Fairness in Scheduling

Miklos Ajtai* James Aspnes† Moni Naor‡ Yuval Rabani§
Leonard J. Schulman¶ Orli Waarts‖

October 14, 1997

Abstract

On-line machine scheduling has been studied extensively, but the fundamental issue of fairness in scheduling is still mostly open. In this paper we explore the issue in settings where there are long living processes which should be repeatedly scheduled for various tasks throughout the lifetime of a system. For any such instance we develop a notion of *desired* load of a process, which is a function of the tasks it participates in. The *unfairness* of a system is the maximum, taken over all processes, of the difference between the desired load and the actual load.

An example of such a setting is the *carpool problem* suggested by Fagin and Williams [16]. In this problem, a set of n people form a carpool. On each day a subset of the people arrive and one of them is designated as the driver. A scheduling rule is required so that the driver will be determined in a 'fair' way.

We investigate this problem under various assumptions on the input distribution. We also show that the carpool problems can capture several other problems of fairness in scheduling.

1 Introduction

1.1 Our results

Consider the following *edge orientation problem*: on a set of n nodes labeled $\{1, \ldots, n\}$ there is a (possibly infinite) sequence of edges, i.e. pairs of nodes. Undirected edges arrive one by one, and each edge should be oriented upon its arrival. The goal is devise a method of orienting the edges so that in every node at every point in time the difference between the indegree and outdegree is as small as possible. This is measured by the *unfairness* at any time, defined to be half the maximum over nodes, of the difference between indegree and outdegree.

The problem comes in three flavors:

*IBM Research Division, Almaden Research Center. E-Mail: ajtai@almaden.ibm.com

†Dept. of Computer Science, Yale University. Supported by NSF grants CCR-9410228 and CCR-9415410. During part of this research this author was visiting IBM Almaden Research Center. E-Mail: aspnes@cs.yale.edu

‡Incumbent of the Morris and Rose Goldman Career Development Chair, Dept. of Applied Math and Computer Science, Weizmann Institute, Israel. Research supported by an Alon Fellowship and a grant from the Israel Science Foundation administered by the Israeli Academy of Sciences. E-Mail: naor@wisdom.weizmann.ac.il

§This work was done while the author was at ICSI, Berkeley, and at the Lab. for Computer Science, MIT. Work at ICSI supported by a Rothschild postdoctoral fellowship. Work at MIT supported by ARPA/Army contract DABT63-93-C-0038. Present address: Department of Computer Science, University of Toronto, Toronto, Ontario M5S 1A4, E-mail: rabani@cs.toronto.edu.

¶Computer Science Division, U.C. Berkeley. Work supported by an NSF postdoctoral fellowship. E-Mail: schulman@cs.berkeley.edu

‖Computer Science Division, U.C. Berkeley. Supported by an NSF postdoctoral fellowship. Part of this work was done while this author was at IBM Almaden Research Center. E-Mail: waarts@cs.berkeley.edu

1. Find a *deterministic* rule for orienting the edges and analyze it on the the worst input sequence.

2. Suggest a rule and analyze it under some assumption on the distribution of the sequence, in particular that each edge in the sequence is chosen uniformly from all possible edges and independently of the rest of the sequence.

3. Suggest a *randomized* rule for orienting the edges and analyze its expected performance on the worst sequence.

The greedy algorithm is the one where an edge is oriented from the node with the smaller difference between the outdegree and indegree to the one with the larger difference. In the deterministic version of the rule ties are broken according to the lexicographic order. In the randomized version of the rule ties are broken at random.

We address the three flavors of the problem and obtain the following results:

1. The optimal worst-case unfairness of a deterministic algorithm is linear there is a method (the greedy algorithm) that achieves the bound $\frac{n-1}{2}$ on unfairness, and for any deterministic rule there is a sequence where a difference of $\frac{1}{2}\lceil\frac{n-1}{2}\rceil$ will occur. (In the broader setting of the "carpool" problem, discussed below, a stronger lower bound of $\frac{n-1}{3}$ has been provided in [16].) These results are described in Section 2.

2. There is a randomized rule (local greedy) with expected unfairness $O(\sqrt{n \log n})$ on any sequence. The lower bound is $\Omega(\sqrt[3]{\log n})$. These results are described in Section 3.

3. The expected unfairness of the greedy algorithm on a uniform distribution on the edges is $\Theta(\log \log n)$ and we derive a complete description of the process in this case. This is the main technical contribution of the paper. These results are described in Section 4.

We view the edge orientation problem as a game played between an algorithm that chooses the edge orientations and an adversary that determines the sequence of edges. Each of the above cases corresponds to one of the three main types of adversaries treated in the literature: the *adaptive*, the *oblivious*, and the *uniformly random*, where the distinction is made according to the way the adversary determines which edges appear in the sequence. An adaptive adversary constructs the sequence on the fly, making decisions that may depend on the whole previous history of the game. An oblivious adversary must fix its sequence before the game starts, though it may choose this sequence based on knowledge of the algorithm. Finally, the uniformly random adversary produces a sequence in which each edge is chosen independently and uniformly at random.

In addition, we investigate the relationship between the edge orientation problem and the *vector rounding problem*, a very general problem to which many problems in fair scheduling can be reduced. In this problem, we are given a real matrix column by column and should produce an integer matrix so that each column in the output matrix is a rounding of the corresponding column in the input matrix that preserves the sum. The goal is to minimize the maximum over all rows of the difference between the sum of the rows in the integer and real matrix. (A formal definition can be found in Section 5.) We show:

4. A general transformation from the vector rounding problem to the edge orientation problem, at the price of doubling the expected difference. The transformation applies to both deterministic and randomized algorithms. It is described in Section 5.

1.2 Motivation

On-line machine scheduling has been studied extensively (see e.g. [3, 4, 5, 7, 14, 17, 14, 21]), but the issue of fairness in job allocation has usually not been considered quantitatively (however, see [2, 12, 15, 18, 19]). In a typical on-line scheduling problem, there are n machines and a number of separate jobs; the jobs arrive one by one, and each job must be assigned to exactly one of the machines, thereby increasing the *load* on this machine by an amount that depends on both the job and the machine. The goal of the scheduling problems studied in all the above literature is to minimize the maximum machine load. The situation in which this model seems most applicable is if all machines have one owner that wishes to optimize their utilization. If the machines have different owners, then fairness in allocation may be an additional, or primary, parameter to be optimized by the scheduler; for instance the dispatcher for a number of independently owned taxicabs.

Assuming that machines (or their owners) are reluctant (or eager) to do the required jobs, a "fair" rule, which takes into account the benefit to each machine (owner) of performing each task, must be applied. Thus, when faced with such a problem we should define the *desired* load of a machine (the fair share) and then suggest an algorithm for scheduling the jobs that tries to give each machine a number of jobs corresponding to its fair share.

An interesting property of the results we obtain in studying "fair" scheduling, is that there are scheduling protocols for which the discrepancy between the loads of the machines can be bounded in terms of functions only of the number of the machines, with no dependence on the elapsed time.

1.3 A general view: the Carpool Problem

The issue of fairness in scheduling was first isolated by Fagin and Williams [16], who abstracted it to what they call the *carpool* problem. A rough quotation from [16]: "Suppose that n people, tired of spending their time and money in gasoline lines, decide to form a carpool. Each day a subset of these people will arrive and one of them should drive. A scheduling algorithm is required for determining which person should drive on any given day. The algorithm should be perceived as fair by all members so as to encourage their continued participation." The analogy to fair machine scheduling is that the carpool participants correspond to the machines that are available to carry some task. The driver is the machine to which the task is actually assigned and thus the machine that incurs the cost of executing it.

The first question is how to define fairness. If the driver were not a member of the group, but a hired driver, then the meaning of fairness would be clear: The professional driver charges a fixed price for every ride, and each day the people that show up split the price of the driver equally among them. When the driver is just one of the set of people that show up, this reasoning leads immediately to the following definition of fairness given in [16]: If on a certain day, d people show up, each of them owes the driver $1/d$ of a ride. The *unfairness* of the algorithm at a certain point of the execution is defined as the maximum number of owed rides that anybody has accumulated up to that point or that anybody owes the rest of the group at that point. A scheduling algorithm is *fair* if there is a bound on the unfairness that is a function only of the number of drivers, and not of the schedule of arrivals (or in particular, the time elapsed). (Assuming an initial condition in which no one owes or is owed any rides.)

In case of random arrivals, we will evaluate algorithms according to the expected value of their unfairness when computed throughout the execution, and refer to this as *expected unfairness*.

Fagin and Williams proposed a natural algorithm for this problem, which we call the *global greedy* algorithm. When a set of people shows up, the one to drive will be the one that is currently the poorest. Ties are broken arbitrarily. A key contribution of their paper was to show that this algorithm is fair in the above sense. Namely, for a worst case sequence of requests, they showed that the unfairness of this algorithm is bounded above by a number that is exponential in the number of people, but independent of the number of days, and mentioned that Coppersmith managed to reduce this upper bound to linear (however, this proof is lost [13]). Finally, in collaboration with Coppersmith they showed a linear lower bound on the unfairness in this setting.

The edge orientation problem is simply a special case of the carpool problem, restricted to two people arriving each day. On the other hand, the general carpool problem is a special case of the vector rounding problem: each participant corresponds to a row, each day corresponds to a column; the ith entry of the jth column is 0 if the ith participant did not show up on the jth day, and is $\frac{1}{d_j}$ if d_j participants, including himself, did show up.

Therefore, an immediate byproduct of the results mentioned above on the edge orientation problem and of the general transformation to the vector rounding problem is that the general carpool problem has unfairness of $\Theta(n)$ against an adaptive adversary, and expected unfairness of $O(\sqrt{n \log n})$ against an oblivious adversary. (In fact, we also show directly that the natural greedy algorithm for the carpool problem maintains unfairness n against an adaptive adversary). Finally, against a random adversary, our results show that the carpool problem has expected unfairness $\Theta(\log \log n)$.

1.4 Comparison with competitive analysis

A popular methodology for evaluating the performance of on-line algorithm is the the *competitive analysis* approach of Sleator and Tarjan [20]: the on-line algorithm is compared with a hypothetical optimal off-line and bounds on the *competitive ratio* are obtained (for an adaptive, oblivious or random adversary). For the carpool problem, if one is given in advance a list specifying for each of the days which subset arrives on that day, then it is possible to construct a schedule whose unfairness is bounded by one (see Section 5). Therefore we can treat the results as being about the *competitive difference* of the carpool problem. If, instead, we would have analyzed the ratio between the an evenly distributed load (which is the best an off-line algorithm could hope to accomplish) and the actual load, we would have obtained a $1 + o(1)$ competitive ratio.

1.5 Other Related Work

Our problem is related to a *chip game* analyzed in [1]. In this game chips are placed in stacks on the integers, and in each round, two chips which are in the same stack may be selected, and one of them moved one step to the right while the other is moved one step to the left. This is the same thing that happens in the edge orientation game when a pair of vertices with the same indegree-outdegree difference is given; the difference between the games is that in ours pairs which are not colocated may also be selected (and moved toward each other). While our game can continue ad infinitum, the game of [1] must terminate, and some of the principal results of that paper concern the terminating states. In particular it is shown there that from any initial state of chips, there is a unique terminating position; and when n chips start all at the origin, no chip can be brought to distance more than $\lceil (n-1)/2 \rceil$ from the origin.

One can obtain an upper bound of $\lceil (n-1)/2 \rceil / 2$ on the maximum unfairness of the greedy algorithm for the edge orientation problem by a reduction to the case of the game of [1]. The reduction is to show that for the greedy algorithm, any sequence of requests for pairs of nodes can be replaced by another sequence, which reaches the same unfairness, but which uses no requests involving non-colocated pairs. (This reduction has also been noted recently by Babu Narayanan.) The worst-case performance of the greedy algorithm for the edge orientation problem is therefore exactly $\lceil (n-1)/2 \rceil /2$. An upper bound of $(n-1)/2$ for the more general carpool problem is given in section 2.1.

1.6 Organization of the paper

Section 2 gives linear upper and lower bounds on the unfairness of the global greedy algorithm with an adaptive adversary. Section 3 gives upper and lower bounds on the unfairness of the local greedy algorithm with an oblivious adversary. Section 4 gives a detailed characterization of the behavior of the global greedy algorithm when given a uniform random sequence of requests. Section 5 gives the reduction from the vector rounding problem to the two-person carpool game.

2 The Global Greedy Algorithm and the Adaptive Adversary

In this section we describe the behavior of algorithms for the carpool problem in the face of an adaptive adversary. We first show (Section 2.1) that the deterministic version of the global greedy algorithm guarantees an upper bound of $\frac{n-1}{2}$ on the maximum unfairness for any sequence of requests. This is within a factor of 2 of what is possible against an adaptive adversary; against *any* algorithm an adaptive adversary for edge orientation (2 people per car) can achieve $\lceil (n-1)/2 \rceil /2$ (Section 2.2), while if the adversary can schedule 3 people per car, the lower bound rises to $(n-1)/3$ (see [16]).

2.1 The global greedy algorithm

Against an adaptive adversary it is not much more difficult to solve the general carpool game than the edge orientation game. Thus we concentrate on the more general case here.

We consider the following on-line deterministic strategy for the n-participant carpool game. We maintain the *deviation* d_j for every $j \in [n]$. Initially, $d_j = 0$ for all $j \in [n]$. Given a request r (i.e., a subset of $[n]$ of cardinality 2 or more), the algorithm chooses $j \in r$ such that $d_j = \min_{i \in r} d_i$, breaking ties arbitrarily. The deviations are then updated as follows. d_j increases by $1 - 1/|r|$. For all other elements $i \in r$, $i \neq j$, d_i decreases by $1/|r|$. Other deviations remain the same. This strategy is the *global greedy* strategy of [16].

We show an upper bound on the unfairness resulting from the deterministic global greedy algorithm. We note first that the deviation d_j tracks the cost to each participant in the carpool game; thus for any adversary ϱ, the unfairness of global greedy is given by $\max_{j \in [n]} |d_j|$, where the values of d_j are taken at the end of the game.

Lemma 2.1 *Consider an n-participant carpool game between an adaptive adversary and the global greedy algorithm. For every round of the game there exists a weighted directed graph with node set $[n]$, edge set E and weight function w with the following properties.*

1.
$$\forall e \in E,\ \frac{1}{n!} \leq w(e) \leq \frac{1}{2}.$$

2.
$\forall e \in E$, $w(e) = \frac{p}{q}$, *where p, q are integers, and q divides $n!$.*

3.

$$\forall j \in [n],\ d_j = \sum_{e \in in(j)} w(e) - \sum_{e \in out(j)} w(e)$$

4. The graph contains no anti-parallel edges: at most one of the edges (i,j) or (j,i) is present in E.

where $in(j)$ is the set of incoming edges incident to j and $out(j)$ is the set of outgoing edges incident to j.

Proof: The proof is by induction on the number of rounds. In the basis no rounds have occurred: take an empty graph.

For the induction step, assume the claim holds for $t-1$ rounds. Let the t-th request of the adversary be $X_t = \{i_1, \ldots, i_k\}$.

Define $w(i,j)$ to be the weight of the directed edge from i to j, if such exists, or minus the weight of the directed edge from j to i, if such exists, or 0 otherwise.

Without loss of generality, assume that the global greedy algorithm selects $i = i_1$. We will modify the graph in two steps. The first modification, described below, maintains the conditions of the lemma for round $t-1$ and in addition establishes that there is no edge to i from any other node in X_t. The essential idea of this step is that since i is the poorest member of the group, any debts owed to i (corresponding to an incoming edge) can be redistributed to i's creditors without changing the deviation for any node in the graph. The second step adds an edge with weight $1/k$ from every i_j, $2 \le j \le k$ to i. This step has the effect of adding $1 - 1/k$ to d_{i_1} and subtracting $1/k$ from d_{i_j}. If these new edges create any pairs of anti-parallel edges, we merge each such pair to a single directed edge. Its weight is the difference between the larger and the smaller weight in the pair, and its direction coincides with the larger weighted edge in the pair. (If the two weights are equal, we remove both edges from the graph.)

To complete the proof we show how to do the first step. Suppose there exists $j \in X_t$ such that $w(j,i) > 0$. From the definition of the algorithm we have that after round $t-1$, $d_i \le d_j$. Thus, there exists $l \in [n]$ such that $w(l,j) > w(l,i)$.

We execute the following procedure.

while $w(j,i) > 0$ **do**
    Choose $l \in [n]$ such that $w(l,j) > w(l,i)$
    if $w(l,j) > 0$ **then**
        Let $w = \min\{1/2 - w(l,i), w(j,i), w(l,j)\}$,
        Increase $w(l,i)$ by w,
        Decrease $w(l,j)$ and $w(j,i)$ by w each.
    else $(w(l,j) < 0)$
        Let $w = \min\{1/2 - w(j,l), w(j,i), w(i,l)\}$,
        Increase $w(j,l)$ by w,
        Decrease $w(j,i)$ and $w(i,l)$ by w each.
stop .

