

# Fully dynamic maximal matching in $O(\log n)$ update time

Surender Baswana  
Department of CSE,  
I.I.T. Kanpur, India  
sbaswana@cse.iitk.ac.in

Manoj Gupta  
Department of CSE,  
I.I.T. Delhi, India  
gmanoj@cse.iitd.ernet.in

Sandeep Sen  
Department of CSE,  
I.I.T. Delhi, India  
ssen@cse.iitd.ernet.in

October 24, 2018

## Abstract

We present an algorithm for maintaining a maximal matching in a graph under addition and deletion of edges. Our algorithm is randomized and it takes expected amortized  $O(\log n)$  time for each edge update where  $n$  is the number of vertices in the graph. Moreover, for any sequence of  $t$  edge updates, the total time taken by the algorithm is  $O(t \log n + n \log^2 n)$  with high probability.

**Note:** The previous version of this result appeared in SIAM J. Comp., 44(1): 88-113, 2015. However, the analysis presented there for the algorithm was erroneous. This version rectifies this deficiency without any changes in the algorithm while preserving the performance bounds of the original algorithm.

# 1 Introduction

Let  $G = (V, E)$  be an undirected graph on  $n = |V|$  vertices and  $m = |E|$  edges. A matching in  $G$  is a set of edges  $M \subseteq E$  such that no two edges in  $M$  share any vertex. The study of matchings satisfying various properties has remained the focus of graph theory for decades [21]. It is due to the elegant structure of matching that it also appears in various combinatorial optimization problems [11, 20]. A few well studied matching problems are maximum cardinality matching [10, 17, 22, 23], maximum weight matching [12, 18], minimum cost matching (chapter 7, [19]), stable matching [13], popular matching [2]. Among these problems, the maximum cardinality matching problem has been studied most extensively. A matching  $M$  is of maximum cardinality if the number of edges in  $M$  is maximum. A maximum cardinality matching (MCM) is also referred to as a *maximum matching*. A matching is said to be a *maximal* matching if it cannot be strictly contained in any other matching. It is well known that a maximal matching guarantees a 2-approximation of the maximum matching. Though it is quite easy to compute a maximal matching in  $O(m + n)$  time, designing an efficient algorithm for maximum matching has remained a very challenging problem for researchers [10, 22]. The fastest algorithm, till date, for maximum matching runs in  $O(m\sqrt{n})$  time and is due to Micali and Vazirani [22]. In this paper, we address the problem of maintaining a maximal matching in a dynamic graph.

Most of the graph applications in real life deal with graphs that are not static, viz., the graph changes over time caused by deletion and insertion of edges. This has motivated researchers to design efficient algorithms for various graph problems in dynamic environment. An algorithmic graph problem is modeled in the dynamic environment as follows. There is an online sequence of insertion and deletion of edges and the goal is to update the solution of the graph problem after each edge update. A trivial way to achieve this is to run the best static algorithm for this problem after each edge update; clearly this approach is wasteful. The aim of a dynamic graph algorithm is to maintain some clever data structure for the underlying problem such that the time taken to update the solution is much smaller than that of the best static algorithm. There exist many efficient dynamic algorithms for various fundamental problems in graphs [6, 16, 27, 28, 30].

Baswana, Gupta and Sen [5] had presented a fully dynamic algorithm for maximal matching which achieved  $O(\log n)$  expected amortized time per edge insertion or deletion. Moreover, for any sequence of  $t$  edge updates, the total time taken by the algorithm is  $O(t \log n + n \log^2 n)$  with high probability. Their algorithm improved an earlier result of Onak and Rubinfeld [25] who presented a randomized algorithm for maintaining a  $c$ -approximate (for some unspecified large constant  $c$ ) matching in a dynamic graph with  $O(\log^2 n)$  expected amortized time for each edge update.

This algorithm also implied a similar result for maintaining a two approximate vertex cover. It is also used as a basis for maintaining approximate maximum weight matching in a dynamic graph [3].

Unfortunately, the analysis given in [5] has a crucial flaw, viz., the statement of Lemma 4.10 [5] is not true in general. Although the bound on the update time was critically dependent on Lemma 4.10, we have presented an alternate proof of the claimed update time based on an interesting property of the original algorithm that was not reported earlier. For completeness, we have presented the full details of the algorithm of [5] and the new corrected analysis in this paper. This property is interesting in its own right and may have other useful applications of the algorithm.

## 2 An overview

Let  $M$  denote a matching of the given graph at any moment. Every edge of  $M$  is called a *matched* edge and an edge in  $E \setminus M$  is called an *unmatched* edge. For an edge  $(u, v) \in M$ , we define  $u$  to be the *mate* of  $v$  and vice versa. For a vertex  $x$ , if there is an edge incident to it in the matching  $M$ , then  $x$  is a *matched* vertex; otherwise it is *free* or *unmatched*.

In order to maintain a maximal matching, it suffices to ensure that there is no edge  $(u, v)$  in the graph such that both  $u$  and  $v$  are free with respect to the matching  $M$ . From this observation, an obvious approach will be to maintain the information for each vertex whether it is matched or free at any stage. When an edge  $(u, v)$  is inserted, add  $(u, v)$  to the matching if  $u$  and  $v$  are free. For a case when an unmatched edge  $(u, v)$  is deleted, no action is required. Otherwise, for both  $u$  and  $v$  we search their neighborhoods for any free vertex and update the matching accordingly. It follows that each update takes  $O(1)$  computation time except when it involves deletion of a matched edge; in this case the computation time is of the order of the sum of the degrees of the two vertices. So this trivial algorithm is quite efficient for *small* degree vertices, but could be expensive for *large* degree vertices. An alternate approach to handling deletion of a matched edge is to use a simple randomized technique - a vertex  $u$  is matched with a randomly chosen neighbor  $v$ . Following the standard adversarial model, it can be observed that an expected  $\deg(u)/2$  edges incident to  $u$  will be deleted before deleting the matched edge  $(u, v)$ . So the expected amortized cost per edge deletion for  $u$  is roughly  $O\left(\frac{\deg(u)+\deg(v)}{\deg(u)/2}\right)$ . If  $\deg(v) < \deg(u)$ , this cost is  $O(1)$ ; but if  $\deg(v) \gg \deg(u)$ , then it can be as bad as the trivial algorithm. We combine the idea of choosing a random mate and the trivial algorithm suitably as follows. We first present a fully dynamic algorithm which achieves expected amortized  $O(\sqrt{n})$  time per update. We introduce a notion of *ownership* of edges in which we assign an edge to that endpoint which has *higher* degree. We maintain a partition of the set of vertices into two levels : 0 and 1. Level 0 consists of vertices which own *fewer* edges than an appropriate threshold and we handle the updates in level 0 using the trivial algorithm. Level 1 consists of vertices (and their mates) which own *larger* number of edges and we use the idea of random mate to handle their updates. In particular, a vertex chooses a random mate from its set of owned edges which ensures that it selects a neighbor having a lower degree.

A careful analysis of the  $O(\sqrt{n})$  update time algorithm suggests that a *finer* partition of vertices may help in achieving a better update time. This leads to our final algorithm which achieves expected amortized  $O(\log n)$  time per update. More specifically, our algorithm maintains an invariant that can be informally summarized as follows.

*Each vertex tries to rise to a level higher than its current level, if, upon reaching that level, there are sufficiently large number of edges incident on it from lower levels. Once a vertex reaches a new level, it selects a random edge from this set and makes it matched.*

Note that we say that "*a vertex rises*" to indicate that the vertex moves to a higher level and "*a vertex falls*" to indicate that the vertex moves to a lower level. A vertex may also fall to a lower level if the number of edges incident to it decreases. But a vertex only uses its neighborhood information to decide whether to move to higher or lower level. Overall, the vertices use their local information to reach a global equilibrium state in which after each update each vertex is at the *right* level having no incentive to either move above or below its current level.

## 2.1 Organization of the paper

For a gentle exposition of the ideas and techniques, we first describe a simple but less efficient fully dynamic algorithm for maximal matching in Section 4. We present our final fully dynamic algorithm which achieves expected amortized  $O(\log n)$  time per update in Section 5. In Section 6, we illustrate an example of a dynamic graph that establishes the tightness of the approximation factor guaranteed by our algorithm. In the following section, we describe notations and elementary results from the probability theory that we shall use.

### 3 Preliminaries

We shall use  $\mathcal{M}$  to denote the matching maintained by our algorithm at any stage. Our algorithms maintain a partition of the set of vertices among various levels. We shall use  $\text{LEVEL}(u)$  to denote the level of a vertex  $u$ . We define  $\text{LEVEL}(u, v)$  for an edge  $(u, v)$  as  $\max(\text{LEVEL}(u), \text{LEVEL}(v))$ . Our algorithms will ensure that both the endpoints of each matched edge in  $\mathcal{M}$  are present at the same level. So the matching  $\mathcal{M}$  maintained by our algorithms will be a set of tuples as follows.

$$\mathcal{M} = \{(u, v, \ell) \mid u \text{ is matched with } v \text{ at level } \ell\}$$

The analysis of our algorithms will use a basic result about asymmetric random walk as follows.

#### Asymmetric random walk on a line

Consider a particle performing a discrete random walk on a line. In each step, it moves one unit to the right with probability  $p$  or to the left with probability  $q = 1 - p$ . Each move is independent of the moves made in the past. The following lemma holds if  $p > q$ .

**Lemma 3.1** *Suppose the random walk starts at location  $\ell$  units to the right of the origin. Then the probability that it ever reaches origin within  $L$  steps for any given  $L$  is less than  $\left(\frac{q}{p}\right)^\ell$ .*

The proof of Lemma 3.1 is sketched in Appendix. During the analysis of algorithms, we shall use the terminology *very high probability* for those events whose probability is  $1 - n^c$  for some positive constant  $c$ .

### 4 Fully dynamic algorithm with expected amortized $O(\sqrt{n})$ time per update

The algorithm maintains a partition of the set of vertices into two levels - 0 and 1. We now introduce the concept of *ownership* of the edges. Each edge present in the graph will be owned by one or both of its endpoints as follows. If both the endpoints of an edge are at level 0, then it is owned by both of them. Otherwise it will be owned by exactly that endpoint which lies at higher level. If both the endpoints are at level 1, the tie will be broken suitably by the algorithm. As the algorithm proceeds, the vertices will make transition from one level to another and the ownership of edges will also change accordingly. Let  $\mathcal{O}_u$  denote the set of edges owned by  $u$  at any moment of time. Each vertex  $u \in V$  will keep the set  $\mathcal{O}_u$  in a dynamic hash table [26] so that each search or deletion operation on  $\mathcal{O}_u$  can be performed in worst case  $O(1)$  time and each insertion operation can be performed in expected  $O(1)$  time. This hash table is also suitably augmented with a linked list storing  $\mathcal{O}_u$  so that we can retrieve all edges of set  $\mathcal{O}_u$  in  $O(|\mathcal{O}_u|)$  time.

The algorithm maintains the following three invariants after each update.

1. Every vertex at level 1 is matched. Every free vertex at level 0 has all its neighbors matched.
2. Every vertex at level 0 owns less than  $\sqrt{n}$  edges at any moment of time.
3. Both the endpoints of every matched edge are at the same level.

The first invariant implies that the matching  $\mathcal{M}$  maintained is maximal at each stage. A vertex  $u$  is said to be a *dirty* vertex at a moment if at least one of its invariants does not hold. In order to restore the invariants, each dirty vertex might make transition to some new level and do some processing. This processing involves owning or disowning some edges depending upon whether the level of the vertex has risen or fallen. Thereafter, the vertex will execute `RANDOM-SETTLE` or `NAIVE-SETTLE` to *settle down* at its new level. The pseudocode of our algorithms for handling insertion and deletion of an edge is given in Figure 1 and Figure 2.

## Handling insertion of an edge

Let  $(u, v)$  be the edge being inserted. If either  $u$  or  $v$  are at level 1, there is no violation of any invariant. So the only processing that needs to be done is to assign  $(u, v)$  to  $\mathcal{O}_u$  if  $\text{LEVEL}(u) = 1$ , and to  $\mathcal{O}_v$  otherwise. This takes  $O(1)$  time. However, if both  $u$  and  $v$  are at level 0, then we execute `HANDLING-INSERTION` procedure which does the following (see Figure 1).

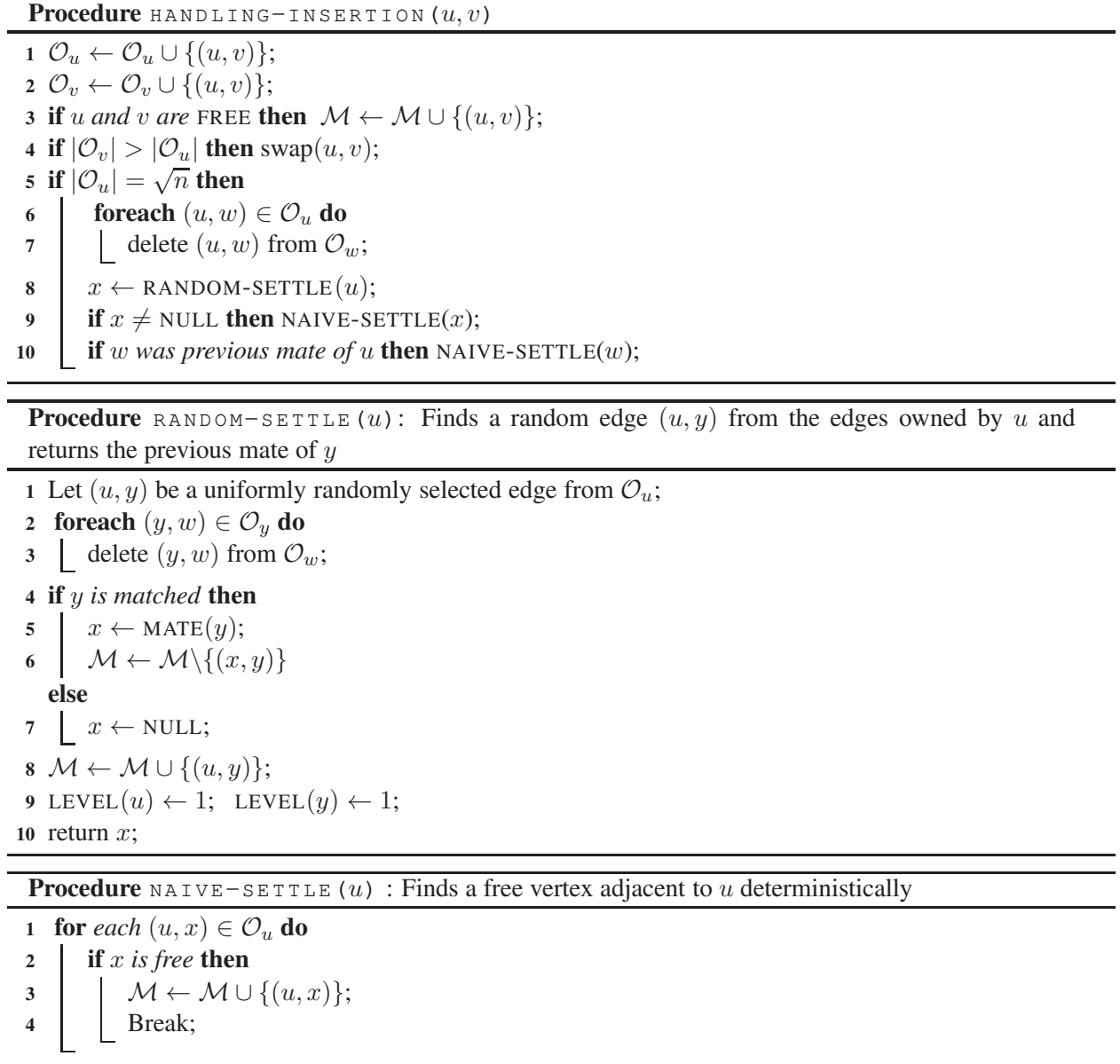


Figure 1: Procedure for handling insertion of an edge  $(u, v)$  where  $\text{LEVEL}(u) = \text{LEVEL}(v) = 0$ .

