above observation demonstrates an analogy of DS-CDMA systems in cellular networks that are considered to be interference-limited. In order to reduce interference, many works have proposed a time-division multiplexing scheme in CDMA for non-realtime data, such that each base station only transmits to at most one user at a time [37], [38], [39]. In the framework of power allocation in CDMA, this is equivalent to allocating all power to one such user. Indeed, this scheme is shown to be energy efficient [40], and maximizes throughput in CDMA [41] and information-theoretic sum-capacity [42]. The inferior system throughput of blind information raining in the interference-limited scenario is explained by the lack of power allocation mechanism.

We can now appreciate the potential of spatial diversity in the railway setting. When very few antennas are installed, we are not exploiting the diversity that is inherited in the system, thus the system throughput attained is inferior, in magnitude, to the achievable system throughput. This is a noteworthy observation; as spectrum licenses auctioned from governmental organization become prohibitively expensive, the pressing need of efficient bandwidth management is evident. Together, our proposed architecture and information raining illustrate that a very high system throughput (per bandwidth W) is realizable in the railway system. Furthermore, because repeaters and antennas are near one another, transmission power is much lower than in cellular systems. Thus, bandwidth assigned to the railway system may be reused in other wireless systems, with minimal interference.

IX. Conclusions

In this paper, we have investigated a novel system architecture that enables high-speed access in railway systems. We further investigate the link layer design approach of the architecture. We have proposed blind and throughput-optimized information raining as transmission schemes between multiple repeaters and vehicle antennas. In throughput optimization, both matching and power allocation problems are individually shown to be NP-complete. Matching heuristics based on Hungarian algorithm and Gale-Shapley algorithm are proposed; Simplex-type approach is proposed in power allocation heuristics. Cyclicity, link rate allocation in information raining, and the role of interference are investigated through simulations of the setting.

Appendix I
Proof of Lemma 7.1

Let $f(\boldsymbol{P}) = \sum_{i=1}^{|X|} \frac{G_{ii}P_i}{N_0+\sum_{j\neq i} G_{ji}P_j}$. Then,

$$\frac{\partial f(\boldsymbol{P})}{\partial P_i} = \frac{G_{ii}}{N_0+\sum_{j\neq i} G_{ji}P_j} - \sum_{\substack{k=1\\k\neq i}}^{|X|} \frac{G_{kk}G_{ik}P_k}{(N_0+G_{ik}P_i+\sum_{l\neq i,l\neq k} G_{lk}P_l)^2}.$$

Therefore,

$$\begin{aligned}
& (\boldsymbol{P}^1 - \boldsymbol{P}^2)^T \nabla f(\boldsymbol{P}^2) \\
= \; & \sum_{i=1}^{|X|} (P_i^1 - P_i^2) \frac{\partial f(\boldsymbol{P}^2)}{\partial P_i} \\
= \; & \sum_{i=1}^{|X|} \frac{G_{ii}(P_i^1 - P_i^2)}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} - \sum_{i=1}^{|X|} \sum_{\substack{k=1 \\ k \neq i}}^{|X|} \frac{G_{kk} G_{ik} P_k^2 (P_i^1 - P_i^2)}{(N_0 + \sum_{l \neq k} G_{lk} P_l^2)^2} \\
& \text{(with a change of variables...)} \\
= \; & \sum_{i=1}^{|X|} \frac{G_{ii}(P_i^1 - P_i^2)}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} - \sum_{i=1}^{|X|} \sum_{\substack{k=1 \\ k \neq i}}^{|X|} \frac{G_{ii} G_{ki} P_i^2 (P_k^1 - P_k^2)}{(N_0 + \sum_{j \neq i} G_{ji} P_j^2)^2} \\
= \; & \sum_{i=1}^{|X|} \frac{G_{ii} P_i^1}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} \\
& - \sum_{i=1}^{|X|} \frac{G_{ii} P_i^2 [(N_0 + \sum_{j \neq i} G_{ji} P_j^2) + \sum_{k \neq i} G_{ki}(P_k^1 - P_k^2)]}{(N_0 + \sum_{j \neq i} G_{ji} P_j^2)^2} \\
= \; & \sum_{i=1}^{|X|} \frac{G_{ii} P_i^1}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} - \sum_{i=1}^{|X|} \frac{G_{ii} P_i^2 (N_0 + \sum_{j \neq i} G_{ji} P_j^1)}{(N_0 + \sum_{j \neq i} G_{ji} P_j^2)^2} \\
< \; & \sum_{i=1}^{|X|} \frac{G_{ii} P_i^1}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} - \sum_{i=1}^{|X|} \frac{G_{ii} P_i^1}{N_0 + \sum_{j \neq i} G_{ji} P_j^2} = 0,
\end{aligned}$$

if $f(\boldsymbol{P}^1) < f(\boldsymbol{P}^2)$.

APPENDIX II
PROOF OF THEOREM 7.2

Our proof shall follow the *reduction algorithm* technique, which is the classic approach in NP-completeness proofs [43, Ch.34].