Recall that by the inductive hypothesis, at the beginning of round t, $w(e) = \frac{p}{q}$, where p, q are integers and q divides $n!$. Clearly, this property is preserved by the above procedure (note that $n \ge 2$). Moreover, it implies that $w(e)$ continues to be $\ge 1/n!$ unless it becomes 0. Thus in both cases of the procedure above, in each iteration the sum of the weights over all edges in the graph decreases by $w \ge 1/n!$. Therefore, this process terminates. ∎

From the above lemma and the observation that unfairness is equal to $\max d_j$, it follows that:

Theorem 2.2 *Against any adversary ϱ, the unfairness of the global greedy algorithm is at most $\frac{n-1}{2}$.*

2.2 Lower bound against an adaptive adversary

In comparison, we note that the results for the *chip game* given in [1] (described in Section 1.5) provide a lower bound on the performance of any deterministic algorithm for the edge orientation game, which we can strengthen to apply to any algorithm facing an adaptive adversary. (Since the edge orientation game is a special case of the more general carpool game this lower bound applies to the carpool game as well.) If we think of a node's location as the difference between its indegree and outdegree, their results show that any sequence of requests which picks only colocated nodes will eventually result in the nodes occupying the entire interval $[-\lceil (n-1)/2\rceil, \lceil (n-1)/2\rceil]$ (except the origin in case n is even). Moreover it is shown there that the number of requests necessary to bring the nodes to this configuration is $n(n+1)(n+2)/24$ if n is even, and $(n-1)n(n+1)/24$ if n is odd. (We use this bound on the length of the request sequence later on to prove lower bounds for randomized algorithms.)

This result applies immediately to any deterministic algorithm against even an oblivious adversary, since the adversary can simulate the algorithm to determine which nodes are colocated. Against a randomized algorithm an adaptive adversary is required, since otherwise it cannot determine which nodes will be colocated. A much weaker lower bound, which applies to a randomized algorithm and an oblivious adversary, is given in Section 3.2.

We state the following general version of the result for later use:

Theorem 2.3 ([1]) *For every deterministic edge orientation algorithm f, for every $k \in \boldsymbol{Z}^+$, $k \leq \frac{1}{2}\lceil\frac{n-1}{2}\rceil$, there exists an oblivious adversary ϱ that gives a sequence of at most k^3 requests, pushing the unfairness achieved by f to at least k.*

3 The Local Greedy Algorithm and the Oblivious Adversary

In this section we consider the case of a sequence of requests supplied by an oblivious adversary. Section 3.1 describes an algorithm, the *local greedy algorithm*, that gives an upper bound of $O(\sqrt{n \log n})$ on the unfairness in the edge orientation game (the results of Section 5 allow this upper bound to be applied to the more general carpool game). Section 3.2 shows how an oblivious adversary can guarantee a lower bound of $\sqrt[3]{\log n}$, using techniques similar to those used by the adaptive adversary.

3.1 The local greedy algorithm

For the upper bound, it is convenient to consider only the edge orientation game described in the introduction. Through the reduction in section 5, our results apply to the general carpool problem as well.

The upper bound is obtained using the following *randomized local greedy* algorithm. For each pair of nodes (i,j) the algorithm keeps track of the difference $\Delta_{i,j}$ between the number of edges $n_{i,j}$ directed from i to j and the number of edges $n_{j,i}$ directed from j to i. When a new (undirected) edge $\{i,j\}$ arrives, it is directed from i to j if $\Delta_{i,j}$ is negative, from j to i if $\Delta_{i,j}$ is positive, and in a direction chosen by a fair coin-flip if $\Delta_{i,j}$ is zero. (Note that this definition is symmetric if we swap i and j, since $\Delta_{j,i} = -\Delta_{i,j}$.)

The algorithm is "locally greedy" in the sense that it always tries to minimize the difference between nodes i and j independently of their relations with other nodes. This guarantees that for each pair of nodes i and j, $|\Delta_{i,j}|$ is at most one.

The following theorem gives an upper bound on the expected unfairness obtained by the randomized local greedy algorithm:

Theorem 3.1 *Against any oblivious adversary, the randomized local greedy algorithm gives an expected maximum unfairness at any step of* $O(\sqrt{n \log n})$.

Proof: Fix a time in the execution and consider a particular vertex i. Let X_i be the random variable whose value is the difference between the indegree and outdegree of j. We can compute X_i as the sum of $\Delta_{i,j}$ over all nodes j distinct from i.

Now for each such j, if $\{i,j\}$ was requested an even number of times, then $\Delta_{i,j} = 0$. If $\{i,j\}$ was requested an odd number of times, then $\Delta_{i,j}$ is 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$ (the value being determined by the coin-flip that orients the last $\{i,j\}$ edge.) Furthermore, the $\Delta_{i,j}$ values are mutually independent for different values of j. By Chernoff's bound, $\forall a > 0$, $\Pr[|X_i| > a] < 2e^{-a^2/2k}$, where $k \leq n-1$ is the number of nodes i for which $\{i,j\}$ was requested an odd number of times. Summing up over all nodes i $\Pr[\exists i,\ |X_i| > a] \leq 2ne^{-a^2/2n}$. Taking $a = 2\sqrt{n \ln n}$ gives a probability of $2/n$. Since $\max_{j \in [n]} X_i \leq n-1$, the theorem follows. ∎

3.2 Lower bound against an oblivious adversary

The upper bound derived in the preceding section can be compared with the following lower bound for any algorithm facing an oblivious adversary.

Theorem 3.2 *For any algorithm for the edge orientation problem, there exists an input sequence that produces an expected maximum unfairness of* $\Omega\left(\sqrt[3]{\log n}\right)$.

Proof: Rather than showing how to construct such a sequence, we will define a single distribution over input sequences such that for any deterministic algorithm, the lower bound holds. By the minimax principle of von Neumann (see [23, 11]), the existence of this "mixed strategy" for the adversary that works equally well against any deterministic algorithm, implies that for each randomized algorithm there is a corresponding "pure strategy" for the adversary that achieves the same bound.

To get the adversary's mixed strategy, we modify the adaptive lower bound of Theorem 2.3. The essential idea is to divide the set $[n]$ of nodes into small disjoint subsets of size k (where k will be determined later). In each such subset we will run the at most k^3-long sequence from Theorem 2.3 based on a random "guess" about how the algorithm breaks ties. If we have enough subsets, with high probability one of these guesses will be correct and we will get the desired unfairness.

Let us make this informal description more precise. For simplicity, assume that n is divisible by k and that k is even. Let $\ell \leq k^3$ be the exact length of the sequence produced by Theorem 2.3 to achieve unfairness k. For each set we generate a random sequence of ℓ pairs of nodes whose distribution is given by following the strategy of Theorem 2.3 against an imaginary algorithm in which edges are oriented according to independent random coin flips. (The theorem applies since having fixed these ℓ independent random coin flips, the behavior of the algorithm is deterministic.)

There are 2^ℓ equiprobable sequences of coin flips. If the coin flips match the decisions made by the real algorithm within a particular group, then by Theorem 2.3, some node in this group reaches unfairness $k/2$. Let X_i, $i = 1, 2, \ldots, n/k$, denote the indicator variable for the event that a node in the i-th set reached unfairness $k/2$. Then $\forall i$, $Prob[X_i = 1] \geq 2^{-\ell} \geq 2^{-k^3}$. Since the X_i's are independent random variables, the probability that any X_i reaches $k/2$ is at least $1 - \left(1 - 2^{-k^3}\right)^{\frac{n}{k}}$. If we take $k \approx \sqrt[3]{\log n}$, this probability is bounded below by a constant. ■

4 The Global Greedy Algorithm and Uniform Requests

In this section we analyze the behavior of the global greedy algorithm for the edge orientation problem, in the case where the adversary schedules the edges uniformly at random. The combination of the adversary and the algorithm is represented as a Markov process. Since requests are uniform, we can forget about the identity of the nodes and describe a single state of the Markov chain by a list consisting of the number of nodes at each *position*, where the *position* of a node is given by its indegree minus its outdegree.

When two nodes are paired, they either each move one step away from the other, if they are at the same position; or each move one step toward the other if they are not. Intuitively, we can think of the process as a balance between a "repulsive force" between colocated nodes and an "attractive force" between distant ones. To stretch this physical analogy further, we would expect that the attractive force, being stronger in spread-out configurations, would tend to gather the nodes into a tight clump held apart only by pressure from the repulsive force.

Our first analysis of the system, in Section 4.1, shows that the nodes do in fact clump together, and that in the stationary distribution of the Markov chain the expected maximal unfairness is $O(\log n)$. We define a potential function on the states of the system, in which each node contributes an amount that is exponential in its deviation from 0 in that state. Since the Markov chain corresponding to the system is ergodic when $n \geq 3$, we can use the fact that in the stationary distribution the expected change in the value of the potential function is 0. We show that at any state where the maximal unfairness exceeds $O(\log n)$, the potential function is likely to drop by a large amount: the expected decrease in the value of the potential function is at least $n - 1$. On the other hand we show that from any state, the potential function can rise by at most 1. For these small rises to balance out the large drops in the states with unfairness greater than $O(\log n)$, the probability of "high" unfairness can be at most $O(1/n)$; and since (as we show) the unfairness of any state cannot exceed n, the expected unfairness is just $O(\log n)$.

However, in our simulations of the process it appeared that the maximum unfairness of the global greedy algorithm against a uniform random adversary was much smaller than $O(\log n)$. An approximation to the process (described in Section 4.2) suggested that the typical maximum unfairness was closer to $O(\log \log n)$. Proving this result without making the unwarranted assumptions needed for this approximation turned out to be quite difficult. However, after examining the process more closely, we obtained the tight asymptotic bounds of $\Theta(\log \log n)$ on the expected maximum unfairness described in Sections 4.3 and 4.4. We also show, in Section 4.5, that regardless of what state the process starts in, it quickly converges to this bound.

As the proofs are rather involved we give a simple overview here. We obtain the $O(\log \log n)$ upper bound through a sequence of tighter and tighter approximations. To start, we pick a time interval of length $n^{\log \log n}$, whose starting point t is any point in the execution. We show that with high probability, the maximal unfairness goes below $\log n$ by time $t + n^4$, and then stays below $2 \log n$ throughout the interval. For the next step we restrict our attention to the subinterval starting at $t + n^4$. For each $\epsilon > 0$, with high probability we can chop off a prefix of this new interval whose length is polynomial in n, leaving a suffix in which the unfairness of all but ϵn nodes is bounded by a constant k. For any such interval we show that with high probability, we can chop off a second prefix, whose length depends polynomially on n but not at all on k or ϵ, to leave a suffix in which at most $\epsilon^2 n$ nodes are above $k+2$. Repeating this operation $\log\log n$ times gives us an interval whose length is only polynomially less than the interval we originally started with, and in which the maximal unfairness is at most $O(\log\log n)$ (with high probability). Since for sufficiently large n the low-unfairness interval is much longer than the high-unfairness interval, it dominates the average and thus gives an $O(\log\log n)$ upper bound on the expected unfairness.

This analysis is tight: by time $t+n^5$, the unfairness is at least $\log\log n$ and stays above $\log\log n$ for at least $n^{\log n}$ additional steps. The proof of this lower bound mirrors the proof of the upper bound. We show that with high probability, any sufficiently long interval throughout which the unfairness of at least ϵn nodes is at least k contains a suffix, whose starting point is polynomially shifted, in which the unfairness of at least $c\epsilon^2 n$ nodes is at least $k+1$; after $\log\log n$ iterations of this process we are left with an interval whose length is close to the length of the original interval we picked, such that with high probability, throughout this resulting interval, the maximal unfairness is at least $\log\log n$.

4.1 A simple $O(\log n)$ upper bound

This section describes a simple $O(\log n)$ upper bound on the maximal fairness. This upper bound is the starting point of the sequence of approximations used to get the $O(\log\log n)$ bound in Section 4.3. The definitions given here of the behavior of the global greedy algorithm and of the state space will also be used in subsequent sections.

4.1.1 The Markov chain

We maintain a *position* d_j for each node j in $[n]$. Initially, $d_j = 0$ for all j. Given a request for a pair of nodes, the algorithm increases by one the position of the node whose current position is the smallest among the two, and decreases by one the position of the other particle in the pair. If two particles in the same position are requested, we flip an unbiased coin to determine which goes up and which goes down. Other positions remain the same. This is a randomized version of the *global greedy* strategy of [16]. We assume that the sequence generated is very long. The exact meaning of "very long" will be explained shortly. (We note that randomization of the on-line player is not essential to the analysis since, against the uniformly random adversary, the nodes may be considered unlabeled.)

Given such random input, the behavior of the global greedy algorithm can be represented as a Markov chain. By our analysis of the deterministic global greedy performance in Section 2.1, we know that $|d_j| \le \lceil (n-1)/2 \rceil$ for all $j \in [n]$. Thus, if the nodes are labeled, then the state space is $\{-\lceil (n-1)/2 \rceil, \ldots, \lceil (n-1)/2 \rceil\}^n$. The i-th coordinate of a state s, denoted s_i, is the position of the i-th node on the line. We now define the transitions and their probabilities. Let s be a state and $\{i,j\}$ a possible request. Without loss of generality, assume $s_i \le s_j$. If $s_i < s_j$, then with probability $\binom{n}{2}^{-1}$ there is a transition to s' with $s'_i = s_i + 1$, $s'_j = s_j - 1$ and for all $k \notin \{i,j\}$, $s'_k = s_k$. If $s_i = s_j$, then with probability $\binom{n}{2}^{-1}/2$ there is a transition to s' as above, and with the same probability there is a transition to s'' with $s''_i = s_i - 1$, $s''_j = s_j + 1$ and for all $k \notin \{i,j\}$, $s''_k = s_k$. For $n \ge 3$ it is easy to see that limited to the set of states reachable from the initial state of the all-zero vector, this Markov chain is ergodic and therefore converges to a stationary distribution. We are interested in the long-term behavior of the chain and therefore assume that the adversary sequence is long enough for the stationary behavior to be dominant.

If the nodes are unlabeled, which we can assume since we are considering the uniformly random adversary, then effectively we are interested in a smaller Markov chain. This smaller chain is a coarsening of the above chain in which each state is simply a count of the nodes lying at each position: thus, the state is represented by a vector $n_{-\lceil (n-1)/2 \rceil}, \ldots, n_{\lceil (n-1)/2 \rceil}$ where each n_i is the number of particles at position i. Let $p_j = n_j/n$. Note that if s is a state reachable from the all 0's vector, then $\sum_i i n_i = 0$.

4.1.2 The potential function

Here we define the potential function that we will use to show that the expected maximum unfairness is $O(\log n)$. Let $\alpha = \frac{3}{2}$ and let the potential function

$$\Phi(s) = \sum_{j=-\lceil (n-1)/2 \rceil}^{\lceil (n-1)/2 \rceil} n_j \cdot \alpha^{|j|}.$$

Let $\Delta\Phi(s) = E_{s'}[\Phi(s')] - \Phi(s)$, where s' denotes the (random) state reached from s in one step of the Markov chain.

We wish to estimate $\Delta\Phi(s)$. The following fact is easily verified:

Fact 4.1 *If s' is any outcome of requesting two nodes occupying the same position, then $\Phi(s') - \Phi(s) > 0$. If s' is the outcome of requesting two nodes that are at distance 1 apart, then $\Phi(s') - \Phi(s) = 0$. Otherwise, $\Phi(s') - \Phi(s) < 0$.*

Estimating $\Delta\Phi(s)$ is done by estimating the contribution of each position separately, and adding up those contributions. The idea is to show that for any $j \neq 0$, the positive contribution due to two nodes in j being requested is overwhelmed by the negative contribution due to a node in j and a node on the other side of 0 being requested. We will ignore other requests. They can only increase the negative contribution. In order to do this correctly, we need to consider disjoint events, so, to evaluate the contribution of position j, we will consider ordered pairs, where the first of the two is from j. The following fact is also easily verified:

Fact 4.2 *Under a uniform distribution over pairs of nodes,*

$$Prob[j,j] = Prob[\text{ordered } j,j] \leq p_j^2,$$

and

$$\begin{aligned} \frac{1}{2} Prob[i,j,i \neq j] &= Prob[\text{ordered } i,j] \\ &= Prob[\text{ordered } j,i] \geq p_i p_j. \end{aligned}$$

These relations are inequalities (rather than equalities) because we draw each pair without replacement; this makes it slightly less likely that we will draw two nodes at the same location.