Both  $u$  and  $v$  become the owner of the edge  $(u, v)$ . If  $u$  and  $v$  are free, then the insertion of  $(u, v)$  has violated the first invariant for  $u$  as well as  $v$ . We restore it by adding  $(u, v)$  to  $\mathcal{M}$ . Note that the insertion of  $(u, v)$  also leads to increase of  $|\mathcal{O}_u|$  and  $|\mathcal{O}_v|$  by one. We process that vertex from  $\{u, v\}$  which owns larger number of edges; let  $u$  be that vertex. If  $|\mathcal{O}_u| = \sqrt{n}$ , then Invariant 2 has got violated. We execute `RANDOM-SETTLE` $(u)$ ; as a result,  $u$  moves to level 1 and gets matched to some vertex, say  $y$ , selected

randomly uniformly from  $\mathcal{O}_u$ . Vertex  $y$  moves to level 1 to satisfy Invariant 3. If  $w$  and  $x$  were respectively the earlier mates of  $u$  and  $y$  at level 0, then the matching of  $u$  with  $y$  has rendered  $w$  and  $x$  free. So to restore Invariant 1, we execute  $\text{NAIVE-SETTLE}(w)$  and  $\text{NAIVE-SETTLE}(x)$ . This finishes the processing of insertion of  $(u, v)$ . Note that when  $u$  rises to level 1,  $|\mathcal{O}_v|$  remains unchanged. Since all the invariants for  $v$  were satisfied before the current edge update, it follows that the second invariant for  $v$  still remains valid.

### Handling deletion of an edge

Let  $(u, v)$  be an edge that is deleted. If  $(u, v) \notin \mathcal{M}$ , all the invariants are still valid. So let us consider the nontrivial case when  $(u, v) \in \mathcal{M}$ . In this case, the deletion of  $(u, v)$  has made  $u$  and  $v$  free. Therefore, potentially the first invariant might have got violated for  $u$  and  $v$ , making them dirty. We do the following processing in this case.

If edge  $(u, v)$  was at level 0, then following the deletion of  $(u, v)$ , vertex  $u$  executes  $\text{NAIVE-SETTLE}(u)$ , and then vertex  $v$  executes  $\text{NAIVE-SETTLE}(v)$ . This restores the first invariant and the vertices  $u$  and  $v$  are *clean* again. If edge  $(u, v)$  was at level 1, then  $u$  is processed using the procedure shown in Figure 2 which does the following ( $v$  is processed similarly).

---

**Procedure**  $\text{HANDLING-DELETION}(u, v)$

---

```

1 foreach  $(u, w) \in \mathcal{O}_u$  and  $\text{LEVEL}(w) = 1$  do
2   | move  $(u, w)$  from  $\mathcal{O}_u$  to  $\mathcal{O}_w$ ;
3 if  $|\mathcal{O}_u| \geq \sqrt{n}$  then
4   |  $x \leftarrow \text{RANDOM-SETTLE}(u)$ ;
5   | if  $x \neq \text{NULL}$  then  $\text{NAIVE-SETTLE}(x)$ ;
   else
6   |  $\text{LEVEL}(u) \leftarrow 0$ ;
7   | foreach  $(u, w) \in \mathcal{O}_u$  and  $\text{LEVEL}(w) = 0$  do
8     | add  $(u, w)$  to  $\mathcal{O}_w$ ;
9   |  $\text{NAIVE-SETTLE}(u)$ ;
10  | foreach  $(u, w) \in \mathcal{O}_u$  do
11  |   | if  $|\mathcal{O}_w| = \sqrt{n}$  then
12  |   |   |  $x \leftarrow \text{RANDOM-SETTLE}(w)$ ;
13  |   |   | if  $x \neq \text{NULL}$  then  $\text{NAIVE-SETTLE}(x)$ ;

```

---

Figure 2: Procedure for processing  $u$  when  $(u, v) \in \mathcal{M}$  is deleted and  $\text{LEVEL}(u)=\text{LEVEL}(v)=1$ .

First,  $u$  disowns all its edges whose other endpoint is at level 1. If  $|\mathcal{O}_u|$  is still greater than or equal to  $\sqrt{n}$ , then  $u$  stays at level 1 and executes  $\text{RANDOM-SETTLE}(u)$ . If  $|\mathcal{O}_u|$  is less than  $\sqrt{n}$ ,  $u$  moves to level 0 and executes  $\text{NAIVE-SETTLE}(u)$ . Note that the transition of  $u$  from level 1 to 0 leads to an increase in the number of edges owned by each of its neighbors at level 0. The second invariant for each such neighbor, say  $w$ , may get violated if  $|\mathcal{O}_w| = \sqrt{n}$ , making  $w$  dirty. So we scan each neighbor of  $u$  sequentially and for each dirty neighbor  $w$  (that is,  $|\mathcal{O}_w| = \sqrt{n}$ ), we execute  $\text{RANDOM-SETTLE}(w)$  to restore the second invariant. This finishes the processing of deletion of  $(u, v)$ .

It can be observed that, unlike insertion of an edge, the deletion of an edge may lead to creation of a large number of dirty vertices. This may happen if the deleted edge is a matched edge at level 1 and at least one of its endpoints move to level 0.

## 4.1 Analysis of the algorithm

While processing the sequence of insertions and deletions of edges, an edge may become matched or unmatched at different update steps. We analyze the algorithm using the concept of *epochs*, which we explain as follows.

**Definition 4.1** *At any time  $t$ , let  $(u, v)$  be any edge in  $\mathcal{M}$ . Then the **epoch** defined by  $(u, v)$  at time  $t$  is the maximal continuous time period containing  $t$  during which it remains in  $\mathcal{M}$ . An epoch is said to belong to level 0 or 1 depending upon the level of the matched edge that defines the epoch.*

The entire life span of an edge  $(u, v)$  consists of a sequence of epochs of  $(u, v)$  separated by the continuous periods when  $(u, v)$  is unmatched. It follows from the algorithm that any edge update that does not change the matching is processed in  $O(1)$  time. An edge update that changes the matching results in the start of new epoch(s) or the termination of some existing epoch(s). For the sake of analysis, we will redistribute the computation performed at any update step  $t$  among the epochs that are created or terminated at step  $t$ . More specifically, let epoch of  $(u_1, v_1), (u_2, v_2), \dots, (u_l, v_l)$  be created and epochs of  $(w_1, x_1), (w_2, x_2), \dots, (w_k, x_k)$  be terminated at step  $t$ . We will redistribute total computation performed at step  $t$  in such a way that:

$$\text{Total computation performed at step } t = \sum_{i=1}^l \text{computation associated with the start of epoch } (u_i, v_i) + \sum_{i=1}^k \text{computation associated with the termination of epoch } (w_i, x_i)$$

Now, we shall analyze the computation involved in each procedure of our algorithm and distribute it suitably among various epochs.

### 1. NAIVE-SETTLE( $u$ )

Observe that whenever the procedure NAIVE-SETTLE( $u$ ) is carried out,  $u$  is present at level 0, and hence  $|\mathcal{O}_u| < \sqrt{n}$ . The procedure NAIVE-SETTLE( $u$ ) searches for a free neighbor of  $u$  by scanning  $\mathcal{O}_u$ . Hence, the time complexity of NAIVE-SETTLE( $u$ ) is  $O(|\mathcal{O}_u|) = O(\sqrt{n})$ . Furthermore, this procedure is called whenever  $u$  loses its mate, say  $v$ . So we can associate the computation cost of NAIVE-SETTLE( $u$ ) with the termination of the previous epoch  $(u, v)$ .

### 2. RANDOM-SETTLE( $u$ )

Observe that whenever RANDOM-SETTLE( $u$ ) is invoked,  $u$  owns at least  $\sqrt{n}$  edges incident from level 0, and hence  $|\mathcal{O}_u| \geq \sqrt{n}$ . During RANDOM-SETTLE( $u$ ),  $u$  finds a random mate from level 0. This is done by selecting a random number  $r \in [1, |\mathcal{O}_u|]$ , and then picking the  $r^{\text{th}}$  edge, say  $(u, y)$ , from the linked list storing  $\mathcal{O}_u$ . This takes  $O(|\mathcal{O}_u|)$  time. Vertex  $u$  then *pulls*  $y$  to level 1 to satisfy the third invariant. In this process,  $y$  becomes the sole owner of all those edges whose other endpoint is at level 0 (line 2,3). Since  $y$  was the owner of at most  $\sqrt{n}$  edges, the total computation time involved in performing this step is  $O(\sqrt{n})$ . Other steps in RANDOM-SETTLE can be executed in  $O(1)$  time. Hence the total computation time is  $O(|\mathcal{O}_u| + \sqrt{n})$  which is  $O(|\mathcal{O}_u|)$  since  $|\mathcal{O}_u| \geq \sqrt{n}$ . We associate this computation time with the start of the epoch  $(u, y)$  that gets created at level 1.

### 3. HANDLING-INSERTION( $u, v$ )

This procedure takes  $O(1)$  time unless one of the endpoints of  $(u, v)$  starts owning  $\sqrt{n}$  edges. In that case, the procedure invokes RANDOM-SETTLE (line 8) and NAIVE-SETTLE (line 9,10). We have already distributed the time taken in these procedures to the respective epochs that get created. Excluding these tasks, the only computation performed in this procedure is in the FOR loop. The purpose of this loop is to make  $u$  the sole owner of all its edges incident from level 0. Since  $u$  owns  $\sqrt{n}$  edges from level 0, the total computation time involved in performing this step is  $O(\sqrt{n})$ . We associate this computation time with the start of the epoch created by  $u$  at level 1.

#### 4. HANDLING-DELETION( $u, v$ )

Procedure HANDLING-DELETION( $u, v$ ) is carried out when the matched edge  $(u, v)$  at level 1 gets deleted. In addition to invoking RANDOM-SETTLE and NAIVE-SETTLE procedures whose computation cost is already assigned to respective epochs, this procedure scans the list  $\mathcal{O}_u$  at most twice. Notice that  $|\mathcal{O}_u|$  can be  $\Theta(n)$ . We associate this computation time of  $O(n)$  with the termination of the epoch  $(u, v)$ .

Excluding the updates that cause the start and termination of an epoch of  $(u, v)$ , every other edge update on  $u$  and  $v$  during the epoch is handled in just  $O(1)$  time. Therefore, we shall focus only on the amount of computation associated with the start and termination of an epoch. Let us now analyze the computation time associated with the epoch at level 0 and level 1.

- Epoch at level 0

As discussed above, it is only the procedure NAIVE-SETTLE whose computation time is associated with an epoch at level 0. This procedure takes  $O(\sqrt{n})$  time. Hence the computation time associated with an epoch at level 0 is  $O(\sqrt{n})$ .

- Epoch at level 1

Consider an epoch at level 1. There are two ways in which this epoch gets created at level 1:

- In HANDLING-INSERTION

An epoch of  $(u, v)$  can be created during the procedure HANDLING-INSERTION( $u, v$ ). In this case, the computation time associated with the start of the epoch of  $(u, v)$  is the computation time incurred in executing the procedure HANDLING-INSERTION and the procedure RANDOM-SETTLE which it invokes. It follows from the above discussion that the computation cost associated with the epoch  $(u, v)$  is  $O(\sqrt{n} + |\mathcal{O}_u|)$  which is  $O(\sqrt{n})$  since  $|\mathcal{O}_u| = \sqrt{n}$  when we invoke HANDLING-INSERTION( $u, v$ ).

- In HANDLING-DELETION

Procedure HANDLING-DELETION( $u, v$ ) invokes RANDOM-SETTLE at lines 4 and 12 to create new epochs at level 1. The execution of RANDOM-SETTLE at line 4 creates a new epoch for  $u$  and its computation time  $O(|\mathcal{O}_u|)$ , which can be  $\Theta(n)$ , gets associated with the start of the new epoch created by  $u$ . The execution of RANDOM-SETTLE at line 12 creates a new epoch for some vertex  $w$  which is some neighbor of  $u$ . Note that  $|\mathcal{O}_w| = \sqrt{n}$ . Its computation time, which is  $O(\sqrt{n})$ , is associated with the start of the epoch at level 1 created by  $w$ .

Now let us calculate the computation cost associated with an epoch, say of an edge  $(u, v)$ , at level 1 when it terminates. It follows from the discussion above that the only computation time associated with the termination of epoch  $(u, v)$  is the computation time of HANDLING-DELETION (excluding the time spent in procedures RANDOM-SETTLE and NAIVE-SETTLE that are already associated with the start of their respective epochs). This cost is at most  $O(n)$ .

From our analysis given above, it follows that the amount of computation time associated with an epoch at level 0 is  $O(\sqrt{n})$  and the computation time associated with an epoch at level 1 is  $O(n)$ .

An epoch of  $(u, v)$  may either terminate (if  $(u, v)$  is removed from the matching) or remain *alive*, i.e.,  $(u, v)$  remain in the matching after the end of all the updates. An epoch of  $(u, v)$ , ends because of exactly one of the following causes.

- (i) if  $(u, v)$  is deleted from the graph.



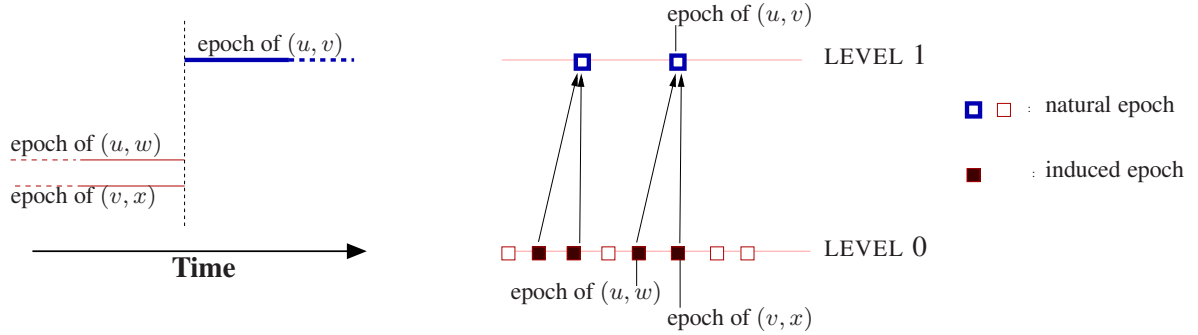


Figure 3: Epochs at level 0 and 1; the creation of an epoch at level 1 can terminate at most two epochs at level 0.