We choose the *clique problem* as our reference NP-complete problem in this proof. A *clique* in a graph $G = (V, E)$ is a subset of vertices in V such that each pair of them is connected by an edge in E. In other words, a clique is a complete subgraph of G. Now, the clique problem is the optimization problem of finding a clique with the maximum size (measured by the number of vertices in the clique) in a graph, which is one of the well-known NP-complete problems [43, p.1003].

Our reduction algorithm is as follows. For every instance of the clique problem with graph $G' = (V, E)$, we index the vertices by 1,2,$\ldots$,$|V|$. Then, we produce our link gain matrix G of size $|V| \times |V|$, such that $G_{ii} = 1 \; \forall \; i$, and $G_{ij} = G_{ji} = \epsilon$ if $(V_i, V_j) \in E$ and $G_{ij} = G_{ji} = \infty$ if $(V_i, V_j) \notin E$, $\forall \; i, j$ that $i \neq j$. Let $\epsilon > 0$ be an arbitrary small number. Next, we set $\gamma_{th} = \infty$ and $N_0 = 1$. Notice that the feasible domain of (11) is now only constrained by $0 \leq P_i \leq P_{max}$, but not the W upper bound.

By our previous claim that the global optimal solution must lie on one of the extreme points of the feasible region, the solution $\boldsymbol{P}^{opt}$ must be of the form $P_i^{opt} = \epsilon$ or $P_{max}, \forall \; i$. Moreover, we perform a reverse-mapping to our clique problem with subset $C \in V$ such that $C = \{V_i | P_i = P_{max}\}$.

We now show that the clique problem has solution C when (11) has solution $\boldsymbol{P}^{opt}$. Suppose that $P_i^{opt} = P_{max}$ for some i. Then, no other $j \neq i, P_j^{opt} = P_{max}$ exists such that $G_{ij} = \infty$. This must be true because otherwise R_{system} will increase by inactivating both link i and j, which contradicts the optimality assumption. Hence, reverse-mapping of the optimal solution $\boldsymbol{P}^{opt}$ to C always yields a clique. In order to obtain the maximum R_{system}, while each link i yields approximately $G_{ii}/(N_0) = 1$ (with no $G_{ij} = \infty$), the number of links that are activated must be maximized. This implies that C is the clique of the maximum size.

Clearly, if there is a polynomial time algorithm that yields $\boldsymbol{P}^{opt}$ for (11), then we can construct our reduction algorithm in polynomial time, and thus create a polynomial time algorithm that yields C for the clique problem. But since the clique problem is known to be NP-complete, then by contradiction, there is no polynomial time algorithm for (11). Therefore, the optimization problem (11) is NP-complete.

REFERENCES

[1] C. Perkins, "IP mobility support for IPv4," RFC 3344, Aug. 2002.

[2] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," IETF Mobile IP Working Group Internet-Draft, June 2003.

[3] T. S. Rappaport, S. Y. Seidel, and R. Singh, "900 MHz multipath propagation measurements for US digital cellular radiotelephone," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 132–139, May 1990.

[4] P. Holmer, "Faster than a speeding bullet train," *IEEE Spectr.*, vol. 40, pp. 30–34, Aug. 2003.

[5] N. W. Whinnett, "Tetra high speed performance for police and rail applications," in *Proc. Seventh IEE European Conf. on Mobile and Personal Communications*, Brighton, UK, Dec. 1993, pp. 104–107.

[6] H. Hayashi, "1.5GHz band wave propagation characteristics and data transmission quality from train to ground," in *Proc. IEEE Vehicular Techn. Conference (VTC '92)*, Denver, CO, May 1992, pp. 192–195.

[7] T. Hattori and K. Abe, "Analyses of propagation characteristics in future railway communication systems using 25GHz band radio," in *Proc. 49th IEEE Vehicular Techn. Conference*, Houston, TX, May 1999, pp. 2288–2292.

[8] T. Hattori, "System design technique in the railway radio communication using microwave and millimeter-wave," in *Proc. IEEE Vehicular Techn. Conference, 2001. (VTC 2001 Fall).*, Atlantic City, NJ, Oct. 2001, pp. 703–707.

[9] (2003) Stay connected with accesszone from bell canada. [Online]. Available: http://www.viarail.ca/wirelessinternet/

[10] B. Brewin. (2003) Wi-fi to ride california passenger rails. [Online]. Available: http://www.computerworld.com/mobiletopics/mobile/story/0,10801,83930,00.html

[11] (2003) Gner mobile office - how does it work? [Online]. Available: http://www.gnermobileoffice.co.uk/GNERMobileOffice/howdoesitwork/

[12] M. Williams. (2003) Nec shows wi-fi at 205mph. [Online]. Available: http://www.infoworld.com/article/03/10/20/HNnecwifi_1.html

[13] E. Kuun and W. Rickard, "Open standards for CBTC and CCTV radio-based communication," *Alcatel Telecommunications Review*, pp. 243–252, 2nd Quarter 2004.

[14] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2105–2115, Oct. 2001.