Let A_j be the event that the first node in a pair is j (given that we are at configuration s). Formally, the *contribution* of position j to $\Delta\Phi(s)$ is $p_j E[\Phi(s') - \Phi(s) | A_j]$. We now show:

Lemma 4.3 *For j, $1 \leq |j| \leq \lceil (n-1)/2 \rceil$, the contribution of position j to $\Delta\Phi(s)$ is at most $-\frac{1}{6} p_j^2 \alpha^{|j|}$.*

Proof: Let $j > 0$. The argument for $j < 0$ is symmetric. If the pair j, j is chosen, the increase in the potential function is $\alpha^{j+1} + \alpha^{j-1} - 2\alpha^j$. On the other hand, if the (ordered) pair chosen is j, i ($i < 0$), the decrease in the potential function is $\alpha^j + \alpha^i - \alpha^{j-1} - \alpha^{i+1}$. We need an estimate on the distribution of nodes on the negative side. Since the sum of the positions of the nodes is 0, the n_j nodes at j must be balanced by nodes in negative positions. Hence:

$$\sum_{i<0}(-i)n_i \geq j \cdot n_j.$$

It is not difficult to see that the worst case (the least decrease in Φ) is when equality holds and when all the negative side nodes are in one position $-x$. Notice that we might need to consider a non-integral x. So, we get $xn_{-x} \geq jn_j$, or $p_{-x} \geq \frac{j}{x}p_j$. The total decrease in the potential function due to position j is at least:

$$p_j^2 \left(\frac{j}{x} \left(\alpha^j + \alpha^x - \alpha^{j-1} - \alpha^{x-1} \right) - \alpha^{j+1} - \alpha^{j-1} + 2\alpha^j \right),$$

for x minimizing this expression. Observe that this decrease is essentially the sum of two first derivatives of α^j, minus its second derivative. The basis of our lower bound on this expression is that for α below some threshold, the increase due to the first derivative (representing the choice of nodes at two different locations) dominates the decrease due to the second derivative (representing the choice of colocated nodes).

We show that for $j > 0$ the decrease is at least $p_j^2\alpha^j/6$, i.e. that for all $j \geq 1$ and $x > 0$,

$$\begin{aligned} & j \left(\alpha^j + \alpha^x - \alpha^{j-1} - \alpha^{x-1} \right) \\ > \; & x \left(\alpha^{j+1} + \alpha^{j-1} - 2\alpha^j + \frac{1}{6}\alpha^j \right), \end{aligned}$$

or

$$j \left(1 - \alpha^{-1} \right) \left(\alpha^j + \alpha^x \right) > x \left(\alpha + \alpha^{-1} - \frac{11}{6} \right) \alpha^j.$$

Recalling that $\alpha = \frac{3}{2}$ we want to prove that

$$\frac{1}{3} j \left(\alpha^j + \alpha^x \right) > \frac{1}{3} x \alpha^j.$$

If $x \leq j$ this is trivial. If $x > j$ write $r = x - j > 0$. We wish to show that $j\alpha^r > r$, and since $j \geq 1$ it suffices to show that $\alpha^r > r$ for all $r > 0$. Let $\delta = \min_r \alpha^r - r$. Some calculus shows that δ is achieved at $r = \frac{-\log\log\alpha}{\log\alpha}$; and moreover that δ varies monotonically in α. Thus we can solve for $\delta = 0$, finding that this is achieved for $\alpha = e^{1/e} \approx 1.4447$, and conclude that $\delta > 0$ for all $\alpha > e^{1/e}$ and in particular for the chosen $\alpha = \frac{3}{2}$. ∎

For $j = 0$ we cannot guarantee a negative contribution. However, we can upper bound the conditional positive contribution by $2\alpha - 2 = 1$, since the probability of choosing a pair in positions $(0, 0)$ is at most 1 and the total increase due to these positions is at most 1.

Concluding the above discussion: the contribution of position 0 is at most $+1$. The contribution of position j, $|j| \geq 1$ is at most $-\frac{1}{6}p_j^2(\frac{3}{2})^{|j|}$.

Let $T = 3\log_{\frac{3}{2}} n + \log_{\frac{3}{2}} 6$ and assume $n \geq 3$. Then, if s has a node whose distance from 0 is more than T, then $\Delta\Phi(s) \leq -n + 1$ (note that if this node's position is j, then $p_j \geq \frac{1}{n}$).

Now, partition the state space into two subsets: A contains those states that do not contain a node beyond T; B contains the other states. We have

Fact 4.4 $\forall a \in A$, $\Delta\Phi(a) \leq 1$. $\forall b \in B$, $\Delta\Phi(b) \leq -n+1$.

Since the total expected change in Φ, under the stationary distribution, must be 0, it must hold that under the stationary distribution, $Prob[B] \leq \frac{1}{n}$. It follows that:

Theorem 4.5 *For $n \geq 3$, in the stationary distribution, the probability that any node is beyond distance $T = 3\log_{\frac{3}{2}} n + \log_{\frac{3}{2}} 6$ from the origin is at most $1/n$. Thus the expectation of the maximum distance of a node from the origin is $\leq T+1 = O(\log n)$.*

4.2 An approximation suggesting the $O(\log\log n)$ bounds

This section describes an approximation by a dynamical system to the Markov chain described in the previous section. This approximation justifies the intuition that the "attractive force" operating between nodes (which we will think of as "particles") at all distances is likely to overwhelm the "repulsive force" operating on colocated nodes. However, it requires some assumptions that are not necessarily warranted in the edge orientation game, and thus must be taken only as a hint of the actual state of affairs. The true bounds are shown in Section 4.3 and 4.4.

Focus on a single particle. At each step, we pair our particle with another random particle, which we will call the *second* particle of the step. We can view our particle as making a random walk along the fairness axis. The probability that the unfairness of our particle will increase by one at a certain step equals (the probability that the step's second particle has higher unfairness) plus (half the probability that the step's second particle has the same unfairness). In other words, the dynamical system below describes our system **under two assumptions**:

1. We assume that the random walk of a single particle converges to a stationary distribution.
2. We assume that the probability of the second particle to have unfairness $\geq j$ is independent on whether our particle is at j.

Neither of these assumptions are necessarily true for the edge orientation process; however, making these assumptions appears to give a good approximation of the process when n is large.

We thus consider the following dynamical system:

$$\forall j,\ -m \leq j \leq m,\ p_j = r_{j-1}p_{j-1} + \ell_{j+1}p_{j+1},$$

$$p_{-m-1} = p_{m+1} = 0,$$

where

$$\forall j,\ -m < j < m,\ 1 - r_j = \ell_j = \sum_{i=-m}^{j-1} p_i + \frac{1}{2}p_j,$$

and

$$\ell_{-m} = r_m = 0.$$

Lemma 4.6 *There exists a (symmetric around 0) stationary distribution of this dynamical system such that*

1. $p_0 \geq \frac{1}{6}$;
2. $\forall j \neq 0,\ p_j \leq (\frac{3}{5})^{|j|-1}$.

Proof: Fold the dynamical system as follows. Let

$$q_j = \begin{cases} p_0 & j = 0 \\ p_j + p_{-j} & j > 0. \end{cases}$$

By symmetry, $\forall j \neq 0$, $q_j = 2p_j$. Also,

$$\Pr[\text{right move from } j] = \begin{cases} 1 & j = 0 \\ 0 & j = m \\ r_j & \text{otherwise.} \end{cases}$$

The following is a stationary distribution of the folded system.

$$q_j = \frac{r_1 r_2 \cdots r_{j-1}}{\ell_1 \ell_2 \cdots \ell_j} q_0.$$

Since

$$0 \le r_{j-1} \le r_{j-2} \le \cdots \le r_1 = \frac{1 - p_0 - p_1}{2} \le \frac{1}{2} \le \frac{1 + p_0 + p_1}{2} = \ell_1 \le \ell_2 \le \cdots \le \ell_j \le 1,$$

we have that

$$q_1 = \frac{q_0}{\ell_1}$$

$$q_0 \le \frac{q_0}{\ell_1} \le 2q_0$$

$$q_j \le \left(\frac{r_1}{\ell_1}\right)^{j-1} 2q_0.$$

Let $t = r_1/\ell_1$. We have that

$$t = \frac{1 - p_0 - p_1}{1 + p_0 + p_1} \le \frac{1 - \frac{3}{2}q_0}{1 + \frac{3}{2}q_0} = 1 - \frac{3q_0}{1 + \frac{3}{2}q_0}.$$

Since $3x/(1+3x/2)$ is monotonically increasing, we have that if $q_0 \ge \alpha$ then $t \le 1 - 3\alpha/(1+3\alpha/2)$. We have that

$$1 = \sum_{j=0}^{m} q_j \le q_0 \left(1 + 2\sum_{j=1}^{m} t^{j-1}\right) \le q_0 \left(1 + \frac{2}{1-t}\right) \le q_0 \left(1 + \frac{2(1 + \frac{3}{2}q_0)}{3q_0}\right),$$

where the last inequality follows from the assumption that $q_0 < 2/3$, justified by $t > 0$. We get that $q_0 + 2/3 + q_0 \ge 1$, or $q_0 \ge 1/6$. We also get $t \le 3/5$, which completes the proof. ■

Corollary 4.7 $q_{10} \le \frac{1}{3}$.

Lemma 4.8 *Using the above notation,* $\forall j \ge 10$, $q_j \ge \sqrt{q_{j+1}}$.

Proof: We prove by induction on i that $q_{m-i} \ge \sqrt{q_{m-i+1}}$.
Basis: $i = 0$. Trivial, since $q_{m+1} = 0$.
Inductive step: Assume correctness for $j = m - i$ and higher indices. We have that

$$q_j = q_{j-1} \frac{r_{j-1}}{\ell_j},$$

or

$$q_{j-1} r_{j-1} = q_j \ell_j.$$

Now,

$$r_j = \frac{1}{4} q_j + \frac{1}{2} \sum_{k=j+1}^{m} q_k,$$

and

$$\ell_j = 1 - r_j.$$

So,

$$q_{j-1} \left[\frac{1}{4} q_{j-1} + \frac{1}{2} \sum_{k=j}^{m} q_k \right] = q_j \left[1 - \frac{1}{4} q_j - \frac{1}{2} \sum_{k=j+1}^{m} q_k \right]. \tag{1}$$

Using the inductive hypothesis,

$$\sum_{k=j}^{m} q_k \leq q_j + q_j^2 + q_j^4 + q_j^8 + \cdots \leq \frac{5}{3} q_j \leq q_{j-1},$$

where the second inequality can be shown using Corollary 4.7 and the third inequality follows from the proof of Lemma 4.6. Similarly, we get from the inductive hypothesis

$$\sum_{k=j+1}^{m} q_k \leq q_j.$$

Plugging these inequalities into 1 gives

$$\frac{3}{4} q_{j-1}^2 \geq q_{j-1} \left[\frac{1}{4} q_{j-1} + \frac{1}{2} \sum_{k=j}^{m} q_k \right] = q_j \left[1 - \frac{1}{4} q_j - \frac{1}{2} \sum_{k=j+1}^{m} q_k \right] \geq \frac{3}{4} q_j,$$

where the last inequality assumes $q_j \leq \frac{1}{3}$. ■

Lemma 4.8 and Corollary 4.7 guarantee a double exponential decline of the distribution, thus providing an $O(\log \log m)$ bound on the expected deviation from 0.

4.3 Expected maximal unfairness is $O(\log \log n)$

This section gives the proof for the upper bound of $O(\log \log n)$ for the expected maximal unfairness of the greedy algorithm running against the uniform random adversary. The argument is based on a chain of implications of a special form described in Section 4.3.1. The argument uses several corollaries of Hoeffding's inequality for martingales (also known as Azuma's inequality); the statement and proof of these lemmas can be found in Appendix A. The argument itself appears in Section 4.3.2.

Since the argument is rather involved, we give here a road map to its intricacies. Lemma 4.11 shows that if we fix an interval $[t_1, t_1 + n^{\mu}]$ by choosing t_1 uniformly at random from the steps of a sufficiently long execution, then with high probability the maximal unfairness is $O(\log n)$ during that interval. Using a potential function argument (Lemmas 4.12 through 4.16) we show that this implies (with high probability, in the sense described in Section 4.3.1) that for all but a polynomially-sized prefix of the interval at most ϵn of the particles have an unfairness above some constant (Corollary 4.17). The next step is to show that if at most ϵn particles have an unfairness above k throughout some interval then (with high probability) at most $O(\epsilon^2)n$ are above $k+2$ beyond a polynomially-sized prefix of this interval (Lemma 4.18). To apply this fact repeatedly we need to prove a uniformity condition on the probabilities that the implication fails (Lemmas 4.19 and 4.20). Iterating it (Lemma 4.24) shows that, with high probability, at most $n^{9/10}$ particles go above some constant unfairness throughout a suffix of the original interval, which implies (from Lemmas 4.21 and 4.22) that with high probability no particle rises above this constant level plus $O(\log\log n)$ during yet another suffix of the original interval. Since this statement is highly qualified both in terms of its probability of occurrence and the interval during which it is likely to hold, a small amount of additional work is required to show that, for a time chosen uniformly from a large enough interval, the expected unfairness at that time is $O(\log\log n)$. This last step is the proof of Theorem 4.25.

The proof in this section does not address the issue of the rate of convergence to low maximal unfairness starting from an arbitrary state. In Section 4.5 we provide an analysis that, when combined with the analysis in this section, will provide the speed of convergence.

4.3.1 Probabilistic Delayed Implication

The proof of the upper bound of $O(\log\log n)$ for the expected maximal unfairness of the greedy algorithm works by analyzing the Markov process generated by the interaction between the deterministic global greedy algorithm and the uniform random adversary. In the end, it is shown that with high probability, for most of any sufficiently long execution the maximal unfairness is $O(\log\log n)$. This fact is the consequence of a chain of intermediate facts that characterize the behavior of the process over large intervals of the execution. For example, we will show that any execution is likely to contain long intervals during which the maximal unfairness is $O(\log n)$; using this fact we can then show that the process will tend to a situation where all but a constant fraction of the particles have at most a constant unfairness; and finally to one in which the expected maximal unfairness is $O(\log\log n)$ and remains so over a long interval.

Knowing only that some condition Γ (e.g., maximal unfairness is $O(\log n)$ holds throughout an interval will often not be enough to guarantee that some other condition Λ (e.g., maximal unfairness is $O(\log\log n)$) holds throughout the same interval, even though Γ describes conditions under which Λ is likely to become true. The reason for this is two-fold. Because the system consists of many small components, it may take time for the effect of Γ to propagate through the system and cause Λ to become true. And because the system is a random process, we will not be able to completely exclude the possibility that Λ does not happen or does not persist despite good conditions for its occurrence. Instead, the most we can say is that it is likely that if Γ holds throughout an interval, then Λ holds throughout a suffix of that interval.

We will express such statements as *probabilistic delayed implications.* Formally, suppose that for each n, Γ_n and Λ_n are unary relations defined on the set of all possible unfairness functions (for n people). If at time t the unfairness function satisfies Γ_n (alternatively, Λ_n) then we will say that $\Gamma_n(t)$ ($\Lambda_n(t)$) holds. If t_1, κ, μ are positive real numbers we will denote by $W(\Gamma_n, \Lambda_n, t_1, \kappa, \mu)$ the following statement:

$(\forall t \in [t_1, t_1+n^\mu], \Gamma_n(t)) \rightarrow (\forall t \in [t_1+n^\kappa, t_1+n^\mu], \Lambda_n(t))$

This says that if Γ_n holds throughout the interval $[t_1, t_1+n^\mu]$, then Λ_n holds throughout a suffix of that interval consisting of all but the first n^κ steps. Such a statement is a *delayed implication.* What turns it into a *probabilistic* delayed implication is its placement in the following context, which forms the basic structure of several of our lemmas:

$\forall\lambda > 0, \exists\kappa > 0, \forall\mu > 0$, *if n is sufficiently large then for all positive integers t_1 we have*

$$\Pr\left[W(\Gamma_n, \Lambda_n, t_1, \kappa, \mu)\right] \geq 1 - n^{-\lambda}.$$

If this statement holds then we will write $\Gamma \succ \Lambda$ or Γ pd-implies Λ. Since all of the variables in the statement except Γ and Λ are bound by the quantifiers this relation is well-defined. It is not difficult to see that if $\Gamma^1 \succ \Gamma^2 \succ \Gamma^3$ then $\Gamma^1 \succ \Gamma^3$.