(ii)  $u$  (or  $v$ ) get matched to some other vertex leaving its current mate free.

An epoch will be called a *natural* epoch if it terminates due to cause (i); otherwise it will be called an *induced* epoch. *Induced* epoch can terminate prematurely since, unlike natural epoch, the matched edge is not actually deleted from the graph when an *induced* epoch terminates.

It follows from the algorithm described above that every epoch at level 1 is a natural epoch whereas an epoch at level 0 can be natural or induced depending on the cause of its termination. Furthermore, each induced epoch at level 0 can be associated with a natural epoch at level 1 whose creation led to the termination of the former. In fact, there can be at most two induced epochs at level 0 which can be associated with an epoch at level 1. It can be explained as follows (see Figure 3).

Consider an epoch at level 1 associated with an edge, say  $(u, v)$ . Suppose it was created by vertex  $u$ . If  $u$  was already matched at level 0, let  $w \neq v$  be its mate. Similarly, if  $v$  was also matched already, let  $x \neq u$  be its current mate at level 0. So matching  $u$  to  $v$  terminates the epoch of  $(u, w)$  as well as the epoch of edge  $(v, x)$  at level 0. We *charge* the overall cost of these two epochs to the epoch of  $(u, v)$ . We have seen that the computational cost associated with an epoch at level 0 is  $O(\sqrt{n})$ . So the overall computation *charged* to an epoch of  $(u, v)$  at level 1 is  $O(n + 2\sqrt{n})$  which is  $O(n)$ .

**Lemma 4.1** *The computation charged to a natural epoch at level 1 is  $O(n)$  and the computation charged to a natural epoch at level 0 is  $O(\sqrt{n})$ .*

In order to analyze our algorithm, we just need to get a bound on the computation *charged* to all natural epochs that get terminated during a sequence of updates or are alive at the end of all the updates. Let us first analyze the computation cost *charged* to all those epochs which are alive at the end of  $t$  updates. Consider an epoch of edge  $(u, v)$  that is alive at the end of  $t$  updates. If this epoch is at level 0, the computation cost associated with the start of this epoch is  $O(1)$ . If this epoch is at level 1, then the computation time associated with the start of this epoch is  $O(|\mathcal{O}_u|)$  and notice that  $|\mathcal{O}_u| \geq \sqrt{n}$ . Note that there can be at most two induced epochs at level 0 whose computation time, which is  $O(\sqrt{n})$ , is also charged to the epoch of  $(u, v)$ . Hence the computation charged to the live epoch of  $(u, v)$  is  $O(|\mathcal{O}_u|)$ . Observe that, at any given moment of time,  $\mathcal{O}_u \cap \mathcal{O}_w = \emptyset$  for any two vertices  $u, w$  present at level 1. Hence the computation time charged to all live epochs at the end of  $t$  updates is of the order of  $\sum_u |\mathcal{O}_u| \leq 2t = O(t)$ . So all we need is to analyse the computation charged to all natural epochs that get terminated during the sequence of updates.

Let  $t$  be the total number of updates. Each natural epoch at level 0 which gets terminated can be assigned uniquely to the deletion of its matched edge. Hence it follows from Lemma 4.1 that the computation *charged* to all natural epochs terminated at level 0 during  $t$  updates is  $O(t\sqrt{n})$ . We shall now analyze the number of epochs terminated at level 1. Our analysis will crucially exploit the following lemma.

**Lemma 4.2** *Suppose vertex  $v$  creates an epoch at level 1 during an update in the graph. and let  $O_v^{init}$  be the set of edges that  $v$  owned at the time of the creation of this epoch. Then, for any arbitrary sequence  $D$  of edge deletions of  $O_v^{init}$ , and for any  $(v, w) \in O_v^{init}$*

$$\Pr[\text{MATE}(v) = w \mid D] = \frac{1}{|O_v^{init}|}$$

We first carry out the analysis for the high probability bound on the total update time taken by our algorithm. Thereafter we carry out the analysis for the expected value of the total update time.

## 4.2 High probability bound on the total update time

The key idea of randomization is that once a vertex  $v$  creates an epoch at level 1, there should be *many* edge deletions from  $O_v^{init}$  before the matched edge of  $v$  is deleted. In order to quantify this key idea, we introduce the following definition that categorizes an epoch as good or bad.

**Definition 4.2** *An epoch is said to be bad if it gets terminated naturally with in the deletion of the first  $1/3$  edges that it owned at the time of its creation. An epoch is said to be good if it is not bad.*

It follows from Definition 4.2 that a good epoch undergoes many edge deletions before getting terminated. So only the bad epochs are problematic. Now using Lemma 4.2, we establish an upper bound on the probability of an epoch to be bad.

**Lemma 4.3** *Suppose vertex  $v$  creates an epoch at level 1 during the  $k$ th update for some  $k \leq t$ . Then this epoch is going to be bad with probability  $1/3$  irrespective of the future updates in the graph and the random bits picked during their processing.*

**Proof:** Consider any sequence of updates in the graph following the creation of this epoch. This sequence defines the sequence  $D$  of edge deletions of  $O_v^{init}$ . The termination of this epoch is fully determined by the mate that  $v$  picked and this sequence  $D$ . This epoch will be bad if the mate of  $v$  is among the endpoints of the first  $1/3$  edges in this sequence. Then, Lemma 4.2 implies that the mate of  $v$  is equally likely to be the endpoint of any edge in this sequence. So the probability of the epoch to be bad is  $1/3$ <sup>1</sup>.  $\square$

For the time complexity analysis, we will show that the number of bad epochs may exceed the number of good epochs by at most  $O(\log n)$  with very high probability. Notice that the number of epochs created at level 1 is itself a random variable whose value may depend upon the updates in the graph as well as the random bits picked during their processing. However, as shown by Lemma 4.3, each newly created epoch at level 1 will be bad with probability  $1/3$  irrespective of the past epochs. The number of epochs created at level 1 during any  $t$  updates is trivially  $O(nt)$ . Therefore, the sequence of epochs at level 1 can be seen as an instance of the asymmetric random walk as follows. The walk starts at location  $2 \log_2 n$  to the right of the origin. Each step of the walk is one unit to the right of the current location with probability  $2/3$  or one unit to the left with probability  $1/3$  independent of the past moves. We need to find the probability that the walk ever reaches the origin during any time in the algorithm. It follows from Lemma 3.1 that the probability of this event is less than  $1/n^2$ . So the following lemma holds immediately.

**Lemma 4.4** *During any sequence of  $t$  updates, the number of bad epoch at level 1 can exceed the number of good epochs by  $2 \log_2 n$  with probability at most  $1/n^2$ .*

<sup>1</sup>The analysis assumed that each edge from  $O_v^{init}$  is going to be deleted sometime in future. If not, place all such edges arbitrarily at the end of the sequence  $D$ . In this case, the probability of the epoch to be bad will be even less than  $1/3$ .

As stated in Lemma 4.1, each epoch at level 1 has a computation cost of  $O(n)$  charged to it. Let  $t$  be the total number of updates in the graph. For each epoch at level 1, the number of owned edges at the time of its creation is at least  $\sqrt{n}$ . As a result the number of good epochs during  $t$  updates is bounded by  $4t/\sqrt{n}$  deterministically. So the computation cost of good epochs at level 1 is bounded by  $O(t\sqrt{n})$ . Lemma 4.4 implies that the computation cost of all bad epochs at level 1 can exceed the computation cost of all good epochs at most by  $cn \log n$  amount for some constant  $c$  with probability  $\geq 1 - 1/n^2$ . So overall the cost of all epochs at level 1 is bounded by  $O((t\sqrt{n} + n \log n))$  with high probability. The computation cost of all epochs at level 0 is bounded deterministically by  $O(t\sqrt{n})$ . Hence the total computation time taken by our algorithm for any sequence of  $t$  updates is  $O(t\sqrt{n} + n \log n)$  with high probability.

### 4.3 Expected value of the total update time

Let  $X_{v,i,k}$  be a random variable which is 1 if  $v$  creates an epoch at level  $i$  at update step  $k$ , otherwise it is 0. We denote this epoch as  $\text{EPOCH}(v, i, k)$ . Let  $Z_{v,i,k}$  denote the number of edges from  $\mathcal{O}_v^{init}$  that are deleted during the epoch. (If  $\text{EPOCH}(v, i, k)$  is not created,  $Z_{v,i,k}$  is defined as 0). Since each edge deletion at level 1 is uniquely associated to the epoch that owned it. Therefore,  $\sum_{v,k} Z_{v,1,k} \leq t$ . Hence,

$$\sum_{v,k} \mathbf{E}[Z_{v,1,k}] \leq t \quad (1)$$

We shall now derive a bound on the expected value of  $Z_{v,1,k}$  in an alternate way.

**Lemma 4.5**  $\mathbf{E}[Z_{v,1,k}] \geq \sqrt{n}/2 \cdot \Pr[X_{v,1,k} = 1]$ .

**Proof:** We shall first find the expectation of  $Z_{v,1,k}$  conditioned on the event that  $v$  creates an epoch at level 1 during  $k$ th update. That is, we shall find  $\mathbf{E}[Z_{v,1,k} | X_{v,1,k} = 1]$ . Let  $\mathcal{O}_v^{init}$  be the set of edges owned by  $v$  at the moment of creation of  $\text{EPOCH}(v, 1, k)$ , and let  $D$  be the deletion sequence associated with  $\mathcal{O}_v^{init}$ . It follows from Lemma 4.2 that the matched edge of  $v$  is distributed uniformly over  $\mathcal{O}_v^{init}$ . So  $\mathbf{E}[Z_{v,1,k} | X_{v,1,k} = 1] = |\mathcal{O}_v^{init}|/2 \geq \sqrt{n}/2$  since  $|\mathcal{O}_v^{init}|$  for an epoch at level 1 is at least  $\sqrt{n}$ . Using conditional expectation, we get

$$\mathbf{E}[Z_{v,1,k}] = \mathbf{E}[Z_{v,1,k} | X_{v,1,k} = 1] \cdot \Pr[X_{v,1,k} = 1] \geq \sqrt{n}/2 \cdot \Pr[X_{v,1,k} = 1]$$

□

Notice that the computation cost of an epoch at level 1 is at most  $cn$  for some constant  $c$ . So the expected value of the computation cost associated with all natural epochs that get terminated at level 1 during  $t$  updates is

$$\begin{aligned} \sum_{v,k} cn \cdot \Pr[X_{v,1,k} = 1] &= 2c\sqrt{n} \sum_{v,k} \sqrt{n}/2 \cdot \Pr[X_{v,1,k} = 1] \\ &\leq 2c\sqrt{n} \sum_{v,k} \mathbf{E}[Z_{v,1,k}] \quad \text{using Lemma 4.5} \\ &\leq 2c\sqrt{nt} \quad \text{using Equation 1} \end{aligned}$$

We can thus conclude with the following theorem.

**Theorem 4.1** *Starting with a graph on  $n$  vertices and no edges, we can maintain maximal matching for any sequence of  $t$  updates in  $O(t\sqrt{n})$  time in expectation and  $O(t\sqrt{n} + n \log n)$  with high probability.*

#### 4.4 On improving the update time beyond $O(\sqrt{n})$

In order to extend our 2-LEVEL algorithm for getting a better update time, it is worth exploring the reason underlying  $O(\sqrt{n})$  update time guaranteed by our 2-LEVEL algorithm. For this purpose, let us examine the second invariant more carefully. Let  $\alpha(n)$  be the threshold for the maximum number of edges that a vertex at level 0 can own. Consider an epoch at level 1 associated with some edge, say  $(u, v)$ . The computation associated with this epoch is of the order of the number of edges  $u$  and  $v$  own which can be  $\Theta(n)$  in the worst case. However, the expected duration of the epoch is of the order of the minimum number of edges  $u$  can own at the time of its creation, i.e.,  $\Theta(\alpha(n))$ . Therefore, the expected amortized computation per edge deletion for an epoch at level 1 is  $O(n/\alpha(n))$ . Balancing this with the  $\alpha(n)$  update time at level 0, yields  $\alpha(n) = \sqrt{n}$ .

In order to improve the running time of our algorithm, we need to decrease the ratio between the maximum and the minimum number of edges a vertex can own during an epoch at any level. It is this ratio that determines the expected amortized time of an epoch. This insight motivates us for having a finer partition of vertices – the number of levels should be increased to  $O(\log n)$  instead of just 2. When a vertex creates an epoch at level  $i$ , it will own at least  $2^i$  edges, and during the epoch it will be allowed to own at most  $2^{i+1} - 1$  edges. As soon as it owns  $2^{i+1}$  edges, it should migrate to higher level. Notice that the ratio of maximum to minimum edges owned by a vertex during an epoch gets reduced from  $\sqrt{n}$  to a constant.

We pursue the approach sketched above combined with some additional techniques in the following section. This leads to a fully dynamic algorithm for maximal matching which achieves expected amortized  $O(\log n)$  update time per edge insertion or deletion.

### 5 Fully dynamic algorithm with expected amortized $O(\log n)$ time per update

This algorithm maintain a partition of vertices among various levels. We describe the difference in this partition vis-a-vis 2-LEVEL algorithm.

1. The fully dynamic algorithm maintains a partition of vertices among  $\lfloor \log_4 n \rfloor + 2$  levels. The levels are numbered from  $-1$  to  $L_0 = \lfloor \log_4 n \rfloor$ . During the algorithm, when a vertex moves to level  $i$ , it owns at least  $4^i$  edges. So a vantage point is needed for a vertex that does not own any edge. As a result, we introduce a level -1 that contains all the vertices that do not own any edge.
2. We use the notion of ownership of edges which is slightly different from the one used in the 2-LEVEL algorithm. In the 2-LEVEL algorithm, at level 0, both the endpoints of the edge are the owner of the edge. Here, at every level, each edge is owned by exactly one of its endpoints. If the endpoints of the edge are at different levels, the edge is owned by the endpoint that lies at the higher level. If the two endpoints are at the same level, then the tie is broken appropriately by the algorithm.

Like the 2-LEVEL algorithm, each vertex  $u$  will maintain a dynamic hash table storing the edges  $\mathcal{O}_u$  owned by it. In addition, the generalized fully dynamic algorithm will maintain the following data structure for each vertex  $u$ . For each  $i \geq \text{LEVEL}(u)$ , let  $\mathcal{E}_u^i$  be the set of all those edges incident on  $u$  from vertices at level  $i$  that are not owned by  $u$ . The set  $\mathcal{E}_u^i$  will be maintained in a dynamic hash table. However, the onus of maintaining  $\mathcal{E}_u^i$  will not be on  $u$ . For any edge  $(u, v) \in \mathcal{E}_u^i$ ,  $v$  will be responsible for the maintenance of  $(u, v)$  in  $\mathcal{E}_u^i$  since  $(u, v) \in \mathcal{O}_v$ . For example, suppose vertex  $v$  moves to level  $j$ . If  $j > \text{LEVEL}(u)$ , then  $v$  will remove  $(u, v)$  from  $\mathcal{E}_u^i$  and insert it to  $\mathcal{E}_u^j$ . Otherwise ( $j \leq \text{LEVEL}(u)$ ),  $v$  will remove  $(u, v)$  from  $\mathcal{E}_u^i$  and insert it to  $\mathcal{O}_u$ .