[15] Y. Bejerano, "Efficient integration of multi-hop wireless and wired networks with QoS constraints," in *Proc. of the 8th annual international conference on Mobile computing and networking*. Atlanta, GA: ACM Press, Sept. 2002, pp. 215–226.

[16] J. Boyer, D. Falconer, and H. Yanikomeroglu, "A theoretical characterization of the multihop wireless communications channel with diversity," in *Proc. IEEE GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 841–845.

[17] H. Yanikomeroglu, "Cellular multihop communications: infrastructure-based relay network architecture for 4g wireless systems," in *Proc. the 22nd Queen's Biennial Symposium on Communications (QBSC'04)*, Queen's University, Kingston, Ontario, Canada, 2004, pp. 76–78.

[18] A. Tsirigos and Z. J. Haas, "Multipath routing in the presence of frequent topological changes," *IEEE Commun. Mag.*, vol. 39, pp. 132–138, Nov. 2001.

[19] E. Ayanoğlu, C.-L. I, R. D. Gitlin, and J. E. Mazo, "Diversity coding for transparent self-healing and fault-tolerant communication networks," *IEEE Trans. Commun.*, vol. 41, pp. 1677–1686, Nov. 1993.

[20] S. M. Cherry, "Broadband a go-go," *IEEE Spectr.*, vol. 40, pp. 20–25, June 2003.

[21] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[22] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1121–1128, May 1999.

[23] P. Djukic, "Optimum resource allocation in multipath ad hoc networks," MASc thesis, University of Toronto, Toronto, Canada, Aug. 2003.

[24] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," *Journal of the ACM*, vol. 36, pp. 335–348, Apr. 1989.

[25] A. J. McAuley, "Reliable broadband communication using a burst erasure correcting code," in *Proc. ACM SIGCOMM '90; (Special Issue Computer Communication Review)*, Philadelphia, PA, Sept. 1990, pp. 297–306.

[26] C. H. Papadimitriou, *Combinatorial optimization : algorithms and complexity*. Englewood Cliffs, N.J.: Prentice Hall, 1982.

[27] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[28] U. Derigs, *Programming in Networks and Graphs: On the Combinatorial Background and Near-Equivalence of Network Flow and Matching Algorithms*, ser. Lecture Notes in Economics and Mathematical Systems. Berlin Heidelberg: Springer-Verlag, 1988.

[29] J. E. Hopcroft and R. M. Karp, "An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs," *SIAM Journal on Computing*, pp. 2:225–231, 1973.

[30] G. Mitra and K. Darby-Dowman, "An investigation of algorithms used in restructuring of linear programming basis matrices prior to inversion, studies on graphs and discrete programming," in *Studies on graphs and discrete programming, Annals of Discrete Mathematics 11*, Amsterdam, North Holland, 1981, pp. 69–93.

[31] D. Gale and L. Shapley, "College admissions and the stability of marriage," *American Mathematical Monthly*, vol. 69, pp. 9–15, 1962.

[32] D. E. Knuth, *Stable Marriage and Its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms*, ser. CRM Proceedings & Lecture Notes. Providence, R.I.: American Mathematical Society, 1997, originally published in French in 1981; translated by M. Goldstein.

[33] D. Gusfield and R. W. Irving, *The Stable Marriage Problem: Structure and Algorithms*. Cambridge, MA: The MIT Press, 1989.

[34] D. Ho, "Link-layer design and throughput optimization of mobile hotspot in railway system," MASc Thesis, University of Toronto, Toronto, Canada, 2004.

[35] I. Stancu-Minasian, *Fractional Programming: Theory, Methods and Applications*. Dordrecht: Kluwer Academic Publishers, 1997.

[36] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Piscataway, NJ: IEEE Press, 1996.

[37] A. W. Anpalagan and E. S. Sousa, "A tagging-based medium access scheme for wireless data applications in CDMA/TDM networks," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, San Diego, CA, Sep/Oct 2001, pp. E–152–E–156.

[38] M. L. Honig and U. Madhow, "Hybrid intra-cell TDMA/inter-cell CDMA with inter-cell interference suppression for wireless networks," in *Proc. IEEE Vehicular Techn. Conf.*, Secaucus, NJ, May 1993, pp. 309–312.

[39] K. Navaie and S. Valaee, "Impact of heavy tail call duration on total interference in cellular cdma," in *The Eighth Canadian Workshop on Information Theory*, Waterloo, Ontario, Canada, May 2003, pp. 208–211.

[40] F. Berggren, S.-L. Kim, R. Jantti, and J. Zander, "Joint power control and intracell scheduling of DS-CDMA nonreal time data," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1860–1870, Oct. 2001.

[41] A. Bedekar, S. Borst, K. Ramanan, P. Whiting, and E. Yeh, "Downlink scheduling in CDMA data networks," in *Proc. IEEE GLOBECOM*, Rio de Janeireo, Brazil, Dec. 1999, pp. 2653–2657.

[42] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE International Conference on Communications (ICC 95)*, Seattle, WA, June 1995, pp. 331–335.

[43] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, 2nd ed. Cambridge, Massachusetts: The MIT Press and McGraw-Hill, 2001.