Remark 4.9 *It will be important for our arguments that the entire implication is used to define the event whose probability is being measured. This makes it possible to estimate the probability in question by considering a collection of independent random events. If we put the premise outside, or equivalently if we use the probability of the consequence conditioned on the truth of the premise, then we will lose the independence of these events.*

Because the transitivity of $\succ$ depends on being able to change the exponent κ, it only works if we use it a constant number of times. However, the proof depends on being able to apply a chain of probabilistic delayed implications whose length is a function of n. To do so, we must first apply a uniformity condition. Given, for each n, an index set I_n and a set of pairs $\{\Gamma_n^\iota, \Lambda_n^\iota | \iota \in I_n\}$ of unary relations on the set of unfairness functions for n people, we will say that $\Gamma^\iota \succ \Lambda^\iota$ *uniformly in ι* if the following holds:

$\forall\lambda > 0, \exists\kappa > 0, \forall\mu > 0$ *if n is sufficiently large then for all positive integer t_1 and for all $\iota \in I_n$ we have*

$$\Pr\left[W(\Gamma_n^\iota, \Lambda_n^\iota, t_1, \kappa, \mu)\right] \geq 1 - n^{-\lambda}.$$

Uniformity gives us a stronger version of transitivity. Intuitively, if we have a chain of n uniform pd-implications, we can combine them so that the first relation in each chain pd-implies the last relation. Because the definition of uniform pd-implication is rather complicated this intuitive statement must be expanded on a bit:

Lemma 4.10 *Let that $I_n = \{1, \ldots, r_n\}$ where $r_n \leq n$, and for each n, let Θ_n^i, $i = 1, \ldots, r_n, r_n + 1$ be a sequence of unary relations on the set of unfairness functions for n people. For each $\iota \in I_n$ let $\Gamma_n^\iota = \Theta_n^\iota$, $\Lambda_n^\iota = \Theta_n^{\iota+1}$. Finally let $\bar{\Gamma}_n = \Theta_n^1$, $\bar{\Lambda}_n = \Theta_n^{r_n+1}$. If $\Gamma_n^\iota \succ \Lambda_n^\iota$ uniformly in ι then $\bar{\Gamma} \succ \bar{\Lambda}$.*

Proof: Suppose that $\lambda > 0$. We apply the definition of $\Gamma^\iota \succ \Lambda^\iota$ uniformly in ι, with λ replaced by $\lambda' = \lambda + 1$. Let $\kappa > 0$ be the number whose existence is guaranteed by the definition. We claim that for any t_1 and $\mu > 0$

$$Pr[W(\bar{\Gamma}_n, \bar{\Lambda}_n, t_1, \kappa + 1, \mu)] \geq 1 - n^{-\lambda}$$

The proof is by bounding the probability that $W(\bar{\Gamma}_n, \bar{\Lambda}_n, t_1, \kappa + 1, \mu)$ does not hold. If it does not hold, there is a positive integer j, $1 \leq j < n$ so that it is not true that

$(\forall t \in [t_1 + (j-1)n^\kappa, t_1 + n^\mu], \Theta_n^j(t)) \rightarrow \forall t \in [t_1 + jn^\kappa, t_1 + n^\mu], \Theta_n^{j+1}(t)$,

For any fixed $j \in \{1, \ldots, n-1\}$ the assumption that $\Gamma^\iota \succ \Lambda^\iota$ implies that the probability of the above event is at most $n^{-\lambda'} = n^{-\lambda-1}$. Since there are at most n choices for j, the probability that $W(\bar{\Gamma}_n, \bar{\Lambda}_n, t_1, \kappa + 1, \mu)$ fails is thus at most $n^{-\lambda}$. ∎

4.3.2 The Proof

We denote by $s(t)$ the state of our Markov chain at time t. We now define several random variables on $s(t)$. We abuse notation and define $s(i,t)$ as the position of particle i at time t. (Using this, $s(t) = \{(i, s(i,t)) \mid i \in [n]\}$.) $N_{=k}(t)$ denotes the cardinality of the set $\{i \mid |s(i,t)| = k\}$. Similarly, $N_{\geq k}(t)$ denotes the cardinality of the set $\{i \mid |s(i,t)| \geq k\}$. Since the behavior of the system depends on the position of uniformly-chosen participants, it will be convenient to normalize these quantities by dividing by n; accordingly, let $\rho_{=k}(t) = N_{=k}(t)/n$ and $\rho_{\geq k}(t) = N_{\geq k}(t)/n$. The quantity $\max(t)$ denotes the maximum k for which $N_{=k}(t) > 0$. For every α in the range $1 < \alpha < 2$, we define a potential function Φ_α over the state space of the Markov chain:

$$\Phi_\alpha(s(t)) = \sum_{i \in [n]} \alpha^{|s(i,t)|}.$$

In addition, we define,

$$\Delta\Phi_\alpha(s(t)) = E[\Phi_\alpha(s(t+1)) - \Phi_\alpha(s(t)) \mid s(t)].$$

(In other words, $\Delta\Phi_\alpha(s(t))$ is just the expected change in Φ_α in the next step of the process starting from $s(t)$.)

Let $E_j(t)$ be the event that the first particle in the pair chosen at time t is j. The *contribution* of position j to $\Delta\Phi_\alpha(s(t))$ is $\rho_j(t)E[\Phi_\alpha(s(t+1)) \mid s(t) \wedge E_j(t)] - \Phi_\alpha(s(t))$. Recall that Lemma 4.3 says that the contribution of position j is at most $-\frac{1}{6}p_j^2\alpha^{|j|}$ when $|j| \geq 1$.

Let T be an extremely large integer. We bound $E[\max(t)]$ on an interval whose starting point is chosen uniformly at random in $[T]$, and whose length is sufficiently long and fixed in advance. Let $t_1 \in_U [T]$, and let n^μ be the length of the interval we choose. From now on our goal is to bound $E[\max(t)]$ for $t \in [t_1, t_1 + n^\mu]$.

Let $\alpha = \frac{3}{2}$. The next lemma constitutes the first step in our proof.

Lemma 4.11 *Denote $\epsilon = \frac{\log_2 12}{\log_2 n}$. Let $t_1 \in_U [T]$. Let $\lambda > 0$, $c > 2 + \epsilon + \lambda$, $\mu < c - 2 - \epsilon - \lambda$. Then, if $n \geq 3$,*

$$\Pr[\forall t \in [t_1, t_1 + n^\mu],\ \max(t) \leq c \log_\alpha n] > 1 - n^{-\lambda}.$$

Proof: Lemma 4.3 of Section 4.1 states that the contribution of position j, $|j| \geq 1$ to $\Delta\Phi_\alpha(s(t))$ is at most $-\frac{1}{6}p_j^2\alpha^{|j|}$. If $n \geq 2$, $c \geq 1$, clearly $c \log_\alpha n \geq 1$. Thus, for every $c \geq 1$, the contribution of position $\pm c \log_\alpha n$ (if there are any particles there) is at most $-n^{c-2}/6$.

Partition the state space into two subsets: A contains those states that contain a particle i whose absolute position is at least $c \log_\alpha n$, which by the above discussion contribute at most $-n^{c-2}/6$ each; and B contains the other states. We thus have

$$\forall a \in A, \Delta\Phi_\alpha(a) \leq -n^{c-2}/6; \quad \forall b \in B, \Delta\Phi_\alpha(b) \leq 1.$$

The total expected change in Φ, under the stationary distribution, must be 0. Since with high probability t_1 occurs after we have come arbitrarily near to the stationary distribution, we have that for each $t \in [t_1, t_1 + n^\mu]$, the absolute expected change in $\Phi(s(t))$ is at most 1. Hence, $\Pr[s(t) \in A] \leq 12n^{2-c}$. Thus,

$$\Pr[\exists t \in (t_1, t_1 + n^\mu], \max(t) > c \log_\alpha n] \leq n^\mu 12 n^{2-c} = n^{2+\epsilon-c+\mu} < n^{-\lambda}.$$

Here n^ϵ substitutes for the constant 12, which would otherwise add clutter to the $n^{-\lambda}$ bound. ∎

We will need the following technical lemma:

Lemma 4.12 *Let $1 < \beta < \frac{41}{40}$. Let $\epsilon = \beta - 1$. Let $d > 4\beta^{3/\epsilon^2}$. Let n be sufficiently large. If*

$$\Phi_\beta(s(t)) \geq dn$$

then

$$\Delta\Phi_\beta(s(t)) \leq -\frac{\zeta}{40n}\Phi_\beta(s(t)),$$

where

$$\zeta = \epsilon - \epsilon^2 - \frac{\epsilon}{\beta^2} > 0.$$

Proof: Let C_i be the expected change in Φ conditioned upon i being the first element of the pair that marks the time t transition. We prove that for any fixed i we have

$$C_i < -\frac{\zeta}{40}\beta^{|s(i,t)|}.$$

Since the expected change in $\Phi_\beta(s(t))$ is $\frac{1}{n}\sum_i C_i$ this implies the lemma.

Suppose that i is fixed and let $j = s(i,t)$. Without loss of generality, assume that $j \geq 0$. Let i' denote the other particle that is hit at time t. We distinguish between three cases according to the value of j. We use the following:

$\beta^{j+1} - \beta^j = \epsilon\beta^j$,

$\beta^j - \beta^{j-1} = \epsilon\beta^{j-1} = \frac{\epsilon}{1+\epsilon}\beta^j \geq (\epsilon - \epsilon^2)\beta^j$, provided that $j > 0$.

Case 1. $j < \frac{3}{\epsilon^2}$. If $s(i',t) = j$, then Φ increases by

$$\beta^{j+1} + \beta^{j-1} - 2\beta^j \leq \epsilon\beta^j < \epsilon(1+\epsilon)^{3/\epsilon^2}.$$

In all other cases, Φ either stays the same or decreases. If $s(i',t) > \frac{3}{\epsilon^2}$, Φ must decrease, so we bound the expected decrease by considering particles from the range above $\frac{3}{\epsilon^2}$ only. Let N_i denote the expectation of the change in Φ conditioned on i being the first element in the pair and the other element in the pair being in position other than j. We get

$$\begin{aligned} N_i &\leq -\frac{1}{n}\sum_{i',|s(i',t)|>\frac{3}{\epsilon^2}} \left[(\epsilon - \epsilon^2)\beta^{s(i',t)} - \epsilon\beta^j\right] \\ &\leq -\frac{\zeta}{n}\sum_{i',|s(i',t)|>\frac{3}{\epsilon^2}} \beta^{s(i',t)}. \end{aligned}$$

Since $\Phi_\beta(s(t)) > dn$ and $d > 2\beta^{3/\epsilon^2}$, $\sum_{i',|s(i',t)|>3/\epsilon^2}\beta^{s(i',t)} > dn/2$. Therefore, $N_i < -\frac{\zeta d}{2}$. Thus,

$$\begin{aligned} C_i &< \epsilon(1+\epsilon)^{3/\epsilon^2} + N_i \\ &\leq -\frac{\zeta d}{4} \\ &\leq -\zeta\beta^j. \end{aligned}$$

Case 2. $j \geq \frac{3}{\epsilon^2}$ and $\sum_{j' \in [\epsilon j, \frac{1}{\epsilon}j]} p_{j'} \leq \frac{9}{10}$.

The positive contribution to the change in Φ (due to the case $s(i,t) = j$) at most $p_j((\beta^j - \beta^{j-1}) + (\beta^j - \beta^{j+1})) \leq p_j(-\epsilon\beta^j + (\epsilon^2 - \epsilon)\beta^j) = p_j\epsilon^2\beta^j$. In order to estimate the negative contribution, we ignore particles whose position is in the range $[\epsilon j, \frac{1}{\epsilon}j]$. They can only decrease the negative contribution. Therefore, the negative contribution is bounded above by

$$-\frac{1}{n}\sum_{i',s(i',t)>\frac{1}{\epsilon}j}\epsilon\left(\beta^{s(i',t)-1}-\beta^{j}\right)-\frac{1}{n}\sum_{i',0\leq s(i',t)<\epsilon j}\epsilon\left(\beta^{j-1}-\beta^{s(i',t)}\right)-\frac{1}{n}\sum_{i',s(i',t)<0}\epsilon\beta^{j-1}.$$

Recall that $\beta<\frac{3}{2}$. We use that for $j\geq\frac{3}{\epsilon^2}$, $j'\geq\frac{1}{\epsilon}j$, $\epsilon\leq 3/2$, it holds that

$$\beta^{j'}-\beta^{j}>\frac{1}{2}\beta^{j},$$

since $(1+\epsilon)^{3(1-\epsilon)/\epsilon^3}\geq 2^{3(1-\epsilon)/\epsilon^2}>\frac{3}{2}$. Also, for j,ϵ as above, $0\leq j'<\epsilon j$, we have

$$\beta^{j-1}-\beta^{j'}>\frac{1}{2}\beta^{j},$$

since

$$\frac{1}{2}\beta+\beta^{-3(1-\epsilon)/\epsilon^2}<\frac{3}{4}+\frac{1}{4}=1.$$

Obviously,

$$\beta^{j-1}>\frac{1}{2}\beta^{j}.$$

We conclude that

$$N_i<-\sum_{i',s(i',t)\notin[\epsilon j,\frac{1}{\epsilon}j]}\frac{\epsilon}{2n}\beta^{j}.$$

Since $\sum_{j'\in[\epsilon j,\frac{1}{\epsilon}j]}p_{j'}\leq 9/10$, we get $N_i<-\frac{1}{20}\epsilon\beta^j$, and therefore, since $\epsilon\leq\frac{1}{40}$, which gives $p_j\epsilon^2\beta^j\leq\frac{1}{40}\epsilon\beta^j$, we get

$$C_i<-\frac{1}{40}\epsilon\beta^{j}.$$

Case 3. $j\geq\frac{3}{\epsilon^2}$ and $\sum_{j'\in[\epsilon j,\frac{1}{\epsilon}j]}p_{j'}>\frac{9}{10}$.

The positive contribution to C_i is again at most $p_j\epsilon^2\beta^j$. In order to get an upper bound on the negative contribution we consider two subcases:

Case 3.a. $\sum_{i',s(i',t)\in[\epsilon j,\frac{1}{\epsilon}j]}s(i',t)\leq\frac{2nj}{10}$.

According to the assumption $\sum_{j'\in[\epsilon j,\frac{1}{\epsilon}j]}p_{j'}>\frac{9}{10}$, so the set $Y=\{i'|s(i',t)\in[\epsilon j,j/\epsilon]\}$ has at least $\frac{9}{10}n$ elements. Since the sum of the values of the function s on this set is at most $\frac{2}{10}nj$ there must be a $Y'\subseteq Y$, $|Y'|\geq\frac{4}{10}n$ so that for all $i'\in Y'$ we have $s(i',t)\leq\frac{j}{2}$. We consider the contribution due to elements of Y' only.

The contribution of a pair i,i' where $i'\in Y'$ is at most $-\epsilon(\beta^{j-1}-\beta^{s(i',t)})$. Since $s(i',t)\leq\frac{j}{2}$ and $j\geq 3/\epsilon^2$ We have that $\beta^{s(i',t)}\leq\frac{1}{2}\beta^{j-1}$ (using $\beta^{j/2-1}>2$). Since $\beta^{j-1}>\frac{2}{3}\beta^j$, we conclude that the contribution of the pair i,i' is at most $-\frac{1}{3}\epsilon\beta^j$. Since Y' has at least $\frac{4}{10}n$ elements we get that $N_i<-\frac{4}{30}\epsilon\beta^j<-\frac{1}{20}\epsilon\beta^j$. We conclude as in case 2 that

$$C_i<-\frac{1}{40}\epsilon\beta^{j}.$$

Case 3.b. $\sum_{i',s(i',t)\in[\epsilon j,j/\epsilon]}s(i',t)>\frac{2nj}{10}$.

The negative contribution is at most

$$-\frac{1}{n}\sum_{i',s(i',t)<0}\epsilon(\beta^j+\beta^{|s(i',t)|}).$$

Since $\sum_{i'} s(i',t)=0$, the assumption of case 3.b. implies that

$$K=\sum_{i',s(i',t)<0}|s(i',t)|\geq\sum_{i',s(i',t)\in[\epsilon j,j/\epsilon]}s(i',t)>\frac{2nj}{10}.$$

We will use the following fact.

Fact 4.13 *If H is a finite set with at most u elements and h is a nonnegative function on H and $\gamma>1$, $K>0$ and $\gamma^{K/u}>3$, then $\sum_{x\in H}\gamma^{h(x)}$ has its minimum over all nonnegative functions h and sets H with the conditions $\sum_{x\in H}h(x)=K$, $|H|\leq u$ if h is a constant and $|H|=u$.*

Observe that since $\sum_{j'\in[\epsilon j,\frac{1}{\epsilon}j]}p_{j'}>\frac{9}{10}$, the set $Y=\{i'|s(i',t)<0\}$ has at most $\frac{1}{10}n$ elements. Thus, we apply the above fact with $H:=\{i'|s(i',t)<0\}$, $u:=\lceil n/10\rceil$ and $\gamma:=\beta$. (Notice that $\beta^{K/u}\geq\beta^{2j}\geq\beta^{6/\epsilon^2}>3$.) Since $j\geq\frac{3}{\epsilon^2}$, we get that the negative contribution is at most

$$-\frac{1}{10}n\frac{1}{n}\epsilon\beta^{2j}\leq-\frac{1}{10}\epsilon\beta^j,$$

and hence

$$C_i<-\frac{1}{40}\epsilon\beta^j.$$

■

Next we can show:

Lemma 4.14 *Let $c>0$ such that $\beta=\alpha^{1/8c}<\frac{41}{40}$. (Notice that since $\alpha=\frac{3}{2}>1$ we have $\beta>1$.) Let $d>8\beta^{3/(\beta-1)^2}$. Let $\lambda>0$ and n sufficiently large. Then*

$$\Pr\left[\begin{array}{l}(\forall t\in[t_1,t_1+n^3],\ \max(t)\leq c\log_\alpha n)\rightarrow\\(\exists t\in[t_1,t_1+n^3],\ \Phi_\beta(s(t))\leq\frac{dn}{2})\end{array}\right]\geq 1-n^{-2\lambda}.$$

Proof: For all t, let $X_t=\Phi_\beta(s(t+t_1))$. Let Y_t be the indicator of

$$\max(t+t_1)\leq c\log_\alpha n \text{ and } X_t>dn/2.$$

We define random variables Z_t recursively as follows. For $t=0$, let $Z_0=\min\{X_0,n^2\}$. Notice that if $Y_0=1$ then $Z_0=X_0$. For $t>0$, if $Y_j=1$ for each $j\leq t$, then $Z_t=X_t$; otherwise, $Z_t=Z_{t-1}-1/\log n$. In effect, Z_t tracks X_t until the condition above is violated, after which it decays at the rate $-1/\log n$.