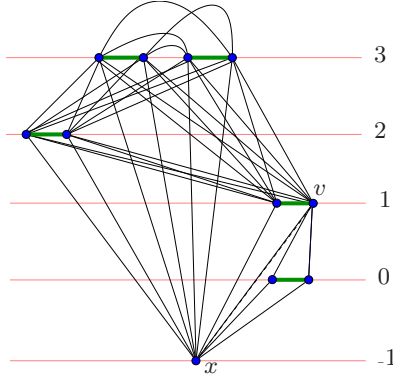


Figure 4: A snapshot of the algorithm on  $K_9$ : all vertices are matched( thick edges) except vertex  $x$  at level  $-1$ .  $\phi_v(2) = 4 < 4^2$  and  $\phi_v(3) = 6 < 4^3$ , so  $v$  cannot rise to a higher level.

### 5.1 Invariants and a basic subroutine used by the algorithm

As can be seen from the 2-level algorithm, it is advantageous for a vertex to get settled at a higher level once it owns a *large* number of edges. Pushing this idea still further, our fully dynamic algorithm will allow a vertex to rise to a higher level if it can own *sufficiently large* number of edges after moving there. In order to formally define this approach, we introduce an important notation here.

For a vertex  $v$  with  $\text{LEVEL}(v) = i$ ,

$$\phi_v(j) = \begin{cases} |\mathcal{O}_v| + \sum_{i \leq k < j} |\mathcal{E}_v^k| & \text{if } j > i \\ 0 & \text{otherwise} \end{cases}$$

In other words, for any vertex  $v$  at level  $i$  and any  $j > i$ ,  $\phi_v(j)$  denote the number of edges which  $v$  can own if  $v$  rises to level  $j$ . Our algorithm will be based on the following key idea. If a vertex  $v$  has  $\phi_v(j) \geq 4^j$ , then  $v$  would rise to the level  $j$ . In case, there are multiple levels to which  $v$  can rise,  $v$  will rise to the highest such level. With this key idea, we now describe the three invariants which our algorithm will maintain.

1. Every vertex at level  $\geq 0$  is matched and every vertex at level  $-1$  is free.
2. For each vertex  $v$  and for all  $j > \text{LEVEL}(v)$ ,  $\phi_v(j) < 4^j$  holds true.
3. Both the endpoints of a matched edge are at the same level.

It follows that the free vertices, if any, will be present at level  $-1$  only. Any vertex  $v$  present at level  $-1$  can not have any neighbor at level  $-1$ . Otherwise, it would imply that  $\phi_v(0) \geq 1 = 4^0$ , violating the second invariant. Hence, every neighbor of a free vertex must be matched. This implies that the algorithm will always maintain a maximal matching. Furthermore, the key idea of our algorithm is captured by the second invariant – after processing every update there is no vertex which fulfills the criteria of rising. Figure 4 depicts a snapshot of the algorithm.

An edge update may lead to the violation of the invariants mentioned above and the algorithm basically restores these invariants. This may involve rise or fall of vertices between levels. Notice that the second invariant of a vertex is influenced by the rise and fall of its neighbors. We now state and prove two lemmas which quantify this influence more precisely.

**Lemma 5.1** *The rise of a vertex  $v$  does not violate the second invariant for any of its neighbors.*

**Proof:** Consider any neighbor  $u$  of  $v$ . Let  $\text{LEVEL}(u) = k$ . Since the second invariant holds true for  $u$  before the rise of  $v$ , so  $\phi_u(i) < 4^i$  for all  $i > k$ . It suffices if we can show that  $\phi_u(i)$  does not increase for any  $i$  due to the rise of  $v$ . We show this as follows.

Let vertex  $v$  rise from level  $j$  to  $\ell$ . If  $\ell \leq k$ , the edge  $(u, v)$  continues to be an element of  $\mathcal{O}_u$ , and so there is no change in  $\phi_u(i)$  for any  $i$ . Let us consider the case when  $\ell > k$ . The rise of  $v$  from  $j$  to  $\ell$  causes removal of  $(u, v)$  from  $\mathcal{O}_u$  (or  $\mathcal{E}_u^j$  if  $j \geq k$ ) and insertion to  $\mathcal{E}_u^\ell$ . As a result  $\phi_u(i)$  decreases by one for each  $i$  in  $[\max(j, k) + 1, \ell]$ , and remains unchanged for all other values of  $i$ .  $\square$

**Lemma 5.2** *Suppose a vertex  $v$  falls from level  $j$  to  $j - 1$ . As a result, for any neighbor  $u$  of  $v$ ,  $\phi_u(i)$  increases by at most 1 for  $i = j$  and remains unchanged for all other values of  $i$ .*

**Proof:** Let  $\text{LEVEL}(u) = k$ . In case  $k \geq j$ , there is no change in  $\phi_u(i)$  for any  $i$  due to fall of  $v$ . So let us consider the case  $j > k$ . In this case, the fall of  $v$  from level  $j$  to  $j - 1$  leads to the insertion of  $(u, v)$  in  $\mathcal{E}_u^{j-1}$  and deletion from  $\mathcal{E}_u^j$ . Consequently,  $\phi_u(i)$  increases by one only for  $i = j$  and remains unchanged for all other values of  $i$ .  $\square$

In order to detect any violation of the second invariant for a vertex  $v$  due to rise or fall of its neighbors, we shall maintain  $\{\phi_v(i) | i \leq L_0\}$  in an array  $\phi_v[]$  of size  $L_0 + 2$ . The updates on this data structure during the algorithm will involve the following two types of operations.

- **DECREMENT- $\phi(v, I)$ :** This operation decrements  $\phi_v(i)$  by one for all  $i$  in interval  $I$ . This operation will be executed when some neighbor of  $v$  rises. For example, suppose some neighbor of  $v$  rises from level  $j$  to  $\ell$ , then  $\phi_v(i)$  decreases by one for all  $i$  in interval  $I = [\max(j, \text{LEVEL}(v)) + 1, \ell]$ .
- **INCREMENT- $\phi(v, i)$ :** this operation increases  $\phi_v(i)$  by one. This operation will be executed when some neighbor of  $v$  falls from  $i$  to  $i - 1$ .

It can be seen that a single **DECREMENT- $\phi(v, I)$**  operation takes  $O(|I|)$  time which is  $O(\log n)$  in the worst case. On the other hand any single **INCREMENT- $\phi(v, i)$**  operation takes  $O(1)$  time. However, since  $\phi_v(i)$  is 0 initially and is non-negative always, we can conclude the following.

**Lemma 5.3** *The computation cost of all **DECREMENT- $\phi()$**  operations over all vertices is upper-bounded by the computation cost of all **INCREMENT- $\phi()$**  operations over all vertices during the algorithm.*

**Observation 5.1** *It follows from Lemma 5.3 that we just need to analyze the computation involving all **INCREMENT- $\phi()$**  operations since the computation involved in **DECREMENT- $\phi()$**  operations is subsumed by the former.*

If any invariant of a vertex, say  $u$ , gets violated, it might rise or fall, though in some cases, it may still remain at the same level. However, in all these cases, eventually the vertex  $u$  will execute the procedure, **GENERIC-RANDOM-SETTLE**, shown in Figure 5. This procedure is essentially a generalized version of **RANDOM-SETTLE( $u$ )** which we used in the 2-level algorithm. **GENERIC-RANDOM-SETTLE( $u, i$ )** starts with moving  $u$  from its current level ( $\text{LEVEL}(u)$ ) to level  $i$ . If level  $i$  is higher than the previous level of  $u$ , then  $u$  performs the following tasks. For each edge  $(u, w)$  already owned by it,  $u$  informs  $w$  about its rise to level  $i$  by updating  $\mathcal{E}_w^i$ . In addition  $u$  acquires the ownership of all the edges whose other endpoint lies at a level  $\in [\text{LEVEL}(u), i - 1]$ . For each such edge  $(u, w)$  that is now owned by  $u$ , we perform **DECREMENT- $\phi(w, [\text{LEVEL}(w) + 1, i])$**  to reflect that the edge is now owned by vertex  $u$  which has moved to level  $i$ . Henceforth, the procedure then resembles **RANDOM-SETTLE**. It finds a random edge  $(u, v)$  from  $\mathcal{O}_u$  and moves  $v$  to level  $i$ . The procedure returns the previous mate of  $v$ , if  $v$  was matched. We can thus state the following lemma.

**Lemma 5.4** Consider a vertex  $u$  that executes `GENERIC-RANDOM-SETTLE( $u, i$ )` and selects a mate  $v$ . Excluding the time spent in `DECREMENT- $\phi$`  operations, the computation time of this procedure is of the order of  $|\mathcal{O}_u| + |\mathcal{O}_v|$  where  $\mathcal{O}_u$  and  $\mathcal{O}_v$  is the set of edges owned by  $u$  and  $v$  just at the end of the procedure.

---

```

Procedure GENERIC-RANDOM-SETTLE( $u, i$ )
1 if LEVEL( $u$ ) <  $i$  then // $u$  rises to level  $i$ 
2   for each  $(u, w) \in \mathcal{O}_u$  do // $u$  informs  $w$  about its rise
3     transfer  $(u, w)$  from  $\mathcal{E}_w^{\text{LEVEL}(u)}$  to  $\mathcal{E}_w^i$ ;
4     DECREMENT- $\phi(w, [\text{LEVEL}(u) + 1, i])$ ;
5   for each  $j = \text{LEVEL}(u)$  to  $i - 1$  do // $u$  gains ownership of some more edges
6     for each  $(u, w) \in \mathcal{E}_u^j$  do
7       transfer  $(u, w)$  from  $\mathcal{E}_u^j$  to  $\mathcal{E}_w^i$ ;
8       transfer  $(u, w)$  from  $\mathcal{O}_w$  to  $\mathcal{O}_u$ ;
9       DECREMENT- $\phi(w, [j + 1, i])$ ;
10  foreach  $j = \text{LEVEL}(u) + 1$  to  $i$  do  $\phi_u(j) \leftarrow 0$ ;
11  LEVEL( $u$ )  $\leftarrow i$ ;
12  Let  $(u, v)$  be a uniformly randomly selected edge from  $\mathcal{O}_u$ ;
13  if  $v$  is matched then
14     $x \leftarrow \text{MATE}(v)$ ;
15     $\mathcal{M} \leftarrow \mathcal{M} \setminus \{(v, x)\}$ ;
16  else
17     $x \leftarrow \text{NULL}$ ;
18  for each  $(v, w) \in \mathcal{O}_v$  do // $v$  informs  $w$  about its rise
19    transfer  $(v, w)$  from  $\mathcal{E}_w^{\text{LEVEL}(v)}$  to  $\mathcal{E}_w^i$ ;
20    DECREMENT- $\phi(w, [\text{LEVEL}(v) + 1, i])$ ;
21  for each  $j = \text{LEVEL}(v)$  to  $i - 1$  do // $v$  gains ownership of some more edges
22    for each  $(v, w) \in \mathcal{E}_v^j$  do
23      transfer  $(v, w)$  from  $\mathcal{E}_v^j$  to  $\mathcal{E}_w^i$ ;
24      transfer  $(v, w)$  from  $\mathcal{O}_w$  to  $\mathcal{O}_v$ ;
25      DECREMENT- $\phi(w, [j + 1, i])$ ;
26   $\mathcal{M} \leftarrow \mathcal{M} \cup \{(u, v)\}$ ;
27  foreach  $j = \text{LEVEL}(v) + 1$  to  $i$  do  $\phi_v(j) \leftarrow 0$ ;
28  LEVEL( $v$ )  $\leftarrow i$ ; /* $v$  rises to level  $i$ */
29  return  $x$ ;

```

---

Figure 5: Procedure used by a free vertex  $u$  to settle at LEVEL  $i$ .

## 5.2 Handling edge updates by the fully dynamic algorithm

Our fully dynamic algorithm will employ a generic procedure called `PROCESS-FREE-VERTICES()`. The input to this procedure is a sequence  $S$  consisting of ordered pairs of the form  $(x, k)$  where  $x$  is a free vertex at level  $k \geq 0$ . Observe that the presence of free vertices at level  $\geq 0$  implies that matching  $\mathcal{M}$

is not necessarily maximal. In order to preserve maximality of matching, the procedure PROCESS-FREE-VERTICES restores the invariants of each such free vertex till  $S$  becomes empty. We now describe our fully dynamic algorithm.

### Handling deletion of an edge

Consider deletion of an edge, say  $(u, v)$ . For each  $j > \max(\text{LEVEL}(u), \text{LEVEL}(v))$ , we decrement  $\phi_u(j)$  and  $\phi_v(j)$  by one. If  $(u, v)$  is an unmatched edge, no invariant gets violated. So we only delete the edge  $(u, v)$  from the data structures of  $u$  and  $v$ . Otherwise, let  $k = \text{LEVEL}(u) = \text{LEVEL}(v)$ . We execute the Procedure PROCESS-FREE-VERTICES( $\langle(u, k), (v, k)\rangle$ ).

### Handling insertion of an edge

Consider insertion of an edge, say  $(u, v)$ . Without loss of generality, assume that initially  $u$  was at the same level as  $v$  or a higher level than  $v$ . So we add  $(u, v)$  to  $\mathcal{O}_u$  and  $\mathcal{E}_v^{\text{LEVEL}(u)}$ . For each  $j > \max(\text{LEVEL}(u), \text{LEVEL}(v))$ , we increment  $\phi_u(j)$  and  $\phi_v(j)$  by one. We check if the second invariant has got violated for either  $u$  or  $v$ . This invariant may get violated for  $u$  (likewise for  $v$ ) if there is any integer  $i > \max(\text{LEVEL}(u), \text{LEVEL}(v))$ , such that  $\phi_u(i)$  has become  $4^i$  just after the insertion of edge  $(u, v)$ . In case there are multiple such integers, let  $i_{\max}$  be the largest such integer. To restore the invariant,  $u$  leaves its current mate, say  $w$ , and rises to level  $i_{\max}$ . We execute GENERIC-RANDOM-SETTLE( $u, i_{\max}$ ), and let  $x$  be the vertex returned by this procedure. Let  $j$  and  $k$  be respectively the levels of  $w$  and  $x$ . Note that  $x$  and  $w$  are two free vertices now. We execute PROCESS-FREE-VERTICES( $\langle(x, k), (w, j)\rangle$ ).