For sufficiently large n, the Z's satisfy the conditions of Corollary A.2 with $A=n^2$, $B=O(n^{1/8})$, and $C=-1/\log n$. (Notice that if n is sufficiently large, $-1/\log n$ is larger than the negative drift guaranteed by Lemma 4.12 in the case $X_t>dn/2$.) Take $\delta=\sqrt{4\lambda\ln n}$. From Corollary A.2 we get

$$\Pr[Z_{n^3}<0]>1-n^{-2\lambda}.$$

(We can use here $n/\log n>1+4\sqrt{\lambda\ln n}$.) By Fact A.3 we may conclude that

$$\Pr[\exists t\in[t_1,t_1+n^3],(\max(t)>c\log_\alpha n)\vee(\Phi_\beta(s(t))\leq dn/2)]\geq 1-n^{-2\lambda}.$$

■

Lemma 4.15 *Let c, d, λ be as in the previous lemma and n sufficiently large. Then*

$$\Pr\left[\begin{array}{l} [(\forall t \in [t_1, t_1+n^\mu],\ \max(t) \le c\log_\alpha n)\wedge \\ (\exists t \in [t_1, t_1+n^3],\ \Phi_\beta(s(t)) \le \frac{dn}{2})] \rightarrow \\ \forall t \in [t_1+n^3, t_1+n^\mu],\ \Phi_\beta(s(t)) \le dn \end{array}\right] \ge 1 - n^{-2\lambda}.$$

Proof: Let t_2 be the smallest $t \in [t_1, t_1+n^3]$ such that $\Phi_\beta(s(t)) \le dn/2$. If no such t exists, let $t_2 = \infty$. (Notice that t_2 is a random variable.) We show that

$$\Pr\left[\begin{array}{l} ((t_2 < \infty) \wedge (\forall t \in [t_2, t_1+n^\mu], \max(t) \le c\log_\alpha n)) \rightarrow \\ (\forall t \in [t_2, t_1+n^\mu],\ \Phi_\beta(s(t)) \le dn) \end{array}\right] \ge 1 - n^{-2\lambda}. \qquad (2)$$

Let $X_t = \Phi_\beta(s(t+t_2))$. Let Y_t be the indicator of

$$\boldsymbol{t_2 < \infty} \text{ and } \max(\boldsymbol{t+t_2}) \le \boldsymbol{c}\log_\alpha \boldsymbol{n}.$$

Define random variables Z_t as follows. If $t_2 < \infty$, then $Z_0 = X_0$; otherwise $Z_0 = dn/2$. For all $t > 0$, if $\forall j \le t$, $Y_j = 1$, then $Z_t = X_t$; otherwise $Z_t = Z_{t-1} - 1/\log n$. The Z's satisfy the conditions of Lemma A.5 with $\nu := 2(\lambda+\mu)$, $A, D := dn/2$, $B := n^{1/8}$, $C := 1/\log n$. (We can use here $n^{3/4}/(\log n) \ge 8(\lambda+\mu)/d$.) Therefore, we get that

$$\Pr[\forall t \in [0, n^\mu],\ Z_t \le dn] > 1 - n^{-2\lambda}.$$

Using Fact A.4, inequality 2 follows. ■

Lemma 4.16 *Let $c > 0$ such that $\beta = \alpha^{1/8c} < \frac{41}{40}$. Let $d > 8\beta^{3/(\beta-1)^2}$. Let*

$$\Gamma_n(t) \equiv \max(t) \le c\log_\alpha n,$$

$$\Lambda_n(t) \equiv \Phi_\beta(s(t)) \le dn.$$

Then $\Gamma_n \succ \Lambda_n$.

Proof: Suppose that $\lambda > 0$ is given. Let $\kappa = 3$ and let $\mu > 0$ be arbitrary. We show that for every sufficiently large n and for every fixed t_1,

$$\Pr\left[\begin{array}{l} (\forall t \in [t_1, t_1+n^\mu], \max(t) \le c\log_\alpha n) \rightarrow \\ (\forall t \in [t_1+n^\kappa, t_1+n^\mu],\ \Phi_\beta(s(t)) \le dn) \end{array}\right] \ge 1 - n^{-\lambda}.$$

This bound follows from

$$\begin{array}{rl} & \Pr\left[\begin{array}{l} (\forall t \in [t_1, t_1+n^\mu], \max(t) \le c\log_\alpha n) \rightarrow \\ (\forall t \in [t_1+n^3, t_1+n^\mu],\ \Phi_\beta(s(t)) \le dn) \end{array}\right] \\ \ge & \Pr\left[\begin{array}{l} \left(\begin{array}{l} (\forall t \in [t_1, t_1+n^\mu], \max(t) \le c\log_\alpha n) \rightarrow \\ (\exists t \in [t_1, t_1+n^3],\ \Phi_\beta(s(t)) \le \frac{dn}{2}) \end{array}\right) \wedge \\ \left(\begin{array}{l} ((\forall t \in [t_1, t_1+n^\mu], \max(t) \le c\log_\alpha n)\wedge \\ (\exists t \in [t_1, t_1+n^3],\ \Phi_\beta(s(t)) \le \frac{dn}{2})) \rightarrow \\ (\forall t \in [t_1+n^3, t_1+n^\mu],\ \Phi_\beta(s(t)) \le dn) \end{array}\right) \end{array}\right] \\ \ge & 1 - 2n^{-2\lambda} \\ \ge & 1 - n^{-\lambda} \text{ (we use here that } n^{2\lambda} > 2n^\lambda), \end{array}$$

where the inequality before the last follows from Lemmas 4.14 and 4.15. ■

Corollary 4.17 *Let $c > 0$ such that $\beta = \alpha^{1/8c} < \frac{41}{40}$. Let $\epsilon, c_{init} > 0$ such that $d = \epsilon\beta^{c_{init}} > 8\beta^{3/(\beta-1)^2}$. Let*

$$\Gamma_n(t) \equiv \max(t) \leq c\log_\alpha n,$$

$$\Lambda_n(t) \equiv N_{\geq c_{init}}(t) \leq \epsilon n$$

then $\Gamma \succ \Lambda$.

Proof: Suppose that $\lambda > 0$ is given. Let $\kappa = 3$ and let $\mu > 0$. Then for every sufficiently large n and for each fixed t_1, we have to show that

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1 + n^\mu], \max(t) \leq c\log_\alpha n) \rightarrow \\ (\forall t \in [t_1 + n^\kappa, t_1 + n^\mu], N_{\geq c_{init}}(t) \leq \epsilon n)\end{array}\right] \geq 1 - n^{-\lambda}.$$

Let $d = \epsilon\beta^{c_{init}} = \epsilon\alpha^{c_{init}/8c}$. By Lemma 4.16,

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1 + n^\mu], \max(t) \leq c\log_\alpha n) \rightarrow \\ (\forall t \in [t_1 + n^\kappa, t_1 + n^\mu], \Phi_\beta(s(t)) \leq dn)\end{array}\right] \geq 1 - n^{-\lambda}.$$

Now, if $\Phi_\beta \leq dn$, then there are at most ϵn particles whose absolute position is $\geq c_{init}$. ■

Corollary 4.17 completes the second step of our proof. We now proceed to the third step.

Lemma 4.18 *Let $\tau < \frac{1}{2}$, $t_1 \in \boldsymbol{Z}^+$. For every sufficiently large n, for every positive integer k,*

$$\Pr\left[\begin{array}{l}\forall t \in [t_1, t_1 + n^5],\ N_{\geq k}(t) \leq \tau n \rightarrow \\ \exists t \in [t_1, t_1 + n^5],\ N_{\geq k+2}(t) \leq 100\tau^2 n\end{array}\right] \geq 1 - 2^{-n+2}.$$

Proof: Let

$$U(t) = \sum_{i,|s(i,t)|\geq k+2} (|s(i,t)| - k - 1).$$

Let $\Delta U(t) = U(t+1) - U(t)$. Notice that $\Delta U(t) \in \{-1, 0, +1\}$ and that $\Delta U(t) = 1$ if and only if the time t transition is marked by a pair of particles in the same position $k+1$ or $-k-1$; $\Delta U(t) = -1$ if and only if the time t transition is marked by a pair of particles such that one is in position $\geq k+2$ and the other in position $\leq k$ or one in position $\leq -k-2$ and the other in position $\geq -k$; $\Delta U(t) = 0$ in all other cases. If $N_{\geq k}(t) \leq \tau n$ and $N_{\geq k+2}(t) > 100\tau^2 n$ then $\Pr[\Delta U(t) = 1] \leq \rho_{=k+1}(t)\rho_{\geq k+1}(t) \leq \tau^2$; and $\Pr[\Delta U(t) = -1] \geq \rho_{\geq k+2}(t)(1 - \rho_{\geq k+1}) \geq 100\tau^2(1-\tau) \geq 50\tau^2$.

Let W be the set of all $t > t_1$ with $\Delta U(t) \neq 0$ and let $w_1 < w_2 < \cdots$ be an enumeration of W in increasing order. (Notice that the w's are random variables.) Define a sequence of random variables $X_i = \Delta U(w_i)$. Notice that $\forall i$, $\sum_{j\leq i} X_j \geq -n^2$, since $\forall t$, $0 \leq U(t) \leq n^2$. Let Y_i be the indicator of

$$\forall t \in [t_1, t_1 + n^5],\ N_{\geq k}(t) \leq \tau n \text{ and } N_{\geq k+2}(w_i) > 100\tau^2 n.$$

Define a sequence of random variables Z_i as follows. If $\forall j \leq i$, $Y_j = 1$, then $Z_i = X_i$. Otherwise, $Z_i \in \{-1, +1\}$ is distributed independently of other Z's with $\Pr[Z_i = 1] = 1/50$.

The Z's satisfy the conditions of Lemma A.6. Therefore,

$$\Pr[\sum_{j=1}^{10n^2} Z_j > -5n^2] \leq 2\left(\frac{24}{50}\right)^{\frac{5}{2}n^2} < 2^{-n}.$$

By Fact A.3,

$$\Pr[(\exists t \in [t_1, t_1+n^5],\ N_{\geq k}(t) > \tau n) \vee (\exists i \in [1, 10n^2], N_{\geq k+2}(w_i) \leq 100\tau^2 n)] > 1 - 2^{-n}.$$

Since $\Pr[w_{10n^2} \leq n^5] > 1-2^{-n}$, the event in the lemma fails to hold with probability at most $3\cdot 2^{-n}$ and the lemma follows. ■

Lemma 4.19 *Let $c > 0$ such that $\beta = \alpha^{1/8c} < 2$. Let $\epsilon > 0$ such that $\beta(1-\epsilon) > 1$ and $700\epsilon \geq 2$. Let $t_1 \in \boldsymbol{Z}^+$, n sufficiently large, $n^{-1/10} < \tau < \frac{\beta(1-\epsilon)-1}{3\beta(\beta-1)}$, $\mu > 0$, $k \in \boldsymbol{Z}^+$, $\lambda > 0$. Then*

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1+n^\mu],\ \max(t) \leq c\log_\alpha n \wedge N_{\geq k}(t) \leq \tau n)\wedge \\ (\exists t \in [t_1, t_1+n^5],\ N_{\geq k+2}(t) \leq 100\tau^2 n) \rightarrow \\ (\forall t \in [t_1+n^5, t_1+n^\mu],\ N_{\geq k+2}(t) \leq 1000\tau^2 n)\end{array}\right] > 1 - n^{-2\lambda}.$$

Proof: Let t_2 be the smallest $t \in [t_1, t_1+n^5]$ such that $N_{\geq k+2}(t) \leq 100\tau^2 n$, or ∞, if no such t exists. (t_2 is a random variable.) We show that

$$\Pr\left[\begin{array}{l}(t_2 < \infty)\wedge \\ (\forall t \in [t_1, t_1+n^\mu],\ \max(t) \leq c\log_\alpha n)\wedge \\ (\forall t \in [t_1, t_1+n^\mu],\ N_{\geq k}(t) \leq \tau n) \rightarrow \\ \forall t \in [t_2, t_1+n^\mu],\ N_{\geq k+2}(t) \leq 1000\tau^2 n\end{array}\right] > 1 - n^{-2\lambda}, \tag{3}$$

which proves the lemma.

In order to show inequality 3, consider the following potential function:

$$\Psi(t) = \sum_{i, |s(i,t_2)| < k+2 \wedge |s(i,t)| \geq k+2} \beta^{(|s(i,t)|-k-2)}. \tag{4}$$

Notice that $\Psi(t_2) = 0$. Since $N_{\geq k+2}(t_2) \leq 100\tau^2 n$, we have that if $N_{\geq k+2}(t) > 1000\tau^2 n$, then $\Psi(t) > 900\tau^2 n$, so we would like to bound the probability that such an increase in Ψ occurs. We shall use the notation $\Delta\Psi(t) = \Psi(t+1) - \Psi(t)$.

Suppose that at time t, $\max(t) \leq c\log_\alpha n$ and $N_{\geq k}(t) \leq \tau n$ and $\Psi(t) \geq 700\tau^2 n$. Clearly, $\Delta\Psi(t) \leq n^{1/8}$. We need to estimate $E[\Delta\Psi(t)]$ in this case. Ψ may increase by 1 if a particle i with $|s(i,t)| = k+1$ moves away from 0. Under our assumptions, this happens with probability no greater than τ^2. Other contributions to the expected change in Ψ come from particles i that participate in the sum in equation 4. Consider such a particle i. The probability that it moves away from 0 is bounded above by $2\tau/n$. The probability that it moves towards 0 is bounded below by $(1-\tau)/n$. The expected change due to this particle is therefore bounded above by

$$\beta^{|s(i,t)|}\left(\frac{2\tau}{n}(\beta-1) - \frac{1-\tau}{n}\left(1-\frac{1}{\beta}\right)\right) \leq -\beta^{|s(i,t)|}\frac{\epsilon}{n}. \tag{5}$$

Since we are guaranteed that $\Psi(t) \geq 700\tau^2 n$, therefore if we sum up the bound in 5 for all i that contribute to Ψ we get that the expected change due to these particles is at most $-700\epsilon\tau^2 \leq -2\tau^2$. Thus, we may conclude that $E[\Delta\Psi(t)] \leq -2\tau^2 + \tau^2 \leq -n^{-1/5}$.

Now, define random variables $X_t = \Psi(t+t_2)$. Let Y_t be the indicator of

$$\boldsymbol{t_2 < \infty} \text{ and } \max(\boldsymbol{t+t_2}) \leq \boldsymbol{c}\log_\alpha \boldsymbol{n} \text{ and } \boldsymbol{N}_{\geq k}(\boldsymbol{t+t_2}) \leq \boldsymbol{\tau n}.$$

Define a sequence of random variables $Z_0, Z_1, \ldots$ as follows. $Z_0 = X_0$. For $t > 0$, if $\forall j \leq t$, $Y_j = 1$, then $Z_t = X_t$; otherwise, $Z_t = Z_{t-1} - n^{-1/5}$. The Z's satisfy the conditions of Lemma A.5 with $\nu := 2(\lambda+\mu)$, $A := 700\tau^2 n$, $B := n^{1/8}$, $C := n^{-1/5}$, $D := 200\tau^2 n$. Therefore, $\Pr[\exists t \in [n^\mu],\ Z_t > 900\tau^2 n] < 1 - n^{-2\lambda}$. Fact A.4 gives the required inequality 3. ■

Lemma 4.20 *For each positive integer n, let I_n be the set of all pairs $\langle k, \tau\rangle$, where k is a positive integer and τ is as in the previous lemma. Let*

$$\Gamma_n^{\langle k,\tau\rangle}(t) \equiv N_{\geq k}(t) \leq \tau n,$$

$$\Lambda_n^{\langle k,\tau\rangle}(t) \equiv N_{\geq k+2}(t) \leq 1000\tau^2 n,$$

for all $\langle k, \tau\rangle \in I_n$. Then $\Gamma_n^\iota \succ \Lambda_n^\iota$ uniformly in ι

Proof: Take $\kappa = 5$ and use Lemmas 4.18 and 4.19 ∎

Lemma 4.21 *Let j be an arbitrary particle. Let $t_1 \in \boldsymbol{Z}^+$. Let n be sufficiently large and k a positive integer. Then*

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1 + n^4],\ N_{\geq k}(t) \leq n^{9/10}) \rightarrow \\ (\exists t \in [t_1, t_1+n^4],\ |s(j,t)| \leq k)\end{array}\right] > 1 - 2^{-n/2}.$$

Proof: Let $\Delta(t) = |s(j, t+1)| - |s(j,t)|$. $\Delta(t) \in \{-1, 0, +1\}$. Let W be the set of all $t \geq t_1$ such that $\Delta(t) \neq 0$ and let $w_1 < w_2 < \cdots$ be an enumeration of W in increasing order. Let $X_i = \Delta(w_i)$. If $w_i \leq n^4$ and $|s(j, w_i)| > k$, then $\Pr[X_i = 1] \leq n^{-1/10}$, because the pair $\{j, j'\}$ that marks the time w_i transition must have $|s(j', w_i)| \geq k$. Notice also that for all i, $\sum_{i' \leq i} X_{i'} \geq -n$, because $|s(j, w_1)| = |s(j, t_1)| \leq n$ and $|s(j, w_{i+1})| \geq 0$. Let Y_i be the indicator of

$$\boldsymbol{\forall t \in [t_1, t_1 + n^4],\ N_{\geq k}(t) \leq n^{9/10}} \text{ and } \boldsymbol{|s(j, w_i)| > k}$$

Define a sequence of random variables $Z_1, Z_2, \ldots$ as follows. If $\forall i' \leq i$, $Y_{i'} = 1$, then $Z_i = X_i$; otherwise $Z_i \in \{-1, +1\}$ is distributed independently of other Z's with $\Pr[Z_i = 1] = n^{-1/10}$. The Z's satisfy the conditions of Lemma A.6 with $p := n^{-1/10}$. We get that

$$\Pr[\sum_{i=1}^{10n} Z_i > -5n] < 2(24n^{-1/10})^{\frac{5}{2}n} < 2^{-n}.$$

By Fact A.3,

$$\Pr[(\exists t \in [t_1, t_1+n^4],\ N_{\geq k}(t) > n^{9/10}) \vee (\exists i \in [1, 10n],\ |s(j, w_i)| \leq k)] > 1 - 2^{-n}.$$

Since $\Pr[w_{10n} \leq n^4] > 1 - 2^{-n}$, the lemma follows. ∎

Lemma 4.22 *Let j be an arbitrary particle. Let $t_1 \in \boldsymbol{Z}^+$. Let n be sufficiently large and k a positive integer. Let $\mu > 0$. Then*

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1+n^\mu],\ N_{\geq k}(t) \leq n^{9/10}) \wedge \\ (\exists t \in [t_1, t_1 + n^4],\ |s(j,t)| \leq k) \rightarrow \\ (\forall t \in [t_1 + n^4, t_1 + n^\mu],\ |s(j,t)| \leq k + \log\log n)\end{array}\right] > 1 - n^{-\frac{1}{100}\log\log n}.$$

Proof: Let t_2 be the smallest $t \in [t_1, t_1 + n^4]$ such that $|s(j,t)| \leq k$, or ∞, if no such t exists. We show that

$$\Pr\left[\begin{array}{l}(t_2 < \infty) \wedge \\ (\forall t \in [t_1, t_1 + n^\mu],\ N_{\geq k}(t) \leq n^{9/10}) \rightarrow \\ \forall t \in [t_2, t_1 + n^\mu],\ |s(j,t)| \leq k + \log\log n\end{array}\right] > 1 - n^{-\frac{1}{100}\log\log n}, \tag{6}$$

which proves the lemma.

In order to prove inequality 6, we apply an argument similar to that used in the proof of Lemma A.5 of considering at most $n^{2\mu}$ pairs t_3, t_4 and showing the following bound:

$$\Pr\left[\begin{array}{l}(t_2 < \infty)\wedge \\ (\forall t \in [t_1, t_1+n^\mu],\ N_{\geq k}(t) \leq n^{9/10})\wedge \\ (|s(j,t_3)| \leq k)\wedge \\ (\forall t \in (t_3, t_4],\ |s(j,t)| > k) \rightarrow |s(j,t_4)| > k + \log\log n\end{array}\right] < n^{-\frac{1}{100}\log\log n - 2\mu} \tag{7}$$

Using the same notation as in the previous lemma, let W be the set of $t \geq t_3$ such that $\Delta(t) \neq 0$. Let $w_1 < w_2 < \cdots$ be an enumeration of W in increasing order. Define $X_i = \Delta(w_i)$. If $w_i \leq t_1 + n^\mu$ and $\forall t \in [t_1, t_1+n^\mu]$, $N_{\geq k}(t) \leq n^{9/10}$, then $\Pr[X_i = 1] \leq n^{-1/10}$. Let Y_i be the indicator of

$$\begin{array}{l}\boldsymbol{t_2 < \infty} \text{ and } \forall \boldsymbol{t} \in [\boldsymbol{t_1}, \boldsymbol{t_1} + \boldsymbol{n^\mu}],\ \boldsymbol{N_{\geq k}(t) \leq n^{9/10}} \text{ and} \\ |\boldsymbol{s(j,t_3)}| \leq \boldsymbol{k} \text{ and } \forall \boldsymbol{t} \in (\boldsymbol{t_3}, \boldsymbol{t_4}],\ |\boldsymbol{s(j,t)}| > \boldsymbol{k} \text{ and } \boldsymbol{w_i} \leq \boldsymbol{t_4}.\end{array}$$

Define a sequence of random variables $Z_1, Z_2, \ldots$ as follows. If $\forall i' \leq i$, $Y_{i'} = 1$, then $Z_i = X_i$; otherwise, $Z_i \in \{-1, +1\}$ is distributed independently of the other Z's with $\Pr[Z_i = 1] = n^{-1/10}$. Let i be the largest such that $w_i \leq t_4$. If $i \leq \log\log n$, then

$$\Pr\left[\sum_{i' \leq i} Z_{i'} > \log\log n\right] = 0.$$

If $i > \log\log n$, we use Lemma A.6 with $p := n^{-1/10}$ to get

$$\Pr\left[\sum_{i' \leq i} Z_{i'} > \log\log n\right] < 2\left(24n^{-1/10}\right)^{i/4} < n^{-\frac{1}{100}\log\log n - 2\mu}.$$

(we use here that n is sufficiently large.)

The use of Fact A.4 completes the proof. ∎

Lemma 4.23 *Suppose that for every positive integer n, k_n is a positive integer and*

$$\Gamma_n(t) \equiv N_{\geq k_n}(t) \leq n^{9/10},$$

$$\Lambda_n(t) \equiv \max(t) \leq k_n + \log\log n$$

then $\Gamma_n \succ \Lambda_n$.

Proof: We use the same arguments that were used in the proofs of Lemmas 4.16 and 4.20. Fix $\lambda > 0$. Take $\kappa = 4$. The probabilities of the bad events are given by Lemmas 4.21 and 4.22. Notice that these lemmas consider a single particle, whereas here we have to consider all n particles (we assume the worst case, that the bad events for the individual particles are mutually disjoint) We get that

$$\begin{aligned}&\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1+n^\mu],\ N_{\geq k}(t) \leq n^{9/10}) \rightarrow \\ (\forall t \in [t_1+n^4, t_1+n^\mu],\ \max(t) \leq k_n + \log\log n)\end{array}\right] \\ > \ & 1 - n\left(2^{-n/2} + n^{-\frac{1}{100}\log\log n}\right) \\ > \ & 1 - n^{-\lambda}.\end{aligned}$$

∎

Lemma 4.24 *Let $c_{init} = 2 \times 10^5$. Let $t_1 \in_U [0,T]$. Then,*

$$\Pr\left[\forall t \in [t_1+n^6, t_1+n^7],\ \max(t) \le c_{init} + 3\log\log n\right] > 1 - n^{-1}.$$

Proof: Let $c = 12$, $\epsilon = 10^{-4}$. Let j be the largest such that $(1000\epsilon)^{2^{(j-1)}}/1000 \ge n^{-1/10}$. Notice that $j < \log\log n$. Let

$$\begin{array}{rcl} \Gamma_n^0(t) & \equiv & \max(t) \le c\log_\alpha n, \\ \Lambda_n^0(t) = \Gamma_n^1(t) & \equiv & N_{\ge c_{init}}(t) \le \epsilon n, \\ \Lambda_n^1(t) = \Gamma_n^2(t) & \equiv & N_{\ge c_{init}+2}(t) \le 1000\epsilon^2 n, \\ \Lambda_n^2(t) = \Gamma_n^3(t) & \equiv & N_{\ge c_{init}+4}(t) \le (1000)^3\epsilon^4 n, \\ & \vdots & \\ \Lambda_n^{j-1}(t) = \Gamma_n^j(t) & \equiv & N_{\ge c_{init}+2(j-1)}(t) \le (1000\epsilon)^{2^{(j-1)}} n/1000, \\ \Lambda_n^j(t) & \equiv & \max(t) \le c_{init} + 2(j-1) + \log\log n \end{array}$$

Then, for $\iota \in \{0,1,\ldots,j\}$, $\Gamma_n^\iota \succ \Lambda_n^\iota$ uniformly in ι with $\kappa = 5$. (This follows by applying Lemma 4.20 for each step in the chain but the last, which instead follows from Lemma 4.22 and the choice of j.) Therefore, by Lemma 4.10, $\Gamma_n^0 \succ \Lambda_n^j$ with $\kappa = 6$. Therefore, for sufficiently large n,

$$\Pr\left[\begin{array}{l} (\forall t \in [t_1, t_1+n^7],\ \max(t) \le c\log_\alpha n) \rightarrow \\ (\forall t \in [t_1+n^6, t_1+n^7],\ \max(t) \le c_{init} + 2(j-1) + \log\log n) \end{array}\right] > 1 - n^{-2}. \tag{8}$$

Let $\mu = 7$, $\lambda = 2$. μ, λ, c satisfy the conditions of Lemma 4.11. Therefore,

$$\Pr\left[\forall t \in [t_1, t_1+n^7],\ \max(t) \le c\log_\alpha n\right] > 1 - n^{-2}. \tag{9}$$

Combining inequalities 8 and 9 gives

$$\Pr\left[\forall t \in [t_1+n^6, t_1+n^7],\ \max(t) \le c_{init} + 2(j-1) + \log\log n\right] > 1 - 2n^{-2} > 1 - n^{-1}.$$

Since $2(j-1) + \log\log n < 3\log\log n$, the lemma follows. ∎

Theorem 4.25 *For sufficiently large n,*

$$E[\max(t)] < 4\log\log n.$$

Proof: Denote by $E_{a,b}$ the expectation of $\max(t)$ when $t \in_U [a,b]$. We bound E_{t_1,t_1+n^7} for $t_1 \in_U [0,T]$. Using Lemma 4.24, $E_{t_1,t_1+n^7} \le n^{-1}E_{t_1,t_1+n^6} + (1-n^{-1})E_{t_1+n^6,t_1+n^7}$. Clearly, $E_{t_1,t_1+n^6} \le n$. Moreover, from Lemma 4.24 we deduce that for sufficiently large n, $E_{t_1+n^6,t_1+n^7} \le 4\log\log n - 1$, which proves the claim. ∎

4.4 Expected Maximal Unfairness is $\Omega(\log\log n)$

The proof of the $\log\log n$ lower bound mirrors the proof of the $\log\log n$ upper bound. Pick a time interval whose length is fixed in advance and is polynomial in n, and whose starting point is any point in the execution. Clearly, the number of people with unfairness ≥ 0 throughout this interval is n. Next we show that with high probability, for each k, τ, any sufficiently long interval throughout which the unfairness of at least τn people is at least k, contains an interval whose starting point is polynomially shifted by a polynomial with a smaller exponent that the original interval, and whose ending point is the same as that of the original interval, and so that throughout the contained interval the unfairness of at least $\tau^2 n$ people is at least $k+1$. The shift does not depend on k, τ. Thus, after $\log\log n$ steps we are left with an interval whose length is close to the length of the original interval we picked, and so that with high probability, throughout this resulting interval, the maximal unfairness is at least $\log\log n$.

We proceed with the actual proof.

Lemma 4.26 *Let $\tau > n^{-1/4}$, $t_1 \in \boldsymbol{Z}^+$, $\lambda > 0$. For every sufficiently large n, for every positive integer k,*

$$\Pr\left[\begin{array}{l} \forall t \in [t_1, t_1+n^4],\ N_{\geq k}(t) \geq \tau n \rightarrow \\ \exists t \in [t_1, t_1+n^4],\ N_{\geq k+1}(t) \geq \frac{1}{16}\tau^2 n \end{array}\right] \geq 1 - n^{-2\lambda}.$$

Proof: Let $X_i = N_{\geq k+1}(t_1+i)$. We wish to compute a lower bound on $E[X_i | X_{i-1}]$. Let us first consider the contribution to the expected change due to decreases in $N_{\geq k+1}$. This value can drop only if one or both of the particles chosen is at $\pm(k+1)$. The probability that the first particle is at $\pm(k+1)$ is at most $\rho_{=k+1}(t_1+i-1)$ and similarly the probability that the second particle is at $\pm(k+1)$ is also at most $\rho_{=k+1}(t_1+i-1)$. So the expected decrease due to particles moving from $\pm(k+1)$ is bounded by $2\rho_{=k+1}(t_1+i-1)$. Next let us consider the contribution to the expected change due to increase in $N_{\geq k+1}$. One of position k or $-k$ contains at least $\frac{1}{2}N_{=k}(t_1+i-1)$ particles. Thus two of these particles are paired with probability at least

$$\binom{\frac{1}{2}N_{=k}(t_1+i-1)}{2} \Big/ \binom{n}{2}$$

which is bounded below by

$$\left(\frac{\rho_{=k}(t_1+i-1)}{2}\right)^2 - \frac{1}{4(n-1)}.$$

Thus we have

$$E[X_i|X_{i-1}] \geq X_{i-1} - 2\rho_{=k+1}(t_1+i-1) + \left(\frac{\rho_{=k}(t_1+i-1)}{2}\right)^2 - \frac{1}{4(n-1)}.$$

If $N_{\geq k}(t_1+i-1) \geq \tau n$ and $X_{i-1} < \tau^2 n/16$, we have that

$$\begin{aligned} \rho_{=k}(t_1+i-1) &= \frac{N_{=k}(t_1+i-1)}{n} \\ &= \frac{N_{\geq k}(t_1+i-1) - N_{\geq k+1}(t_1+i-1)}{n} \\ &= \rho_{\geq k}(t_1+i-1) - \rho_{\geq k+1}(t_1+i-1) \\ &\geq \tau - \frac{1}{8}\tau^2 \\ &\geq \frac{7}{8}\tau, \end{aligned}$$

$$\rho_{=k+1}(t_1+i-1) \leq \rho_{\geq k+1}(t_1+i-1) < \frac{1}{16}\tau^2,$$

and since $\tau > n^{-1/4}$,

$$\frac{1}{4(n-1)} \leq \frac{1}{2n} \leq \frac{1}{256\sqrt{n}} \leq \frac{1}{256}\tau^2$$

(assuming n is sufficiently large.)

Therefore,

$$E[X_i - X_{i-1} \mid X_{i-1}] \geq \frac{1}{16}\tau^2.$$

Let Y_i be the indicator of

$$N_{\geq k}(t_1 + i) \geq \tau n \text{ and } X_i < \tau^2 n/16.$$

Define a sequence of random variables Z_i as follows. If $\forall j < i$, $Y_j = 1$, then $Z_i = X_i$. Otherwise, $Z_i = Z_{i-1} + \tau^2/16$. The Z's satisfy the conditions of Corollary A.7 with $A := X_0$, $B := 1$, $C := \tau^2/16$. Therefore, taking $\delta := 2\sqrt{\lambda \ln n}$,

$$\Pr[Z_{n^4} < X_0 + \frac{1}{16}n^4\tau^2 - 4n^2\sqrt{\lambda \ln n}] < n^{-2\lambda}.$$

Since $X_0 \geq 0$ and $X_{n^4} \leq n$, the lemma follows from Fact A.3. ∎

Lemma 4.27 *Let $t_1 \in \boldsymbol{Z}^+$, n sufficiently large, $\tau > n^{-1/5}$, $\mu > 0$, $k \in \boldsymbol{Z}^+$, $\lambda > 0$. Then*

$$\Pr\left[\begin{array}{l}(\forall t \in [t_1, t_1 + n^\mu],\ N_{\geq k}(t) \geq \tau n)\wedge \\ (\exists t \in [t_1, t_1 + n^4],\ N_{\geq k+1}(t) \geq \frac{1}{16}\tau^2 n) \rightarrow \\ (\forall t \in [t_1 + n^4, t_1 + n^\mu],\ N_{\geq k+1}(t) \geq \frac{1}{32}\tau^2 n)\end{array}\right] > 1 - n^{-2\lambda}.$$

Proof: Let t_2 be the smallest $t \in [t_1, t_1 + n^4]$ such that $N_{\geq k+1}(t) \geq \tau^2 n/16$, or ∞, if no such t exists. We show that

$$\Pr\left[\begin{array}{l}(t_2 < \infty)\wedge \\ (\forall t \in [t_1, t_1 + n^\mu],\ N_{\geq k}(t) \geq \tau n) \rightarrow \\ \forall t \in [t_2, t_1 + n^\mu],\ N_{\geq k+1}(t) \geq \frac{1}{32}\tau^2 n\end{array}\right] > 1 - n^{-2\lambda}, \tag{10}$$

which proves the lemma.