If the insertion of edge  $(u, v)$  violates the second invariant for both  $u$  and  $v$ , we proceed as follows. Let  $j$  be the highest level to which  $u$  can rise after the insertion of  $(u, v)$ , that is,  $\phi_u(j) = 4^j$ . Similarly, let  $\ell$  be the highest level to which  $v$  may rise, that is,  $\phi_v(\ell) = 4^\ell$ . If  $j \geq \ell$ , we allow only  $u$  to rise to level  $j$ ; otherwise we allow only  $v$  to rise to  $\ell$ . Note that after  $u$  moves to level  $j$ , edge  $(u, v)$  becomes an element of  $\mathcal{E}_v^j$ . So  $\sum_{\text{LEVEL}(v) \leq k < \ell} |\mathcal{E}_v^k|$  decreases by 1. As a result,  $\phi_v(\ell) = |\mathcal{O}_v| + \sum_{\text{LEVEL}(v) \leq k < \ell} |\mathcal{E}_v^k|$  also decreases by 1 and is now strictly less than  $4^\ell$ ; thus the second invariant for  $v$  is also restored.

#### 5.2.1 Description of Procedure PROCESS-FREE-VERTICES

The procedure receives a sequence  $S$  of ordered pairs  $(x, i)$  such that  $x$  is a free vertex at level  $i$ . It processes the free vertices in a decreasing order of their levels starting from  $L_0$ . We give an overview of this processing at level  $i$ . For a free vertex at level  $i$ , if it owns *sufficiently* large number of edges, then it settles at level  $i$  and gets matched by selecting a random edge from the edges owned by it. Otherwise the vertex falls down by one level. Notice that the fall of a vertex from level  $i$  to  $i - 1$  may lead to rise of some of its neighbors lying at level  $< i$ . However, as follows from Lemma 5.2, for each such vertex  $v$ , only  $\phi_v(i)$  increases by one and  $\phi_v()$  value for all other levels remains unchanged. So the second invariant may get violated only for  $\phi_v(i)$ . This implies that  $v$  will rise only to level  $i$ . After these rising vertices move to level  $i$  (by executing GENERIC-RANDOM-SETTLE), we move onto level  $i - 1$  and proceed similarly. Overall, the entire process can be seen as a wave of free vertices falling level by level. Eventually this wave of free vertices reaches level  $-1$  and fades away ensuring maximal matching. With this overview, we now describe the procedure in more details and its complete pseudocode is given in Figure 6.

The procedure uses an array  $Q$  of size  $L_0 + 2$ , where  $Q[i]$  is a pointer to a queue (initially empty) corresponding to level  $i$ . For each ordered pair  $(x, k) \in S$ , it inserts  $x$  into queue  $Q[k]$ . The procedure executes a FOR loop from  $L_0$  down to 0 where the  $i$ th iteration extracts and processes the vertices of queue  $Q[i]$  one by one as follows. Let  $v$  be a vertex extracted from  $Q[i]$ . First we execute the function FALLING( $v$ ) which does the following.  $v$  disowns all its edges whose other endpoint lies at level  $i$ . If  $v$  owns less than  $4^i$



---

```

Procedure PROCESS-FREE-VERTICES ( $S$ )
1 for each  $(x, i) \in S$  do ENQUEUE( $Q[i], x$ );
2 for  $i = L_0$  to 0 do
3   while ( $Q[i]$  is not EMPTY) do
4      $v \leftarrow$  DEQUEUE( $Q[i]$ );
5     if FALLING( $v$ ) then                                     //  $v$  falls to  $i-1$ 
6       LEVEL( $v$ )  $\leftarrow$   $i - 1$ ;
7       ENQUEUE( $Q[i - 1], v$ );
8       for each  $(u, v) \in \mathcal{O}_v$  do
9         transfer  $(u, v)$  from  $\mathcal{E}_u^i$  to  $\mathcal{E}_u^{i-1}$ ;
10        INCREMENT- $\phi(u, i)$ ;
11        INCREMENT- $\phi(v, i)$ ;
12        if  $\phi_u(i) \geq 4^i$  then                               //  $u$  rises to  $i$ 
13           $x \leftarrow$  GENERIC-RANDOM-SETTLE( $u, i$ );
14          if  $x \neq$  NULL then
15             $j \leftarrow$  LEVEL( $x$ );
16            ENQUEUE( $Q[j], x$ );
17
18      else                                                     //  $v$  settles at level  $i$ 
19         $x \leftarrow$  GENERIC-RANDOM-SETTLE( $v, i$ );
20        if  $x \neq$  NULL then
21           $j \leftarrow$  LEVEL( $x$ );
22          ENQUEUE( $Q[j], x$ );

```

---

```

Function FALLING ( $v$ )
1  $i \leftarrow$  LEVEL( $v$ );
2 for each  $(u, v) \in \mathcal{O}_v$  such that LEVEL( $u$ ) =  $i$  do         //  $v$  disowns all edges at level  $i$ 
3   transfer  $(u, v)$  from  $\mathcal{O}_v$  to  $\mathcal{O}_u$ ;
4   transfer  $(u, v)$  from  $\mathcal{E}_u^i$  to  $\mathcal{E}_v^i$ ;
5 if  $|\mathcal{O}_v| < 4^i$  then return TRUE else return FALSE;

```

---

Figure 6: Procedure for processing free vertices given as a sequence  $S$  of ordered pairs  $(x, i)$  where  $x$  is a free vertex at LEVEL  $i$ .

edges then  $v$  falls to level  $i - 1$ , otherwise  $v$  will continue to stay at level  $i$ . The processing of the free vertex  $v$  for each of these two cases is done as follows.

1.  $v$  has to stay at level  $i$ .

$v$  executes GENERIC-RANDOM-SETTLE and selects a random mate, say  $w$ , from level  $j < i$  (if  $w$  is present in  $Q[j]$  then it is removed from it and is raised to level  $i$ ). If  $x$  was the previous mate of  $w$ , then  $x$  is a falling vertex. Vertex  $x$  gets added to  $Q[j]$ . This finishes the processing of  $v$ .

2.  $v$  has to fall.

In this case,  $v$  falls to level  $i - 1$  and is inserted to  $Q[i - 1]$ . At this stage,  $\mathcal{O}_v$  consists of neighbors of  $v$  from level  $i - 1$  or below. It follows from Lemma 5.2 that the fall of  $v$  from  $i$  to  $i - 1$  leads to increase in  $\phi_u(i)$  by one for each neighbor  $u$  of  $v$  which is present at a level lower than  $i$ . Moreover,  $\phi_v(i)$ , that was 0 initially, has to be set to  $|\mathcal{O}_v|$ . So all the vertices of  $\mathcal{O}_v$  are scanned, and for each  $(u, v) \in \mathcal{O}_v$ , we increment  $\phi_u(i)$  and  $\phi_v(i)$  by 1. In case  $\phi_u(i)$  has become  $4^i$ ,  $u$  has to rise to level  $i$  and is processed as follows.  $u$  executes GENERIC-RANDOM-SETTLE( $u, i$ ) to select a random mate, say  $w$ , from level  $j < i$ . If  $w$  was in  $Q[j]$  then it is removed from it. If  $x$  was the previous mate of  $w$ , then  $x$  is a falling vertex, and so it gets added to queue  $Q[j]$ .

**Remark 5.1** Notice a stark similarity between the above procedure for handling a free vertex and the procedure for handling a free vertex at level 1 in the 2-level algorithm.

In case 1,  $v$  remains at level  $i$  and  $w$  moves to the level  $i$  from some level  $j < i$ . This renders vertex  $x$  (earlier mate of  $w$ ) free and the first invariant of  $x$  is violated. So  $x$  is added to the queue at level  $j$ . The processing of  $v$  does not change  $\phi_u()$  for any neighbor  $u$  of  $v$ . Furthermore, the rise of  $w$  to level  $i$  does not lead to violation of any invariant due to Lemma 5.1. In case 2,  $v$  falls to level  $i - 1$  and as a result some vertices may rise to level  $i$ . Each such rising vertex executes GENERIC-RANDOM-SETTLE. As in case 1, the processing of these rising vertices may create some free vertices only at level  $< i$ . We can thus state the following lemma.

**Lemma 5.5** After  $i$ th iteration of the for loop of PROCESS-FREE-VERTICES, the free vertices are present only in the queues at level  $< i$ , and for all vertices not belonging to these queues the three invariants holds.

Lemma 5.5 establishes that after termination of procedure PROCESS-FREE-VERTICES, there are no free vertices at level  $\geq 0$  and all the invariants get restored globally.

### 5.3 Analysis of the algorithm

Processing the deletion or insertion of an edge  $(u, v)$  begins with decrementing or incrementing  $\phi_u(i)$  and  $\phi_v(i)$  for each level  $j > \max(\text{LEVEL}(u), \text{LEVEL}(v))$ . Since there are  $O(\log n)$  levels, the computation associated with this task over any sequence of  $t$  updates will be  $O(t \log n)$ . This task may be followed by executing the procedure PROCESS-FREE-VERTICES that restores the invariants and updates the matching accordingly. The updates in the matching can be seen as creation of new epochs and termination of some of the existing epochs. Like 2-level algorithm, for the purpose of analysis, we visualize the entire algorithm as a sequence of creation and termination of various epochs. Excluding the  $O(t \log n)$  time for maintaining  $\phi$ , the total computation performed by the algorithm can be associated with all the epochs that get terminated and those that remain alive at the end of the sequence of updates. Along exactly similar lines as in 2-level algorithm, the computation associated with all the epochs that are alive at the end of  $t$  updates is  $O(t)$  only. So we just need to focus on the epochs that get terminated and the computation associated with each of them.

Let us first analyse the computation associated with an epoch of a matched edge  $(u, v)$ . Suppose this epoch got created by vertex  $v$  at level  $j$ . So  $v$  would have executed GENERIC-RANDOM-SETTLE and selected  $u$  as a random mate from level  $< j$ . Note that  $v$  must be owning less than  $4^{j+1}$  edges and  $u$  would be owning at most  $4^j$  edges at that moment. This observation and Lemma 5.4 imply that the computation involved in the creation of the epoch is  $O(4^j)$ . Once the epoch is created, any update pertaining to  $u$  or  $v$  will be performed in just  $O(1)$  time until the epoch gets terminated. Let us analyze the computation performed when the epoch gets terminated. At this moment either one or both  $u$  and  $v$  become free vertices. If  $v$  becomes free,  $v$  executes the following task (see procedure PROCESS-FREE-VERTICES in Figure 6):  $v$  scans all edges owned by it, which is less than  $4^{j+1}$ , and disowns those edges incident from vertices of level  $j$ . Thereafter, if  $v$  still owns at least  $4^j$  edges, it settles at level  $j$  and creates a new epoch at level  $j$ . Otherwise,  $v$  keeps falling one level at a time. For a single fall of  $v$  from level  $i$  to  $i - 1$ , the computation performed involves the following tasks: scanning the edges owned by  $v$ , disowning those incident from vertices at level  $i$ , incrementing  $\phi_w$  values for each neighbor  $w$  of  $v$  lying at level less than  $i$ , and updating  $\phi_v(i)$  to  $|\mathcal{O}_v|$ . All this computation is of the order of the number of edges  $v$  owns at level  $i$  which is less than  $4^{i+1}$ . Eventually either  $v$  settles at some level  $k \geq 0$  and becomes part of a new epoch or it reaches level  $-1$ . The total computation performed by  $v$  is, therefore, of the order of  $\sum_{i=k}^j 4^{i+1} = O(4^j)$ . This entire computation involving  $v$  (and  $u$ ) in this process is associated with the epoch of  $(u, v)$ . Hence we can state the following Lemma.

**Lemma 5.6** *For any  $i \geq 0$ , the computation associated with an epoch at level  $i$  is  $O(4^i)$ .*

An epoch corresponding to edge  $(u, v)$  at level  $i$  could be terminated if the matched edge  $(u, v)$  gets deleted. Such an epoch is called a natural epoch. However, this epoch could be terminated due to one of the following reasons also.

- $u$  (or  $v$ ) get selected as a random mate by one of their neighbors present at LEVEL  $> i$ .
- $u$  (or  $v$ ) starts owning  $4^{i+1}$  or more edges.

Each of the above factors render the epoch to be an induced epoch. For any level  $i > 0$ , the creation of an epoch causes termination of at most two epochs at levels  $< i$ . It can be explained as follows: Consider an epoch at level  $i$  associated with an edge, say  $(u, v)$ . Suppose it was created by vertex  $u$ . If  $u$  was already matched at some level  $j < i$ , let  $w \neq v$  be its mate. Similarly, if  $v$  was also matched already at some level  $k < i$ , let  $x \neq u$  be its mate. So matching  $u$  to  $v$  terminates the epoch of  $(u, w)$  and  $(v, x)$  at level  $j$  and  $k$  respectively. We can thus state the following lemma.

**Lemma 5.7** *Creation of an epoch at a level  $i$  may cause termination of at most 2 epochs at level  $< i$ .*

### 5.3.1 Analysing an epoch

Consider an epoch created by a vertex  $v$  at level  $i$ . At the time of the creation of the epoch, let  $\mathcal{O}_v^{init}$  be the set of edges owned by  $v$ , and let  $w = \text{MATE}(v)$ . This epoch may terminate much before the deletion of  $(v, w)$ . This happens when  $v$  or  $w$  moves to some level  $> i$  before the deletion of  $(v, w)$ . In order to analyse termination of an epoch, therefore, we associate an update sequence with it as follows. For each edge  $(v, x) \in \mathcal{O}_v^{init}$ , we consider the first time in future that  $x$  moves to some higher level<sup>2</sup>. The *update label* associated with edge  $(v, x)$  is defined as

if  $x$  moves to a level  $> i$  before its deletion then it is classified as *upward* else it is *deletion*.

---

<sup>2</sup>vertex  $x$  may move to level  $> i$  (and down) multiple times while the algorithm processes a sequence of updates. However, it is only the first time (after the creation of the epoch) when  $x$  moves to a level  $> i$  that is relevant as far as the possibility of the termination of the epoch by the upward movement of  $x$  is concerned.

Likewise, we also consider the first time in future that  $v$  moves to a level  $> i$ . If  $v$  never moves to any level  $> i$  in future, we just append  $v$  at the end of all the updates associated with  $\mathcal{O}_v^{init}$ . The update sequence  $U$  for the epoch is the sequence of these updates on the edges of  $\mathcal{O}_v^{init}$  and vertex  $v$  arranged in the chronological order. Consider the following example. Suppose  $\mathcal{O}_v^{init}$  has 10 edges and let the corresponding neighbors of  $v$  be  $\{w_1, \dots, w_{10}\}$ . Let the updates in the chronological order be : the deletion of  $(v, w_4)$ , upward movement of  $w_1$ , upward movement of  $w_9$ , the deletion of  $(v, w_5)$ , and so on. The corresponding update sequence will be

$$U : \langle \overset{\bullet}{w}_4, \overset{\uparrow}{w}_1, \overset{\uparrow}{w}_9, \overset{\bullet}{w}_5, \overset{\bullet}{w}_8, \overset{\bullet}{w}_3, \overset{\uparrow}{w}_2, \overset{\uparrow}{v}, \overset{\bullet}{w}_7, \overset{\uparrow}{w}_{10}, \overset{\bullet}{w}_8 \rangle$$

**Observation 5.2** *If the update associated with (the owner)  $v$  appears at  $\ell$ th location in  $U$ , then the epoch will terminate on or before the  $\ell$ th update in  $U$ . Therefore, the updates at location  $> \ell$  in  $U$  will have no influence on the termination of the epoch.*

Unlike the 2-level algorithm, the update sequence associated with an epoch is not uniquely defined by the sequence of updates in the graph after the creation of the epoch. Rather, it also depends upon the current matching as well as the random bits chosen by the algorithm while processing the updates. So there is a probability distribution defined over all possible update sequences that depends upon these two factors. Consequently, the analysis of an epoch in our final algorithm is more complex compared to the 2-level algorithm. In particular, it is not obvious whether there is any dependence between the random mate picked by a vertex while creating an epoch and the sequence of updates associated with the epoch. However, using an interesting non-trivial property of our algorithm, we will establish that there is no dependence between the two.