In order to show inequality 10, define random variables $X_i = N_{\geq k+1}(t_2 + i)$. As in the previous lemma, if $N_{\geq k}(t_2 + i - 1) \geq \tau n$ and $X_{i-1} \leq \tau^2 n/16$, we have that

$$E[X_i - X_{i-1} \mid X_{i-1}] \geq \frac{1}{16}\tau^2.$$

Let Y_i be the indicator of

$$t_2 < \infty \text{ and and } N_{\geq k}(t_2 + i) \geq \tau n.$$

Define a sequence of random variables $Z_0, Z_1, \ldots$ as follows. $Z_0 = X_0$. For $i > 0$, if $\forall j < i$, $Y_j = 1$, then $Z_i = X_i$; otherwise, $Z_i = Z_{i-1} + \tau^2/16$. The Z's satisfy the conditions of Lemma A.8 with $\nu := 2(\lambda + \mu)$, $A := \tau^2 n/16$, $B := 1$, $C := \tau^2/16$, $D := \tau^2 n/32$. (we may use here $\sqrt[5]{n} \geq 2048(\lambda + \mu)\ln n$.) Therefore, $\Pr[\exists i \in [n^\mu],\ Z_i < \tau^2 n/32] < n^{-2\lambda}$. Fact A.4 gives the required inequality 10. ∎

Lemma 4.28 *For each positive integer n, let I_n be the set of all pairs $\langle k, \tau\rangle$, where k is a positive integer and $\tau > n^{-1/5}$. Let*

$$\Gamma_n^{\langle k,\tau\rangle}(t) \equiv N_{\geq k}(t) \geq \tau n,$$

$$\Lambda_n^{\langle k,\tau\rangle}(t) \equiv N_{\geq k+1}(t) \geq \frac{1}{32}\tau^2 n,$$

for all $\langle k, \tau\rangle \in I_n$. Then $\Gamma_n^\iota \succ \Lambda_n^\iota$ uniformly in ι

Proof: Take $\kappa = 4$ and use Lemmas 4.26 and 4.27. ∎

Theorem 4.29 *$E[\max(t)] > \log\log n - K$, where K is an absolute constant.*

Proof: Let j be the largest such that $32^{-2^j+1} > n^{-1/5}$. Notice that $j \geq \lfloor \log\log n \rfloor - 5$. Let

$$\begin{aligned}
\Gamma_n^0(t) &\equiv N_{\geq 0}(t) \geq n, \\
\Lambda_n^0(t) = \Gamma_n^1(t) &\equiv N_{\geq 1}(t) \geq 32^{-1}n, \\
\Lambda_n^1(t) = \Gamma_n^2(t) &\equiv N_{\geq 2}(t) \geq 32^{-3}n, \\
\Lambda_n^2(t) = \Gamma_n^3(t) &\equiv N_{\geq 3}(t) \geq 32^{-7}n, \\
&\vdots \\
\Lambda_n^{j-1}(t) = \Gamma_n^j(t) &\equiv N_{\geq j}(t) \geq 32^{-2^j+1}n,
\end{aligned}$$

Using Lemma 4.10, we conclude from Lemma 4.28 that $\Gamma_n^0 \succ \Gamma_n^j$ with $\kappa = 5$. Since $N_{\geq 0}(t)$ is always n, and since $N_{\geq j}(t) \geq 32^{-2^j+1}n$ implies that $N_{\geq j}(t) > 0$, we conclude that for every $\lambda, \mu > 0$, for every $t_1 \in \boldsymbol{Z}^+$, for every sufficiently large n,

$$\Pr\left[\forall t \in [t_1 + n^5, t_1 + n^\mu],\ N_{\geq \lfloor \log\log n \rfloor - 5}(t) > 0\right] \geq 1 - n^{-\lambda},$$

which proves the theorem. ∎

4.5 Convergence to Low Maximal Unfairness

In the previous sections we showed that at a randomly-chosen time in a sufficiently long execution the expected maximal unfairness is small. In this section we consider the question of how quickly the process will converge to low maximal unfairness starting from an arbitrary configuration (with the constraint that this configuration must be reachable from the initial state.) We show that starting at any reachable configuration at time t_1, with high probability the maximal unfairness goes below $\log n$ by time $t_1 + n^4$, and then stays below $2\log n$ for an interval of length at least $n^{\log n}$. (This result in fact implies that the system will quickly converge to a maximum unfairness of $O(\log\log n)$, using the proof of Theorem 4.25 starting with Corollary 4.17.)

The basic idea of the proof is to show that the value of our usual potential function Φ_β is likely to drop when it is too large. We do this in the following technical lemma:

Lemma 4.30 *For all sufficiently small $\epsilon > 0$, there exists a constant $c > 0$ such that for all sufficiently large n and $\beta = 1 + \epsilon$, if $\Phi_\beta = \sum_i \beta^{|s(i)|} > cn$ then the expected value of the change in $\log \Phi_\beta$ is smaller than $-\zeta$, where ζ is a constant greater than zero.*

Proof: Observe that

$$\log \Phi(s(t+1)) - \log \Phi(s(t)) = \log \frac{\Phi(s(t+1))}{\Phi(s(t))} = \log(1 + \frac{\Phi(s(t+1)) - \Phi(s(t))}{\Phi(s(t))}).$$

Thus, the proof of the lemma follows from Lemma 4.12 and the fact that $\ln(1+z) < z$ for $0 < |z| < 1$. ∎

Theorem 4.31 *For all $\lambda > 0$, there exists $\delta > 0$ such that for all sufficiently small ϵ, all sufficiently large n, and any point t_1 in the execution, we have*

$$Pr(\forall t \in [t_1 + n^4, t_1 + n^{\delta \log n}], \Phi_\beta(s(t)) = \sum_i \beta^{|s(i,t)|} \leq cn) > 1 - n^{-\lambda}.$$

where $\beta = 1 + \epsilon$ and c is a constant independent of n.

Proof: Let c be twice the constant from Lemma 4.30. We'll show that Φ_β drops below $cn/2$ with high probability by some $t < t_1 + n^4$ and that once below $cn/2$ it is unlikely to rise above cn for an additional $n^{\delta \log n}$ steps.

Lemma 4.30 implies that until Φ_β is below $cn/2$, the expected change in $\log \Phi_\beta$ is less than $-\zeta$.

Note that 2 is an upper bound on the change of $\log \Phi$. Since $0 < \log \Phi(s(t)) \leq n^2$, applying Corollary A.2 we get that with probability at least $1 - n^{-2\lambda}$, there is some $t < t_1 + n^4$, so that $\Phi_\beta(s(t)) \leq cn/2$.

Let t_2 be the smallest t with this property.

For any fixed t we will denote the event "$\Phi_\beta(s(t)) \geq cn/2$ and $\Phi_\beta(s(t)) < cn$" by $Q(t)$. It is enough to show that for any fixed $t_3, t_4 \in [t_2, t_1 + n^{\delta \log n}]$ we have:

$$Pr((\Phi_\beta(s(t_3)) = \frac{cn}{2} \wedge \forall t \in [t_3, t_4), Q(t)) \rightarrow (\Phi_\beta(s(t_4)) < cn)) > 1 - n^{-2\lambda - 2\delta \log n}. \tag{11}$$

If $t_4 \leq t_3 + n^{3/2}$, it follows from Corollary A.2 that $\Phi_\beta(s(t_4))$ is below $\frac{3}{4}cn$ with high probability. Thus to complete the proof it is enough to consider only the case $t_4 > t_3 + n^{3/2}$.

We will prove (11) in two steps. In the first step we show:

$$Pr((\forall t \in [t_3, t_4), Q(t)) \rightarrow (\forall t \in [t_3 + n^{3/2}, t_4], \max_i |s(i,t)| < 2\log_2 n)) > 1 - n^{-4\lambda - 4\delta \log n}. \tag{12}$$

This will mean that on $[t_3, t_4]$ we have a good upper bound on the changes of Φ_β. In the second step we use use this fact to prove (11) by another application of Corollary A.2.

The proof of (12) is similar to the proof of Lemma 4.22 in that we consider the behavior of each particle individually. First we prove that for each fixed i, with a probability of at least $1 - n^{-n^{1/4}}$, $s(i)$ drops from $\log_\beta(cn) = \bar{c} \log n$ at t_3 to $\log n$ in less than $n^{3/2}$ time units. The reason is that in an interval of this length, i will be paired at least $n^{1/4}$ times with exponentially high probability. As long as $s(i) > \log n$, $\Phi_\beta < cn$ implies that the number of people with unfairness at least $\log n$ can be at most $\frac{cn}{\beta^{\log_2 n}} = n^{1-c_3}$, where $c_3 > 0$ depends only on c and β. Therefore the change of $s(i)$ will be -1 with a probability of at least $1 - n^{-c_3}$. Applying Lemma A.6 we get that $s(i)$ goes down to $\log n$.

Let t_5 be the first time where particle a goes below $\log n$. We now show that it it is likely to remain below $2 \log n$. Specifically:

$$\begin{aligned} Pr((s(i,t_5) = \log n \quad &\wedge \quad \forall t \in [t_3, t_1 + n^{\delta \log n}), Q(t)) \rightarrow (\forall t \in [t_5, t_1 + n^{\delta \log n}], |s(i,t)| < 2 \log n)) \\ &\geq \quad 1 - n^{-4\lambda - 4\delta \log n - 1}. \end{aligned}$$

Let $\Delta_t = |s(i,t)| - |s(i,t-1)|$. The value of Δ_t can be $-1, 0$ or 1. Let W be the set of all $t > t_1$ with $\Delta_t \neq 0$ and let w_j be the j-th element of W so that $w_1 < w_2 < \cdots$ are all of the elements of W in increasing order. Let $X_k = \Delta_{w_k}$. X_k is a random variable which takes only the values $-1, 1$. We define a random variable Z_k in the following way. If $Q(w_k - 1)$ holds then $Z_k = X_k$; if $Q(w_k - 1)$ does not hold then the value of Z_k is chosen at the w_k-th step of the randomization independently of earlier steps and $Z_k = 1$ with a probability of n^{-c_3}, and $Z_k = -1$ with probability $1 - n^{-c_3}$.

Given k, We apply Lemma A.6 to the random variables $Z_1, \ldots, Z_k$ with $p = n^{-c_3}$. The fact that $p_1 \geq 1 - n^{-c_3}$ and the definition of each Z_j implies that the conditions of the lemma are met with $p = n^{-c_3}$.

Clearly, if $k < \log n$ then $P(\sum_{j=1}^{k} Z_j > \log n) = 0$. If $k > \log n$, then applying Lemma A.6 as before we get that

$$P(\sum_{j=1}^{k} Z_j > \log n) < 2(24n^{-c_3})^{i/4} \leq 2(24n^{-c_3})^{\frac{\log n}{4}}.$$

Therefore

$$P(\exists k, \sum_{j=1}^{k} Z_j > \log n) < n^{\delta \log n} 2(24n^{-c_3})^{k/4} \leq n^{\delta \log n} n^{-c' \log n} \leq n^{-4\delta \log n - 4\lambda - 1}.$$

Expanding out the definition of Z_k and Q, this implies (12). ■

Corollary 4.32 *There exists a constant a such that for all $\lambda > 0$, there exists $\delta > 0$ such that for all sufficiently large n, and any point t_1 in the execution,*

$$Pr(\forall t \in [t_1 + n^4, t_1 + n^{\delta \log n}], \max(t) \leq a \log n) > 1 - n^{-\lambda}.$$

Proof: Immediate from the definition of Φ_β. ■

5 Reducing Vector Rounding to the 2-Person Carpool Game

In this section we show that the general carpool problem can be reduced to one where each day only two people arrive, or, equivalently, to the edge orientation game. This is done by a reduction from the still more general *vector rounding problem.* The n-dimensional vector rounding problem is this: the input is a list of vectors $(V_1, V_2, \ldots)$, where each $V_t = (v_t^1, v_t^2, \ldots v_t^n)$ is a vector of length n over the reals. The output is a list of integer vectors $(Z_1, Z_2, \ldots)$ where $Z_t = (z_t^1, z_t^2, \ldots z_t^n)$ is a rounding of V_t that preserves the sum, i.e. for all $1 \leq i \leq n$ we have that $z_t^i \in \{\lfloor v_t^i \rfloor, \lceil v_t^i \rceil\}$ and that

$$\sum_{i=1}^{n} z_t^i \in \{\lfloor \sum_{i=1}^{n} v_t^i \rfloor, \lceil \sum_{i=1}^{n} v_t^i \rceil\}$$

The goal is to make the accumulated difference in each entry as small as possible, i.e. for every t we want $\max_{1 \leq i \leq n} |\sum_{j=1}^{t} z_j^i - \sum_{j=1}^{t} v_j^i|$ to be as small as possible. For input vectors $(V_1, V_2, \ldots)$ and output vectors $(Z_1, Z_2, \ldots)$ the associated cost at time t is

$$\max_{1 \leq i \leq n} |\sum_{j=1}^{t} z_j^i - \sum_{j=1}^{t} v_j^i|.$$

As before, we can consider the off-line problem where we are given the vectors $(V_1, V_2, \ldots)$ ahead of time and the on-line problem where we are given the vectors $(V_1, V_2, \ldots)$ one at a time and have to decide on the corresponding $(Z_1, Z_2, \ldots)$. As in the carpool problem, in the on-line version we consider deterministic algorithms as well as randomized algorithms against the oblivious adversary.

Tijdeman [22] has considered the vector rounding problem and has shown that the off-line version has a solution of difference 1, i.e. for every sequence of real vectors $(V_1, V_2, \ldots)$ there exist integer $(Z_1, Z_2, \ldots)$ such that for all $t \geq 1$ we have $\max_{1 \leq i \leq n} |\sum_{j=1}^{t} z_j^i - \sum_{j=1}^{t} v_j^i| \leq 1$.

One can cast the carpool problem as a vector rounding problem: for a sequence $(X_1, X_2, \ldots)$ create the vectors $(V_1, V_2, \ldots)$ where for all $t \geq 1$ and all $1 \leq i \leq n$ we have

$$v_t^i = \begin{cases} 1/|X_t| & \text{if } i \in X_t \\ 0 & \text{if } i \notin X_t. \end{cases}$$

Therefore if we can connect the performance of the 2-person carpool problem to the vector rounding problem then we will have reduced the general carpool problem to the 2-person problem.

Before we show the reduction we will make some simplifying assumptions, which can be easily justified: We assume that for every t and $1 \leq i \leq n$, v_t^i is non-negative (since we can add the absolute value of $\lceil v_t^i \rceil$ and then subtract it from z_t^i) and that $\sum_{i=1}^n v_t^i$ is an integer (if not, then we can add a "dummy" entry to the vector in order to make the sum an integer; this increases n to $n+1$). Furthermore we assume an a priori bound T on the number of vectors, i.e. $t < T$ (otherwise we will increase T as we go along in multiples of 2).

Our reduction is applicable to both deterministic and randomized algorithms.

Theorem 5.1 *The statement of the result differs slightly depending on the nature of the algorithm used for the two-person carpool problem:*

- *Deterministic Algorithms: Suppose that we have a deterministic algorithm f for the n-participant carpool problem where every day two people show up that maintains unfairness at most $\alpha(n)$, then we can construct a deterministic algorithm f' to the vector rounding problem that maintains an accumulated difference of at most $2\alpha(n)$ for every sequence $\varphi = (V_1, V_2, \ldots)$.*

- *Randomized algorithms against the oblivious adversary: Suppose that we have a randomized algorithm $\tilde{f}$ for the n-participant carpool problem where every day two people show up that maintains unfairness $\alpha(n)$, then we can construct a randomized algorithm $\tilde{f}'$ for the vector rounding problem that maintains an accumulated difference of at most $2\alpha(n)$, for every sequence $\varphi = (V_1, V_2, \ldots)$.*

Proof: The reduction is made by a "scaling" argument, similar in flavor to the bit-by-bit rounding of Beck and Fiala [9, 8]. The constructions of the deterministic and randomized algorithms for the vector rounding problem from the corresponding algorithms for the 2-person carpool problem are similar, only the analysis is a bit different. Consider the binary representation of the v_t^i's. We will make it only $\ell = 2 \log T$ bits long by ignoring the rest of the bits and adjusting one of the v_t^i. This can hardly affect the outcome (by a $\frac{1}{T}$ additive term only, and this can be made arbitrarily small by making ℓ larger). Run ℓ carpool instances simultaneously, one corresponding to each of the ℓ bit positions. For each instance apply the strategy for the carpool problem (f or $\tilde{f}$ depending on the case). Each problem has n participants and the accounting and decisions (but not the inputs!) of each instance are done independently.

We start by describing how the ℓ-th instance is defined and then how the rest of the instances follow. Consider the ℓ-th bits of the entries of V_t. Since $\sum_i v_t^i \in \boldsymbol{Z}^+$, there must be an even number of i's such that the ℓ-th bit of v_t^i is 1. Partition them into pairs arbitrarily and schedule those pairs as requests. If v_t^i and v_t^j are paired, request $\{i, j\}$. Those i's that were chosen to drive by the carpool strategy f or $\tilde{f}$ are rounded up, i.e. we add 2^{-l} to v_t^i. Those i's that were not chosen as drivers are rounded down, i.e. we simply throw away (that is, replace with a 0) the ℓ-th bit of v_t^i. It is easy to verify that this procedure preserves the sum of entries (i.e. $\sum_{i=1}^n v_t^i$) and that the modified V_t requires only $\ell - 1$ bits for its representation. The procedure is repeated now with the

$\ell - 1$st instance and so on. After we do that for all the ℓ carpool instances we are left with an integer V_t which is our Z_t.