**Lemma 5.8** *Suppose a vertex  $v$  creates an epoch and let  $\mathcal{O}_v^{init}$  be the set of its owned edges at the time of the creation of this epoch. Then, for any update sequence  $U$  and for each  $(v, w) \in \mathcal{O}_v^{init}$ ,*

$$\Pr[\text{MATE}(v) = w \mid U] = \Pr[\text{MATE}(v) = w] = \frac{1}{|\mathcal{O}_v^{init}|}$$

Lemma 5.8 can be seen as a generalization of Lemma 4.2 that we stated for our 2-level algorithm. Its proof is given in Section 5.4. The analysis of our algorithm will be critically dependent on this lemma. Using this lemma, we shall first establish a high probability bound on the total update time of the algorithm to process a sequence of updates in the graph.

### 5.3.2 High probability bound on the total update time

Recall Definition 4.2 of a bad epoch. It can be observed from this definition that an induced epoch is always a good epoch. Using Lemma 5.8, the lemma for the probability of a bad epoch extends seamlessly from 2-level algorithm to our final algorithm as follows.

**Lemma 5.9** *Suppose vertex  $v$  creates an epoch at level  $i$  while the algorithm processes  $k$ th update in the graph. This epoch will be bad with probability at most  $1/3$  irrespective of the updates in the graph and the random bits picked during their processing.*

**Proof:** The termination of the epoch is completely determined by the mate that  $v$  picks and the update sequence associated with this epoch. Consider any update sequence  $U$  associated with this epoch. It follows from Lemma 5.8 that conditioned on  $U$ , the mate of  $v$  is equally likely to be the endpoint of any edge in

$\mathcal{O}_v^{init}$ . Now recall from Observation 5.2 that for the epoch to be terminated naturally, the mate of  $v$  must be among the endpoints of the deleted edges that precede  $v$  in  $U$ . We distinguish between the following two cases.

Case1. There are less than  $|\mathcal{O}_v^{init}|/3$  edge deletions preceding  $v$  in  $U$ .

The epoch will be bad only if the matched edge of  $v$  is one of these edge deletions preceding  $v$  in  $U$ . Since the number of these edge deletions is less than  $|\mathcal{O}_v^{init}|/3$ , so using Lemma 5.9 the probability of the epoch to be bad is less than  $1/3$ .

Case2. There are at least  $|\mathcal{O}_v^{init}|/3$  edge deletions preceding  $v$  in  $U$ .

The epoch will be bad if the matched edge of  $v$  is one of the first  $|\mathcal{O}_v^{init}|/3$  edge deletions in  $U$ . From Lemma 5.8, the termination of the epoch is equiprobable for any of the  $\mathcal{O}_v^{init}$ 's, so the probability that the epoch is bad in this case is exactly  $1/3$ .

It follows that the epoch is going to be bad with probability  $1/3$  for each possible update sequence  $U$  associated with the epoch.  $\square$

We will show that the number of bad epochs at a level  $i$  could exceed the number of good epochs at level  $i$  by at most  $O(\log n)$  with very high probability. Notice that the number of epochs created at level  $i$  is itself a random variable. During any update, the number of epochs that will be created at level  $i$  depends upon the past updates in the graph and the random bits picked during their processing. However, Lemma 5.9 implies that each newly created epoch at level  $i$  will be bad with probability at most  $1/3$  independent of these events. Hence, the sequence of epochs at level  $i$  can be seen as an instance of the asymmetric random walk as established in the analysis of the 2-level algorithm. So the bad epochs at any level  $i$  may exceed the good epochs by  $2\log_2 n$  with probability at most  $1/n^2$ . There are  $O(\log n)$  levels in the hierarchy. Hence we get the following lemma using *union bound*.

**Lemma 5.10** *For every level  $i \leq L_0$ , the number of bad epochs will not exceed the number of good epochs by more than  $2\log_2 n$  with probability at least  $1 - (\log n)/n^2 > 1 - 1/n$ .*

Let us temporarily exclude the maximum surplus of  $O(\log n)$  bad epochs at each level from our analysis. Consequently, it follows from Lemma 5.10 that each bad epoch at a level can be mapped to a good epoch at the same level in a unique manner - see Figure 7(i). Also the creation of each epoch at a level  $i + 1$  can terminate at most two (induced) epochs at lower levels as stated in Lemma 5.7. Using this fact and the mapping between the good and bad epochs at a level, we can construct a forest whose nodes will be the epochs terminated across all levels during the algorithm. The intuition for defining this forest is that eventually the computation cost of a bad epoch or an induced epoch will be charged to a good natural epoch. Since a good natural epoch has sufficiently *large* number of edge deletions associated with it, these edge deletions can be charged to pay for all the computation carried out by our algorithm.

With this intuition, we now provide the construction of the forest by defining parent of each epoch using the following rules.

1. Parent of each induced epoch is the epoch at the higher level whose creation led to its termination.
2. Parent of a good epoch is itself (hence it is the root of its tree).
3. If a bad epoch is mapped to an induced epoch, then its parent is the same as the parent of the induced epoch. Otherwise, it is the parent of itself (hence it is the root of its tree).

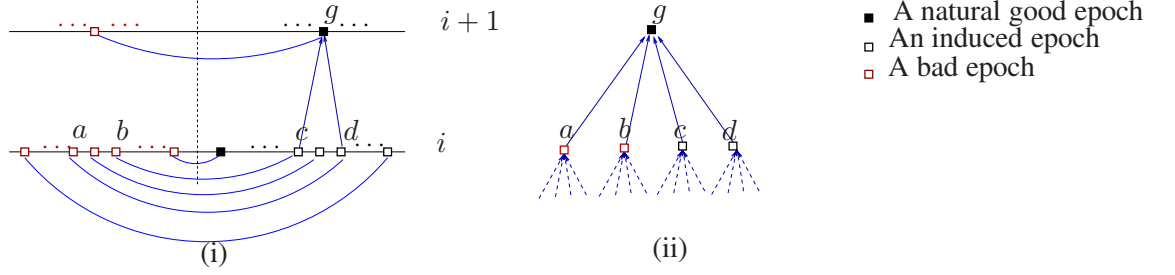


Figure 7: (i) Mapping between bad and good epochs at level  $i$  (ii) Assigning at most 4 epochs from lower levels to an epoch.

It follows from rule 1 and 3 (the if part) that with an epoch at a level, at most 4 epochs from lower levels can be associated. Hence each node in the forest will have at most four children. See Figure 7(ii). Moreover, the root of each tree in the forest of epochs is either a bad epoch or a good natural epoch. Using Lemma 5.6, the computation cost  $C(i)$  associated with a tree of epochs whose root is at level  $i$  obeys the following recurrence for some constant  $a$ .

$$C(i) = a4^i + 4C(i-1)$$

The solution of this recurrence is  $C(i) = O(i4^i)$ . It follows from Lemma 5.10 and rule 3 (Otherwise part) that the trees rooted at good natural epochs at a level  $i$  are at least the number of trees rooted at bad epochs at level  $i$ . Hence, it suffices to analyze the computation cost associated with all the tree rooted at good natural epochs. Now for each good natural epoch at a level  $i$ , there are at least  $4^i/3$  edge deletions associated uniquely to it. This natural epoch will be charged for the computation cost  $C(i) = O(i4^i)$  associated with the tree rooted at it. So if  $t$  is the total number of updates in the graph, then the computation cost associated with all epochs in the forest is  $O(t \log n)$ . The computation cost associated with surplus bad epochs at all levels is  $O(\sum_i i4^i \log n) = O(n \log^2 n)$ . Hence with high probability the computation cost for processing  $t$  edge updates by the algorithm is  $O(t \log n + n \log^2 n)$ . This also implies that the total expected update time is  $O(t \log n)$  for  $t = \Omega(n \log n)$ . In the following subsection, we will establish  $O(t \log n)$  bound on the expected update time for all values of  $t$ .

### 5.3.3 Expected value of the total update time

During a sequence of  $t$  updates in the graph, various epochs get created by various vertices at various levels. Let  $X_{v,i,k}$  be a random variable which is 1 if  $v$  creates an epoch at level  $i$  at update step  $k$ , otherwise it is 0. We denote this epoch as  $\text{EPOCH}(v, i, k)$ . Let  $\mathcal{O}_v^{init}$  denote the edges that  $v$  owned at the time of the creation of the epoch. Let  $Z_{v,i,k}$  denote the number of edges from  $\mathcal{O}_v^{init}$  that are deleted during the epoch. (If  $\text{EPOCH}(v, i, k)$  is not created,  $Z_{v,i,k}$  is defined as 0). The key role in bounding the expected running time is played by a random variable  $B_{v,i,k}$  defined as follows:

$$B_{v,i,k} = \begin{cases} (8Z_{v,i,k} - 2 \cdot 4^i)X_{v,i,k} & \text{if } \text{EPOCH}(v, i, k) \text{ is natural} \\ (4^{i+1} - 2 \cdot 4^i)X_{v,i,k} & \text{if } \text{EPOCH}(v, i, k) \text{ is induced} \end{cases} \quad (2)$$

First observe that  $B_{v,i,k} = 0$  if  $X_{v,i,k} = 0$ . Else (if  $X_{v,i,k} = 1$ ), the random variable  $B_{v,i,k}$  can be seen as credits associated with  $\text{EPOCH}(v, i, k)$  to be used for paying its computation cost. For a natural epoch, the credits is defined in terms of the edges deleted during the epoch. So we define  $B_{v,i,k}$  to be  $8Z_{v,i,k}$ . However, we need to discount for the two epochs at lower levels that may get terminated due to  $\text{EPOCH}(v, i, k)$ . To this end, from the term, we deduct  $2 \cdot 4^i$ . Similarly, if  $\text{EPOCH}(v, i, k)$  is an induced epoch, then it gets  $4^{i+1}$  credits from the epoch that destroyed it. But here again we need to discount for the two epochs at lower

levels that might be terminated by it. To this end, we again deduct  $2 \cdot 4^i$  from  $4^{i+1}$ . The following lemma gives a bound on  $\sum B_{v,i,k}$ .

**Lemma 5.11**  $\sum_{v,i,k} B_{v,i,k} \leq 8 \sum_{v,i,k} Z_{v,i,k} \leq 8t$ , where  $t$  is the total number of updates in the graph.

**Proof:** We need to analyze the sum of  $B_{v,i,k}$ 's for all those  $(i, v, k)$  values for which the  $\text{EPOCH}(i, v, k)$  got created. If this epoch is an induced epoch, it can be associated with an epoch, say  $\text{EPOCH}(v', i', k')$ , at a higher level  $i' > i$  whose creation destroyed it. Notice that the negative  $4^{i'}$  term in  $B_{v',i',k'}$  cancels out the positive  $4^{i+1}$  term in  $B_{v,i,k}$ . Hence, the contribution of induced epochs in  $\sum_{v,i,k} B_{v,i,k}$  is nullified and all that remains is the sum of terms  $8Z_{v,i,k}$  for each natural epoch. Hence  $\sum_{v,i,k} B_{v,i,k} \leq 8 \sum_{v,i,k} Z_{v,i,k}$ . An edge deletion is associated with an epoch in a unique manner, so will contribute to exactly one  $Z_{v,i,k}$ . Therefore,  $\sum_{v,i,k} Z_{v,i,k}$  is upper bounded by the total number of edges deleted.  $\square$

**Corollary 5.0.1**  $\sum_{v,i,k} \mathbf{E}[B_{v,i,k}] \leq 8t$

**Lemma 5.12** For all  $i, v, k$ ,  $\mathbf{E}[B_{v,i,k}] \geq \Pr[X_{v,i,k} = 1] \cdot 4^i$ .

**Proof:** Since  $X_{v,i,k}$  is an indicator random variable,  $\mathbf{E}[B_{v,i,k}] = \Pr[X_{v,i,k} = 1] \mathbf{E}[B_{v,i,k} | X_{v,i,k} = 1]$ . We will first estimate  $\mathbf{E}[B_{v,i,k} | X_{v,i,k} = 1]$ , that is, the expected value of  $B_{v,i,k}$  given that  $\text{EPOCH}(v, i, k)$  got created.

Let  $(\mathcal{U}, P)$  be the probability space of all the update sequences associated with this epoch and let  $U \in \mathcal{U}$  be any update sequence. Suppose among the updates in  $U$  that precede the update associated with  $v$ , only  $d$  are edge deletions. It follows from Lemma 5.8 that the matched edge of  $v$  is distributed uniformly over  $\mathcal{O}_v^{init}$ . So  $\text{EPOCH}(v, i, k)$  will be an induced epoch with probability  $(|\mathcal{O}_v^{init}| - d)/|\mathcal{O}_v^{init}|$  and in that case  $B(v, i, k)$  will be  $4^{i+1} - 2 \cdot 4^i$ . If the epoch is natural, it could be due to any one of the  $d$  edge deletions present in  $U$ . In that case the expected value of  $B_{v,i,k}$  will be  $1/d \sum_{j=1}^d (8j - 2 \cdot 4^i) \geq 4d - 2 \cdot 4^i$ . Considering the cases of induced and natural epoch together,

$$\begin{aligned} \mathbf{E}[B_{v,i,k} | U] &= \frac{|\mathcal{O}_v^{init}| - d}{|\mathcal{O}_v^{init}|} (4^{i+1} - 2 \cdot 4^i) + \frac{d}{|\mathcal{O}_v^{init}|} (4d - 2 \cdot 4^i) = 2 \cdot 4^i - \frac{4^{i+1}d - 4d^2}{|\mathcal{O}_v^{init}|} \\ &\geq 2 \cdot 4^i - \frac{4^i \cdot 4^i}{|\mathcal{O}_v^{init}|} \text{ (for all values of } d) \end{aligned}$$

Therefore

$$\mathbf{E}[B_{v,i,k}] = \sum_{U \in \mathcal{U}} \mathbf{E}[B_{v,i,k} | U] \cdot \Pr[U] \geq \left( 2 \cdot 4^i - \frac{4^i \cdot 4^i}{|\mathcal{O}_v^{init}|} \right) \cdot \sum_{U \in \mathcal{U}} \Pr[U] = 2 \cdot 4^i - \frac{4^i \cdot 4^i}{|\mathcal{O}_v^{init}|}$$

Since  $|\mathcal{O}_v^{init}| \geq 4^i$ , for level  $i$ , the result follows.  $\square$

Let  $W_{v,i,k}$  be a random variable that corresponds to the value of the computation cost of  $\text{EPOCH}(v, i, k)$  if the epoch is created and is 0 otherwise. Notice that the computation cost of an epoch at level  $i$  is  $c4^{i+1}$  for some constant  $c$ . So,  $\mathbf{E}[W_{v,i,k}] = \Pr[X_{v,i,k} = 1]c4^{i+1}$ . Therefore, using Lemma 5.12,

$$\mathbf{E}[W_{v,i,k}] \leq 4c\mathbf{E}[B_{v,i,k}] \tag{3}$$

Using the above equation and Corollary 5.0.1, the total expected computation cost associated with all epochs that get destroyed during the algorithm can be bounded by  $O(t)$  as follows.

$$\sum_{v,i,k} \mathbf{E}[W_{v,i,k}] \leq \sum_{v,i,k} 4c\mathbf{E}[B_{v,i,k}] \leq 32ct = O(t)$$

Since for each update in the graph, we incur  $O(\log n)$  time to update  $\phi$  at various levels, there is an  $O(t \log n)$  overhead for  $t$  updates. We can thus conclude with the following theorem.

**Theorem 5.1** *Starting with a graph on  $n$  vertices and no edges, we can maintain a maximal matching for any sequence of  $t$  updates in  $O(t \log n)$  time in expectation and  $O(t \log n + n \log^2 n)$  with high probability.*

## 5.4 Proof of Lemma 5.8

Our algorithm use randomization to maintain maximal matching. After any given sequence of updates, there is a set of possible maximal matchings that the algorithm may be maintaining and there is probability distribution associated with these maximal matchings. So it is useful to think about the probability space of these matchings as the algorithm proceeds while processing a sequence of updates.

We introduce some notations first. For any matching  $\mathcal{M}$  maintained at any stage by our algorithm, let  $\mathcal{M}_i$  denote the matching at level  $i$ . Let  $\mathcal{M}_{>i} = \cup_{j>i} \mathcal{M}_j$  denote the matchings at all levels  $> i$ . Let  $V_i$  denote the set of all the vertices belonging to levels in the range  $\in [-1, i]$ . We now extend the notations to incorporate the updates in the graph. For any  $k \geq 1$ , let  $G(k)$  denote the graph after a given sequence of  $k$  updates and let  $\mathcal{M}(k)$  denote the maximal matching of  $G(k)$  as maintained by our algorithm. Let  $\mathcal{M}_{>i}(k)$  denote the matching at all levels  $> i$  after a given sequence of  $k$  updates.

After processing certain number of updates by the algorithm, suppose  $M$  and  $M'$  are any two matchings possible such that  $M_{>i} = M'_{>i}$ . Consider any single update in the graph at this stage. In order to process it, suppose we carry out two executions  $I$  and  $I'$  of our algorithm with the initial matching being  $M$  and  $M'$  respectively. That is,  $\mathcal{M}(0) = M$  in the execution  $I$  and  $\mathcal{M}(0) = M'$  in the execution  $I'$ . Our claim is that the probability distribution of matching at levels  $> i$  will be identical at the end of both the executions. More precisely, for any maximal matching  $\mu(1)$  on a subset of vertices in graph  $G(1)$ ,

$$\Pr[\mathcal{M}_{>i}(1) = \mu(1) | \mathcal{M}(0) = M] = \Pr[\mathcal{M}_{>i}(1) = \mu(1) | \mathcal{M}(0) = M']$$

In order to establish our claim, we shall crucially exploit the following lemma.

**Lemma 5.13** *For both the matchings  $M$  and  $M'$ ,  $\phi_v(j)$  is the same for each  $v \in V$  and  $j > i$ .*

**Proof:** It is given that  $M_{>i} = M'_{>i}$ . This implies that for each level  $j > i$  the sets of vertices present are identical in  $M$  and  $M'$ . Hence the set  $V_i$  of all the vertices present at levels  $\in [-1, i]$  is identical in  $M$  and  $M'$ . Hence for any vertex  $v$ , and any level  $j > i$ , the set of all the neighbours of  $v$  at levels  $< j$  is identical; notice that  $\phi_v(j)$  is just the cardinality of this set. So it follows that  $\phi_v(j)$  is the same for each vertex  $v$  and each  $j > i$ .  $\square$

We shall now establish our claim for the deletion of an edge  $e = (u, v)$ . Establishing the claim for the insertion of an edge is similar. Notice that our algorithm does not alter the matching if  $e$  is not a matched edge. If  $e$  is a matched edge, a wave of free vertices originates from  $\text{LEVEL}(e)$  and propagates downward. The following fact follows from our analysis in Section 5.2.1.

- F1.* The algorithm won't alter the matching at level  $> \text{LEVEL}(e)$  while processing the deletion of  $e$ .
- F2.* The matching is updated in the decreasing order of levels, and once the updating of the matching at a level is complete, the matching at that level will remain unchanged during the updates of the matching at lower levels.



It follows from the description of  $M$  and  $M'$  that either  $\text{LEVEL}(e)$  is less than or equal to  $i$  in both the matchings or  $\text{LEVEL}(e)$  is the same in  $M$  and  $M'$ . Let us first consider the (easier) case when  $\text{LEVEL}(e) \leq i$  in  $M$  as well as  $M'$ . It follows from Fact  $F1$  stated above that the only changes in matching  $M$  and  $M'$  will be at levels  $\leq i$ . Hence the matching  $\mathcal{M}_{>i}(1)$  will be identical at the end of both the executions  $I$  and  $I'$ . Let us now consider the more interesting case of  $\text{LEVEL}(e) > i$ . Both the executions  $I$  and  $I'$  invoke the procedure  $\text{PROCESS-FREE-VERTICES}(\langle(u, \text{LEVEL}(e)), (v, \text{LEVEL}(e))\rangle)$  in this case. The reader is recommended to revisit this procedure from Section 5.2.1 before proceeding further.

In order to establish our claim about  $I$  and  $I'$ , we shall establish the following. While the matching at levels  $> i$  is being updated, for each step executed in  $I$ , the identical step can be executed in  $I'$ . Moreover, if the step in  $I$  is executed with some probability, the step will be executed with the same probability in  $I'$  as well. In order to show this, let us analyse the first iteration of the procedure  $\text{PROCESS-FREE-VERTICES}$ . Both  $I$  and  $I'$  will process  $u$  first. After disowning its edges from its present level,  $u$  owns the same set of edges in both the executions. Thereafter,  $u$  will either stay at the same level or fall by one level. If  $u$  stays at the same level, it chooses a random edge to get matched. The probability that any specific random edge is picked by  $u$  is the same in both the executions. Let us consider the case that  $u$  falls by one level. For each neighbour  $z$  of  $u$ , it follows from Lemma 5.13 that  $\phi_\ell(z)$  is the same in the case of  $M$  and  $M'$ . Hence the set of vertices rising to level  $\ell$  are the same in both the executions. In addition, the set of edges that each such vertex owns on rising to level  $j$  is also the same, hence, the probability that any specific random mate is picked is the same in both the executions. So each update in  $M$  and  $M'$  is equally likely during the processing of  $u$ . The reader may note that after each such identical update in  $M$  and  $M'$ , the matchings are identical at each level  $> i$ . Hence, Lemma 5.13 holds again for the updated matchings.

Unlike the first iteration, a generic iteration of the procedure  $\text{PROCESS-FREE-VERTICES}$  may have free vertices at levels  $\leq \text{LEVEL}(e)$  that are kept in respective queues at these levels. Suppose in the beginning of any such iteration of the procedure  $\text{PROCESS-FREE-VERTICES}$  there are two possible configurations such that the matching as well as the queue storing the free vertices are identical at each level  $> i$  but differ at levels  $\leq i$ . Lemma 5.13 will hold for these configurations as well. Therefore, along exactly the same lines as the first iteration analysed above, it can be shown that every update in the matching at level  $> i$  will be carried out with the same probability during any generic iteration for any two configurations that match at all levels  $> i$ .

Therefore, each sequence of updates in the matching is equally likely in both the executions  $I$  and  $I'$  till the last free vertex at level  $i + 1$  is processed. Henceforth, the two executions may differ. But as follows from Fact  $F2$ , it will affect only the matching at levels  $\leq i$  and there won't be any change in the matching at higher levels.

This concludes our claim for a single update. This claim can be invoked appropriately for a sequence of updates giving us the following theorem.

**Theorem 5.2** *Let  $M$  and  $M'$  be any two matchings possible by our algorithm at any time such that  $M_{>i} = M'_{>i}$ . For any sequence of  $t$  update in the graph, suppose we carry out two executions  $I$  and  $I'$  of our algorithm with the initial matching being  $M$  and  $M'$  respectively. The probability distribution of matching at every level  $> i$  will be identical at the end of both the executions. That is,*

$$\Pr[\mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(1) = \mu(1) | \mathcal{M}(0) = M] = \Pr[\mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(1) = \mu(1) | \mathcal{M}(0) = M']$$

where  $\mu(j)$ , for  $1 \leq j \leq t$ , is any maximal matching on a subset of vertices in the graph  $G(j)$ .

For the proof of Theorem 5.2, we shall apply the argument for single update inductively and use the following lemma from elementary probability theory.

**Lemma 5.14** *Suppose  $A, B, C$  are three events defined over a probability space  $(\Omega, P)$ . Then,*

$$\Pr[A \cap B | C] = \Pr[A | B \cap C] \cdot \Pr[B | C]$$

Let us define events  $C$  as  $\mathcal{M}(0) = M$  and  $C'$  as  $\mathcal{M}(0) = M'$ . We have shown that  $\Pr[\mathcal{M}_{>i}(1) = \mu(1) | C] = \Pr[\mathcal{M}_{>i}(1) = \mu(1) | C']$ . If we define event  $B$  as  $\mathcal{M}_{>i}(1) = \mu(1)$  then by another application of the arguments that we used for a single update,

$$\Pr[\mathcal{M}_{>i}(2) = \mu(2) | B, C] = \Pr[\mathcal{M}_{>i}(2) = \mu(2) | B, C']$$

Applying Lemma 5.14, we get

$$\Pr[\mathcal{M}_{>i}(2) = \mu(2), B | C] = \Pr[\mathcal{M}_{>i}(2) = \mu(2) | B, C] \cdot \Pr[B | C]$$

Since  $\Pr[B | C] = \Pr[B | C']$ , it follows that

$$\Pr[\mathcal{M}_{>i}(2) = \mu(2), B | C] = \Pr[\mathcal{M}_{>i}(2) = \mu(2), B | C']$$

The above argument can be inductively applied for every subsequent update. This completes the proof of Theorem 5.2

#### 5.4.1 Connection to the analysis

We first state two lemmas from elementary probability theory that deal with the independence of events. For the sake of completeness, the proof of these lemmas is given in Appendix.

The first lemma deals with conditional probability.

**Lemma 5.15** *Let  $A$  be an event and  $B_1, \dots, B_k$  be  $k$  mutually exclusive events defined over a probability space  $(\Omega, P)$ . If  $\Pr[A | B_j] = \rho$  for each  $1 \leq j \leq k$ , then  $\Pr[A | C] = \rho$  where event  $C = \cup_j B_j$ .*

The second lemma deals with independence of events. Let  $A$  and  $B$  be two events defined over a probability space  $(\Omega, P)$ .  $A$  is said to be independent of  $B$  if  $\Pr[A | B] = \Pr[A | \bar{B}] = \Pr[A]$ . Alternatively,  $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$ . The notion of independence gets carried over from events to random variables in a natural manner as follows.

**Definition 5.1** *An event  $A$  is said to be independent of a random variable  $X$  if for each  $x \in X$ ,  $\Pr[A | X = x] = \Pr[A]$ .*

**Lemma 5.16** *Suppose  $A$  is an event and  $X$  be a random variable defined over probability space  $(\Omega, P)$ . If  $A$  is independent of  $X$ , then for each  $x \in X$ ,*

$$\Pr[X = x | A] = \Pr[X = x]$$

Now we shall establish the connection of Theorem 5.2 to the analysis of our algorithm. In particular, we shall use this theorem to prove Lemma 5.8. Suppose a vertex  $v$  creates an epoch at level  $i$  while the algorithm processes  $k$ th update in the graph for any  $k < t$ . We shall analyse the probability space of the future matchings starting from the time just before the creation of this epoch.

While creating its epoch,  $v$  chooses its mate randomly uniformly out of  $\mathcal{O}_v^{init}$ . Clearly, the change in the matching at levels  $\leq i$  will depend on the mate that  $v$  picks. Let  $\mathbf{M}$  be the set of all possible matchings once the algorithm completes the processing of the  $k$ th update. Now notice that all matchings from the set  $\mathbf{M}$  are identical at each level  $> i$ . So it follows from Theorem 5.2 that for any two matchings  $M, M' \in \mathbf{M}$ ,

$$\begin{aligned} \Pr[\mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(k+1) = \mu(k+1) \mid \mathcal{M}(k) = M] \\ = \Pr[\mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(k+1) = \mu(k+1) \mid \mathcal{M}(k) = M'] \end{aligned}$$

Let this conditional probability be  $\rho$ . For each  $(v, w) \in \mathcal{O}_v^{init}$ , there may be several matchings in  $\mathbf{M}$  in which  $v$  is matched to  $w$ . By applying Lemma 5.15, the following equation holds for every  $(v, w) \in \mathcal{O}_v^{init}$ ,

$$\Pr[\mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(k+1) = \mu(k+1) \mid \text{MATE}(v) = w] = \rho$$

Since this probability is the same for each  $(v, w) \in \mathcal{O}_v^{init}$ , so using Definition 5.1, it follows that the matchings at levels  $> i$  during any sequence of updates is independent of the mate that  $v$  picked during the creation of its epoch. Now applying Lemma 5.16 we get the following lemma.

**Lemma 5.17** *Suppose a vertex  $v$  creates an epoch at level  $i$  while the algorithm processes  $k$ th update in the graph. Consider any sequence of updates in the graph. The mate picked by  $v$  while creating the epoch is independent of the sequence of matchings at levels  $> i$  computed by the algorithm while processing these updates. That is, for any  $t > k$ , and any  $(v, w) \in \mathcal{O}_v^{init}$ ,*

$$\Pr[\text{MATE}(v) = w \mid \mathcal{M}_{>i}(t) = \mu(t), \dots, \mathcal{M}_{>i}(k+1) = \mu(k+1)] = \Pr[\text{MATE}(v) = w] = \frac{1}{|\mathcal{O}_v^{init}|}$$

Consider any given sequence of  $t$  updates in the graph. Subsequent to the time  $v$  creates an epoch at level  $i$  during  $k$ th update, let  $\mu = \langle \mu(k+1), \dots, \mu(t) \rangle$  be the sequence of matching at levels  $> i$  as computed by the algorithm. Notice that the upward movement of  $v$  and each  $(v, w) \in \mathcal{O}_v^{init}$  after the creation of epoch is captured precisely by the corresponding update in the matching at level  $> i$ . Therefore, using  $\mu$  we can define the update sequence associated with the epoch as follows. Consider an edge  $(v, w) \in \mathcal{O}_v^{init}$  and let  $\ell$ th update in the graph be the deletion of  $(v, w)$ . Let  $j < \ell$  be the smallest integer such that  $w \in \mu_j$ , that is,  $w$  appears in the matching at level  $> i$  while processing of  $j$ th update in the graph, then the update associated with  $(v, w)$  is the upward movement. If no such  $j$  exists, the update associated with  $(v, w)$  is its deletion. Likewise, we define the update associated with  $v$ . The update sequence  $U$  for the epoch is the sequence of these updates on  $v$  and the edges of  $\mathcal{O}_v^{init}$  arranged in the chronological order.