How good is this reduction? Let $C(t, j, i, V_1, V_2, \ldots V_t)$ denote the unfairness of the i-th participant at the j-th instance defined by inputs $V_1, V_2, \ldots V_t$. Let $D(t, i, V_1, V_2, \ldots V_t)$ be $\sum_{k=1}^{t}(z_k^i - v_k^i)$ on input $V_1, V_2, \ldots V_t$. We claim that

$$\begin{aligned} & D(t, i, V_1, V_2, \ldots V_t) \\ = & \sum_{j=1}^{\ell} \frac{1}{2^{j-1}} C(t, j, i, V_1, V_2, \ldots V_t) \end{aligned}$$

This can be shown by induction on ℓ. The contribution of the ℓ-th instance is multiplied by $1/2^{\ell-1}$, since we have scaled the ℓ-th instance by $2^{\ell-1}$.

We now turn to the analysis. In the deterministic case we know by assumption on f that $|C(t, j, i, V_1, V_2, \ldots V_t)|$ is bounded by $\alpha(n)$. Therefore,

$$\begin{aligned} & |D(t, i, V_1, V_2, \ldots V_t)| \\ \leq & \sum_{j=1}^{\ell} \frac{1}{2^{j-1}} |C(t, j, i, V_1, V_2, \ldots V_t)| \leq 2\alpha(n) \end{aligned}$$

and therefore $\max_{1 \leq i \leq n} |D(t, i, V_1, V_2, \ldots V_t)| \leq 2\alpha(n)$

For the case of a randomized algorithm, we should first be convinced that the adversary's power is no stronger than that of an oblivious adversary in each of the carpool instances that we have defined. Observe that the inputs to the j instance are determined by $(V_1, V_2, \ldots)$ and the decisions made by the carpool solver on instances $j+1$ through ℓ. The decisions made in instance k, for $1 \leq k \leq j$ at *any* point in time do not effect the inputs to instance j. Therefore, for instance j, the adversary chooses a distribution on $(V_1, V_2, \ldots)$ and can even be given the power to make all the decisions in instances $j+1$ through ℓ and yet all that it would be doing cannot depend on the decision at the j-th instance. Given that we know that

$$\begin{aligned} & \max_{1 \leq i \leq n} |D(t, i, V_1, V_2, \ldots V_t)| \\ \leq & \sum_{j=1}^{\ell} \frac{1}{2^{j-1}} \max_{1 \leq k \leq n} |C(t, j, k, V_1, V_2, \ldots V_t)| \end{aligned}$$

and the expectation of

$$\max_{1 \leq k \leq n} |C(t, j, k, V_1, V_2, \ldots V_t)|$$

is bounded by $\alpha(n)$, we get that the expectation of $\max_{1 \leq i \leq n} |D(t, i, V_1, V_2, \ldots V_t)|$ is at most $2\alpha(n)$. ■

Acknowledgments

We are grateful to Noga Alon, Ron Fagin, Tomás Feder, Alan Frieze, Anna Karlin, Nimrod Megiddo, Babu Narayanan, Gerald Schedler and Joel Spencer for many helpful discussions and remarks.

References

[1] R. Anderson, L. Lovász, P. Shor, J. Spencer, E. Tardos and S. Winograd. Disks, balls, and walls: analysis of a combinatorial game. American Mathematical Monthly **96**, 1989, pp. 481–493.

[2] T. E. Anderson, S. Owicki, J. B. Saxe and C. P. Thacker. High Speed Switch Scheduling for Local Area Networks. *ACM Trans. on Comm. Syst.*, 11(4), 1993, pp. 319–352.

[3] J. Aspnes, Y. Azar, A. Fiat, S. A. Plotkin, and O. Waarts. On-line load balancing with applications to machine scheduling and virtual circuit routing. In *Proc. of the 25th Ann. ACM Symp. on Theory of Computing*, pages 623-631, May 1993.

[4] Y. Azar, A. Broder, and A. Karlin, E. Upfal. Balanced allocations. In *Proc. of the 26th Ann. ACM Symp. on Theory of Computing*, pages 593–602, May 1994.

[5] Y. Azar, J. Naor, and R. Rom. The competitiveness of on-line assignment. In *Proc. of the 3rd Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 203–210, 1992.

[6] Zs. Baranyai. On the factorization of the complete uniform hypergraph. In *Infinite and finite sets*, Colloquia Mathematica Societatis János Bolyai, 1973.

[7] Y. Bartal, A. Fiat, H. Karloff, and R. Vohra. New algorithms for an ancient scheduling problem. In *Proc. of the 24th Ann. ACM Symp. on Theory of Computing*, pages 51–58, 1992.

[8] J. Beck. Balanced two-colorings of finite sets in the cube. *Discrete Math*, 73:13–25, 1988–9.

[9] J. Beck and T. Fiala Integer-making theorems. *Discrete Applied Mathematics*, 3:1–8, 1981.

[10] S. Ben-David, A. Borodin, R.M. Karp, G. Tardos, and A. Wigderson. On the power of randomization in online algorithms. In *Proc. of the 22nd Ann. ACM Symp. on Theory of Computing*, pages 379–386, May 1990.

[11] A. Borodin, N. Linial, and M. Saks. An optimal on-line algorithm for metrical task systems. In *Proc. of the 19th Ann. ACM Symp on Theory of Computing*, pages 373–382, May 1987.

[12] A. Charny. An Algorithm for Rate Allocation in a Packet-Switching Network With Feedback MIT/LCS TR-601, 1994.

[13] D. Coppersmith. Private communication.

[14] E. Davis and J.M. Jaffe. Algorithms for scheduling tasks on unrelated processors. *JACM*, 28:712–736, 1981.

[15] A. Demers, S. Keshav and S. Shenkar Analysis and simulation of a fair queuing algorithm Proc. ACM SIGCOMM 89, pp. 1–12.

[16] R. Fagin and J.H. Williams. A fair carpool scheduling algorithm. In *IBM Journal of Research and Development*, 27(2):133-139, March 1983.

[17] R.L. Graham. Bounds for certain multiprocessing anomalies. *Bell System Technical Journal*, 45:1563–1581, 1966.

[18] E. Hahne. Round-robin scheduling for maxmin fairness in data networks *IEEE J. on Selected areas in Communication, vol 9(7), 1991.*

[19] K. Ramakrishnan and R. Jain. A binary feedback scheme for congestion avoidance in computer networks. *ACM Trans. on Comm. Syst.*, 8(2), 1990, pp. 158–181.

[20] D.D. Sleator and R.E. Tarjan. Amortized efficiency of list update and paging rules. *Communication of the ACM*, 28(2):202–208, 1985.

[21] D. Shmoys, J. Wein, and D. P. Williamson. Scheduling parallel machines on-line. In *Proc. of the 32nd IEEE Ann. Symp. on Foundations of Computer Science*, pages 131–140, 1991.

[22] R. Tijdeman. The chairman assignment problem. *Discrete Math*, 32:323–330, 1980.

[23] A.C. Yao. Probabilistic computation, towards a unified measure of complexity. In *Proc. of the 18th Ann. IEEE Symp. on Foundation of Computer Science*, pages 222–227, 1977.

A Some Consequences of Hoeffding's Inequality

This section describes several extensions of Hoeffding's Inequality for martingales, that will be used in the proofs in Section 4.

Start with Azuma's inequality:

Theorem A.1 (Azuma's inequality) *Let $0 = X_0, \ldots, X_m$ be a martingale with $|X_i - X_{i-1}| \le 1$ for all $i \in [m]$. Let $\delta > 0$ be arbitrary. Then, for all $i \in [m]$,*

$$\Pr[X_i > \delta\sqrt{i}] < e^{-\frac{\delta^2}{2}}.$$

From this it is not difficult to obtain:

Corollary A.2 *Let $X_0, X_1, \ldots, X_m$ be a sequence of random variables that has the following properties:*

1. $X_0 \le A$.
2. $\forall t \in [m]$, $|X_t - X_{t-1}| \le B$.
3. $\forall t \in [m]$, $E[X_t - X_{t-1} \mid X_0, ..., X_{t-1}] \le -C$.

Let $\delta > 0$. Then, for all $i \in [m]$,

$$\Pr[X_i > A - iC + 2B\delta\sqrt{i}] < e^{-\frac{\delta^2}{2}}.$$

Proof: Define

$$R_i = \sum_{t=1}^{i} E[X_t - X_{t-1} \mid X_0, ..., X_{t-1}].$$

Note that R_i is a random variable whose value depends on the values of $X_0, X_1, \ldots, X_{i-1}$. Define a sequence $Y_0, Y_1, \ldots, Y_m$ as follows. $Y_0 = X_0$ and $\forall i \in [m]$, $Y_i = X_i - R_i$. We have that (taking $R_0 = 0$)

$$|Y_i - Y_{i-1}| = |X_i - R_i - X_{i-1} + R_{i-1}| \le B + |E[X_i - X_{i-1} \mid X_0, ..., X_{i-1}]| \le 2B,$$

and

$$\begin{aligned}
E[Y_i \mid Y_0, ..., Y_{i-1}] &= E[(Y_i - Y_{i-1}) + Y_{i-1} \mid Y_0, ..., Y_{i-1}] \\
&= Y_{i-1} + E[(X_i - R_i - X_{i-1} + R_{i-1}) \mid Y_0, ..., Y_{i-1}] \\
&= Y_{i-1} + E[X_i - X_{i-1} \mid Y_0, ..., Y_{i-1}] - E[E[X_i - X_{i-1} \mid Y_{i-1}] \mid Y_0, ..., Y_{i-1}] \\
&= Y_{i-1}.
\end{aligned}$$

Now define $Z_0, Z_1, \ldots, Z_m$ as follows. $\forall i \in \{0\} \cup [m]$, $Z_i = (Y_i - Y_0)/2B$. It is easy to verify that the Z's are a martingale that satisfy the conditions of Azuma's inequality. Therefore, $\forall i \in [m]$,

$$\Pr[Z_i > \delta\sqrt{i}] < e^{-\frac{\delta^2}{2}}.$$

Now, $X_i > A - iC + 2B\delta\sqrt{i} \rightarrow Y_i > A + 2B\delta\sqrt{i} \rightarrow Z_i > \delta\sqrt{i}$. ∎

We will use the following proof paradigms. Let $X_0, X_1, \ldots, X_m$ be nonnegative real valued random variables. Let $Y_0, Y_1, \ldots, Y_m$ be indicator variables. Let $Z_0, Z_1, \ldots, Z_m$ be real valued random variables such that if $\forall j \leq i \; Y_j = 1$, then $Z_i = X_i$. Then, the following facts are trivial:

Fact A.3 *For every predicate P, if not $P(X_0, \ldots, X_m)$ and if $\Pr[P(Z_0, \ldots, Z_m)] > p$, then*

$$\Pr[\exists j \leq m, \; Y_j = 0] > p.$$

Fact A.4 *For every predicate P, if $\Pr[P(Z_0, \ldots, Z_m)] > p$ then,*

$$\Pr[(\forall j \leq m, \; Y_j = 1) \rightarrow P(X_0, \ldots, X_m)] > p.$$

We will also make use of the following consequence of Corollary A.2

Lemma A.5 *Let $\nu, \mu > 0$. Let $A, B, C, D > 0$ such that $2\nu B^2 \ln n / C \leq D$. Let $X_0, X_1, \ldots, X_{n^\mu}$ be a sequence of random variables with the following properties:*

1. *$X_0 \leq A$.*
2. *$\forall i \in [n^\mu]$, $|X_i - X_{i-1}| \leq B$.*
3. *If $X_{i-1} \geq A$ then $E[X_i - X_{i-1} \mid X_0, ..., X_{i-1}] \leq -C$.*

Then,

$$\Pr[\exists i \in [n^\mu], X_i > A + D] < n^{-\nu + 2\mu}.$$

Proof: Consider i, j, $0 < i < j \leq n^\mu$. We show that

$$\Pr[(X_i \leq A \;\wedge\; \forall k \in (i,j), X_k \geq A) \rightarrow (X_j \leq A + D)] > 1 - n^{-\nu}. \tag{13}$$

This proves the lemma, since $(X_t > A + D) \rightarrow (\exists i, j, \; 0 \leq i < j \leq t, \; X_i \leq A \;\wedge\; X_j > A + D \;\wedge\; \forall k \in (i,j), X_k \geq A)$, and there are at most $n^{2\mu}$ pairs i, j that might satisfy this condition for any $t \in [n^\mu]$.

In order to prove inequality 13, let Y_t be the indicator of the event

$$\boldsymbol{X_{i+t} \geq A}.$$

Define a sequence of random variables $Z_0, Z_1, \ldots, Z_{j-i}$ as follows. $Z_0 = X_i$. Let $t \in [1, j-i]$. If $\exists \ell \in [0,t], Y_\ell = 0$, then set $Z_t = Z_{t-1} - C$. Otherwise, set $Z_t = X_{i+t}$. The Z's satisfy the conditions of Corollary A.2 with $A := X_i$, $B := B$, $C := C$. Therefore, we get, for $\delta = \sqrt{2\nu \ln n}$,

$$\Pr[Z_{j-i} < X_i - (j-i)C + 2B\delta\sqrt{j-i}] > 1 - n^{-\nu}.$$

Since $X_i \leq A$, we have that

$$X_i - (j-i)C + 2B\delta\sqrt{j-i} \leq A - (j-i)C + 2B\sqrt{2\nu \ln n}\sqrt{j-i}.$$

The function $f(t) = 2B\sqrt{2\nu \ln n}\sqrt{t} - Ct$ achieves its maximum in the range $[0, \infty)$ either at $t = 0$ or at $t = 2\nu B^2 \ln n / C^2$, which gives $f(t) \leq D$ (using the condition stated in the lemma). This proves inequality 13 because of Fact A.4. ■

We also use the following bound on large deviations.

Lemma A.6 *Assume that i is sufficiently large, $p \leq 1/50$, $Z_1, \ldots, Z_i$ are random variables with values $-1, 1$ only, and for all $j \in \{1, \ldots, i\} \cup \{\delta_1, \ldots, \delta_{j-1}\}$ we have $\Pr[Z_j = 1 | Z_1 = \delta_1, \ldots, Z_{j-1} = \delta_{j-1}] \leq p$. Then*

$$\Pr[(\sum_{j=1}^{i} Z_j) > -i/2] \leq 2(24p)^{i/4}$$

Proof: It is enough to prove the lemma in the case when $Z_1, \ldots, Z_i$ are independent random variables and $\Pr[Z_j = 1] = p$ for $j = 1, \ldots, i$. If $\sum_{j=1}^{i} Z_j > -i/2$ then among the variables $Z_1, \ldots, Z_i$ at least $i/4$ takes the value 1. Let p_r be the probability that exactly r of them takes the value 1.

$$p_r = p^r (1-p)^{i-r} \binom{i}{r} \leq p^r \binom{i}{r}.$$

However, $\binom{i}{r} < i^r / r!$, therefore using the assumption $r > i/4$ and the Stirling formula (if r is sufficiently large) we get

$$\binom{i}{r} \leq i^r / ((1/r)(r/e)^r) \leq i^r / ((r/2e)^r) \leq (2ei/r)^r \leq 24^r.$$

That is, $p_r \leq (24p)^r$. The probability in question is

$$\sum_{r=i/4}^{i} p_r \leq \sum_{r=i/4}^{\infty} (24p)^r \leq 2(24p)^{i/4}.$$

■

By symmetry we have the following reverse versions of Corollary A.2 and Lemma A.5:

Corollary A.7 *Let $X_0, X_1, \ldots, X_m$ be a sequence of random variables that has the following properties:*

1. $X_0 \geq A$.
2. $\forall t \in [m]$, $|X_t - X_{t-1}| \leq B$.

3. $\forall t \in [m]$, $E[X_t - X_{t-1} \mid X_{t-1}] \geq C$.

Let $\delta > 0$. *Then, for all* $i \in [m]$,

$$\Pr[X_i < A + iC - 2B\delta\sqrt{i}] < e^{-\frac{\delta^2}{2}}.$$

Lemma A.8 *Let* $\nu, \mu > 0$. *Let* $A, B, C, D > 0$ *such that* $2\nu B^2 \ln n / C \leq D$. *Let* $X_0, X_1, \ldots, X_{n^\mu}$ *be a sequence of random variables with the following properties:*

1. $X_0 \geq A$.

2. $\forall i \in [n^\mu]$, $|X_i - X_{i-1}| \leq B$.

3. *If* $X_{i-1} \leq A$ *then* $E[X_i - X_{i-1} \mid X_{i-1}] \geq C$.

Then,

$$\Pr[\exists i \in [n^\mu], X_i < A - D] < n^{-\nu+2\mu}.$$