For an update sequence  $U$  associated with an epoch, there may exist many sequences  $\{\mu_1, \dots, \mu_q\}$  such that for each of them, the update sequence associated with the epoch is  $U$ . It follows from Lemma 5.17 that the mate picked by  $v$  during its epoch is independent of each such sequence  $\mu_r, 1 \leq r \leq q$ . Therefore, using Lemma 5.15, the mate picked by  $v$  during its epoch is independent of  $U$  as well. Thus we have established the validity of Lemma 5.8.

## 6 A tight example

We tested our algorithm on random graphs of various densities and found that the matching maintained is very close to the maximum matching. This suggests that our algorithm might be able to maintain *nearly* maximum matching for dynamic graphs appearing in various practical applications. However, it is not hard to come up with an update sequence such that at the end of the sequence, the matching obtained by our algorithm is strictly half the size of maximum matching. In other words, the approximation factor 2 for the matching maintained by our algorithm is indeed tight. We present one such example as follows (see Figure 8).

Let  $G(V \cup W, E)$  be a graph such that  $V = \{v_1, v_2, \dots, v_n\}$  and  $W = \{w_1, w_2, \dots, w_n\}$  for some even number  $n$ . Consider the following update sequence. In the first phase, add edges between every pair of vertices present in  $V$ . This results in a complete subgraph on vertices of  $V$ . The size of any maximal

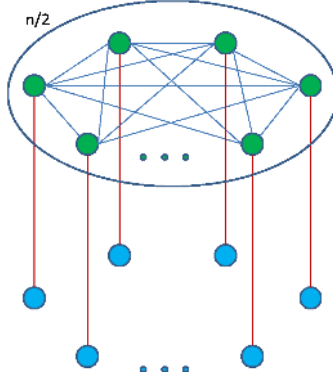


Figure 8: An example where our algorithm gives a 2-approximation. The vertices on top are in  $V$  and form a complete graph. The vertices at the bottom of the figure are in  $W$ .

matching on a complete graph of size  $n$  is  $n/2$ . After the first phase of updates ends, the size of matching obtained by our algorithm is  $n/2$ . In the second phase, add edge  $(v_i, w_i)$  for all  $i$ . Note that the degree of each  $w_i$  is one at the end of the updates. Let us now find the matching which our algorithm maintains. Let  $(v_i, v_j)$  be an edge in the matching after phase 1. Note that both these endpoints are at a level greater than  $-1$ . A vertex in  $W$  is at level  $-1$  as it does not have any adjacent edges after phase 1. When an edge  $(w_i, v_i)$  is added, since  $v_i$  is at a higher level than  $w_i$ ,  $v_i$  becomes the owner of this edge. The second invariant of  $v_i$  is not violated after this edge insertion and nothing happens at this update step and  $w_i$  still remains at level  $-1$ . Using same reasoning, we can show that  $w_j$  also remains at level  $-1$  after the addition of edge  $(v_j, w_j)$ . So matching maintained by the algorithm remains unchanged. It is easy to observe that the maximum matching of the graph  $G$  now has size  $n$  which is twice the size of the matching maintained by our algorithm.

## 7 Postscript

We presented a fully dynamic randomized algorithm for maximal matching which achieves expected amortized  $O(\log n)$  time per edge insertion or deletion. An interesting question is to explore how crucial randomization is for dynamic maximal matching.

Subsequent to an earlier version of this paper [5], Bhattacharya et al. [9] almost answered this question in affirmative by designing a deterministic algorithm that maintains  $(2 + \epsilon)$ -approximate matching in amortized  $O(\text{poly}(\log n, 1/\epsilon))$  update time. Another interesting question is to explore whether we can achieve  $O(1)$  amortized update time. Very recently Solomon [29] answered this question in affirmative as well by designing a randomized algorithm that takes  $O(t + n \log n)$  update time with high probability to process any sequence of  $t$  edge deletions. Though the basic building blocks of his algorithm are the same as ours, the two algorithms are inherently different and so are their analysis.

In our algorithm, a vertex may rise to a higher level and create a new epoch even when its matched edge is intact. But the algorithm of Solomon [29] takes a lazy approach to maintain the hierarchy of vertices wherein a vertex is processed only when it becomes absolutely necessary. Another crucial difference is the following. Our algorithm maintains a function  $\phi_v(j)$  for each vertex  $v$  and each level  $j$ . This function is used to ensure an invariant that each vertex  $v$  is at the highest possible level  $\ell$  such that the edges incident from lower levels is at least  $4^\ell$ . An important property guaranteed by this invariant is that the mate of a vertex while creating an epoch is independent of the update sequence associated with the epoch. The analysis of our algorithm crucially exploits this property. However, the explicit maintenance of  $\phi_v(j)$  imposes an overhead of  $\Theta(\log n)$  in the update time. In order to achieve  $O(1)$  update time, Solomon [29] gets rid of the

maintenance of  $\phi_v(j)$  by taking a lazy approach and a couple of new ideas. As a result, unfortunately, the property of our algorithm no longer holds for the algorithm of Solomon [29] - indeed there is dependence between the update sequence associated with an epoch created by a vertex and the random mate picked by it. Solomon [29] makes use of a new concept called *uninterrupted duration* of an epoch that bypasses the need of our property for the analysis. His analysis can be adapted to our algorithm as well and can be viewed as the correct counterpart of Lemma 4.10 in [5]. However, our new analysis has its own merits since it is based on an insightful property of our algorithm which we believe is of its own independent interest and importance.

Subsequent to the publication of the [5] there has been interesting progress in the area of dynamic matching with approximation less than 2 [7, 8, 24, 14], and dynamic weighted matching [3, 4, 14].

One of the technical challenges in theoretical computer science is to prove lower bounds for algorithmic problems. Recently there has been some progress on proving conditional lower bounds for dynamic graph algorithms [1, 15]. In the light of the lower bound presented by Abboud and Williams [1] based on  $\Omega(n^2)$  hardness of the 3SUM problem, it would be an interesting and challenging problem to see if  $c$ -approximate maximum matching for  $c < 2$  can be maintained in  $o(n)$  update time.

## 8 Acknowledgment

The possibility of an error in Lemma 4.10 of [5] was pointed out by Sayan Bhattacharya and Divyarthi Mohan. The second author would like to thank both of them for discussions on the proof of the expectation bound and the definition of  $B(v, i, k)$ . In [5], we used  $2^i$  as the threshold for raising a vertex to level  $i$ . The possibility of increasing this threshold from  $2^i$  to  $b^i$  for any constant  $b$  without any impact on the time complexity was observed by Shay Solomon [29]. We are thankful to him for this observation.

## References

- [1] Amir Abboud and Virginia Vassilevska Williams. Popular conjectures imply strong lower bounds for dynamic problems. *CoRR*, abs/1402.0054, 2014.
- [2] David J. Abraham, Robert W. Irving, Telikepalli Kavitha, and Kurt Mehlhorn. Popular matchings. *SIAM J. Comput.*, 37(4):1030–1045, 2007.
- [3] Abhash Anand, Surender Baswana, Manoj Gupta, and Sandeep Sen. Maintaining approximate maximum weighted matching in fully dynamic graphs. In *FSTTCS*, pages 257–266, 2012.
- [4] Abhash Anand, Surender Baswana, Manoj Gupta, and Sandeep Sen. Maintaining approximate maximum weighted matching in fully dynamic graphs. *CoRR*, abs/1207.3976, 2012.
- [5] Surender Baswana, Manoj Gupta, and Sandeep Sen. Fully dynamic maximal matching in  $O(\log n)$  update time. *SIAM J. Comput.* (preliminary version appeared in FOCS 2011), 44(1):88–113, 2015.
- [6] Surender Baswana, Sumeet Khurana, and Soumojit Sarkar. Fully dynamic randomized algorithms for graph spanners. *ACM Transactions on Algorithms*, 8(4):35, 2012.
- [7] Aaron Bernstein and Cliff Stein. Fully dynamic matching in bipartite graphs. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 167–179, 2015.

- [8] Aaron Bernstein and Cliff Stein. Faster fully dynamic matchings with small approximation ratios. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 692–711, 2016.
- [9] Sayan Bhattacharya, Monika Henzinger, and Danupon Nanongkai. New deterministic approximation algorithms for fully dynamic matching. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 398–411, 2016.
- [10] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.
- [11] J. Edmonds and E. L. Johnson. Matching, euler tours, and the chinese postman. *Mathematical Programming*, 5:88–124, 1973.
- [12] Harold N. Gabow and Robert Endre Tarjan. Faster scaling algorithms for general graph-matching problems. *J. ACM*, 38(4):815–853, 1991.
- [13] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–14, 1962.
- [14] Manoj Gupta and Richard Peng. Fully dynamic  $(1+\epsilon)$ -approximate matchings. In *54th Annual IEEE Symposium on Foundations of Computer Science*, 2013.
- [15] Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 21–30, 2015.
- [16] Jacob Holm, Kristian de Lichtenberg, and Mikkel Thorup. Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *J. ACM*, 48(4):723–760, 2001.
- [17] John E. Hopcroft and Richard M. Karp. An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.*, 2(4):225–231, 1973.
- [18] Chien-Chung Huang and Telikepalli Kavitha. Efficient algorithms for maximum weight matchings in general graphs with small edge weights. In *SODA*, pages 1400–1412, 2012.
- [19] Jon Kleinberg and Eva Tardos. *Algorithm Design*. Addison Wesley, 2005.
- [20] E. Lawler. *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart & Winston, Newyork, 1976.
- [21] L. Lovasz and M.D. Plummer. *Matching Theory*. AMS Chelsea Publishing, North-Holland, AmsterdamNew York, 1986.
- [22] Silvio Micali and Vijay V. Vazirani. An  $O(\sqrt{|V|}|E|)$  algorithm for finding maximum matching in general graphs. In *FOCS*, pages 17–27, 1980.
- [23] Marcin Mucha and Piotr Sankowski. Maximum matchings via gaussian elimination. In *FOCS*, pages 248–255, 2004.
- [24] Ofer Neiman and Shay Solomon. Deterministic algorithms for fully dynamic maximal matching. *CoRR*, abs/1207.1277, 2012.

- [25] Krzysztof Onak and Ronitt Rubinfeld. Maintaining a large matching and a small vertex cover. In *STOC*, pages 457–464, 2010.
- [26] Rasmus Pagh and Flemming Friche Rodler. Cuckoo hashing. *J. Algorithms*, 51(2):122–144, 2004.
- [27] Liam Roditty and Uri Zwick. Improved dynamic reachability algorithms for directed graphs. *SIAM J. Comput.*, 37(5):1455–1471, 2008.
- [28] Liam Roditty and Uri Zwick. Dynamic approximate all-pairs shortest paths in undirected graphs. *SIAM J. Comput.*, 41(3):670–683, 2012.
- [29] Shay Solomon. Fully dynamic maximal matching in constant update time. *CoRR*, abs/1604.08491, 2016.
- [30] Mikkel Thorup. Fully-dynamic min-cut. *Combinatorica*, 27(1):91–127, 2007.

## 9 Appendix

### Proof of Lemma 5.15

**Proof:**

$$\begin{aligned}
\Pr[A \cap C] &= \Pr[A \cap (\cup_i B_i)] \\
&= \sum_i \Pr[A \cap B_i] \quad \text{since } B_i\text{'s are mutually exclusive} \\
&= \sum_i \Pr[A | B_i] \cdot \Pr[B_i] \quad \text{using the definition of conditional probability} \\
&= \rho \cdot \sum_i \Pr[B_i] \\
&= \rho \cdot \Pr[\cup_i B_i] = \rho \cdot \Pr[C] \quad \text{since } B_i\text{'s are mutually exclusive}
\end{aligned}$$

Hence  $\Pr[A | C] = \Pr[A \cap C] / \Pr[C] = \rho$ . □

### Proof of Lemma 5.16

**Proof:** Since  $A$  is independent of  $X$ , so for each  $x \in X$ ,

$$\Pr[A \cap X = x] = \Pr[A] \cdot \Pr[X = x] \quad (4)$$

Hence

$$\begin{aligned}
\Pr[X = x | A] &= \frac{\Pr[A \cap X = x]}{\Pr[A]} \\
&= \frac{\Pr[A] \cdot \Pr[X = x]}{\Pr[A]} \quad \text{using Equation 4} \\
&= \Pr[X = x]
\end{aligned}$$

□

### Proof of Lemma 3.1

The asymmetric random walk problem can be seen as a special case of the famous Gambler's ruin problem described as follows.

#### Gambler's ruin problem.

There are two players who play a game that goes in rounds. Initially Player 1 has a capital of  $c$  units and Player 2 has a capital of  $c'$  units. Player 1 wins a round with probability  $p$  and loses with probability  $q = 1 - p$  independent of the previous rounds. The winner of a round takes away one unit of the capital from the opponent. The game ends when the capital of one of the players becomes 0.

The following lemma is well-known in many text books on probability theory. A concise and self contained proof is available at the link : [examplehttp://faculty.washington.edu/fm1/394/Materials/Gambler.pdf](http://faculty.washington.edu/fm1/394/Materials/Gambler.pdf).

**Lemma 9.1** *In the Gambler's ruin problem with  $p > q$ , the probability that Player 1 gets ruined is*

$$\frac{1 - (p/q)^{c'}}{1 - (p/q)^{c+c'}}$$

Let us now put an additional restriction in the problem: the total number of rounds allowed in the game is  $L$  for a given number  $L < c'$ . Notice that with this restriction Player 2 will never be ruined. As a result the game will be over when Player 1 gets ruined or when  $L$  rounds are over. The probability of Player 1 getting ruined in this restricted Gambler's problem is strictly less than the probability of Player 1 getting ruined in the original Gambler's problem described above. This is because there is a non-zero probability that Player 1 may be ruined after performing more than  $L$  steps, and the restricted Gambler's problem rules out this possibility. Hence, using Lemma 9.1, it follows that for the restricted version of the Gambler's ruin problem with  $p > q$ , the probability that Player 1 gets ruined is less than

$$\frac{1 - (p/q)^{c'}}{1 - (p/q)^{c+c'}} < \left(\frac{q}{p}\right)^c$$

The restricted version of the Gambler's ruin problem can be formulated as an asymmetric random walk problem: The walk starts at location  $c$  units to the right of the origin. In each step, the particle moves one unit to the right with probability  $p$  or one unit to the left with probability  $q = 1 - p$  independent of the past moves. The walk terminates upon reaching either the origin or when it has performed  $L$  step. This completes the proof of Lemma 3.1.