

# APPROXIMATE HAMILTON DECOMPOSITIONS OF ROBUSTLY EXPANDING REGULAR DIGRAPHS

DERYK OSTHUS AND KATHERINE STADEN

**ABSTRACT.** We show that every sufficiently large  $r$ -regular digraph  $G$  which has linear degree and is a robust outexpander has an approximate decomposition into edge-disjoint Hamilton cycles, i.e.  $G$  contains a set of  $r - o(r)$  edge-disjoint Hamilton cycles. Here  $G$  is a robust outexpander if for every set  $S$  which is not too small and not too large, the ‘robust’ outneighbourhood of  $S$  is a little larger than  $S$ . This generalises a result of Kühn, Osthus and Treglown on approximate Hamilton decompositions of dense regular oriented graphs. It also generalises a result of Frieze and Krivelevich on approximate Hamilton decompositions of quasirandom (di)graphs. In turn, our result is used as a tool by Kühn and Osthus to prove that any sufficiently large  $r$ -regular digraph  $G$  which has linear degree and is a robust outexpander even has a Hamilton decomposition.

## 1. INTRODUCTION

A Hamilton decomposition of a graph or digraph  $G$  is a set of edge-disjoint Hamilton cycles which together cover all the edges of  $G$ . The first result in the area was proved by Walecki in 1892, who showed that a complete graph  $K_n$  has a Hamilton decomposition if and only if  $n$  is odd (see e.g. [17], [3], [4]). Tillson [20] solved the analogous problem for complete digraphs in 1980. Though the area is rich in beautiful conjectures, until recently there were few general results.

Starting with a result of Frieze and Krivelevich [7], a very successful recent direction of research has been to find ‘approximate’ Hamilton decompositions, i.e. a set of edge-disjoint Hamilton cycles which cover almost all the edges of the given (di)graph. The result in [7] concerns dense quasirandom graphs and digraphs. Hypergraph versions of this result were proved by Frieze, Krivelevich and Loh [8] as well as Bal and Frieze [5]. Also, Kühn, Osthus and Treglown [16] proved an approximate version of Kelly’s conjecture. This long-standing conjecture (see [18]) states that every regular tournament has a Hamilton decomposition. In fact, the result in [16] is much more general, namely it states that every regular oriented graph on  $n$  vertices whose in- and outdegree is slightly larger than  $3n/8$  has an approximate Hamilton decomposition. Here an oriented graph is a digraph with at most one edge between each pair of vertices (whereas a digraph may have one edge in each direction between a pair of vertices).

Our main result in turn is a far reaching generalisation of the result in [16]. Instead of a degree condition, it involves an expansion condition that has recently

---

*Date:* February 26, 2013.

Deryk Osthus was supported by the EPSRC, grant no. EP/J008087/1.

been shown to have a close connection with Hamiltonicity. This notion was introduced by Kühn, Osthus and Treglown in [15]. The condition states that for every set  $S$  which is not too small and not too large, its ‘robust’ outneighbourhood is at least a little larger than  $S$  itself. More precisely, suppose that  $G$  is a digraph of order  $n$  and  $S \subseteq V(G)$ . The  $\nu$ -robust outneighbourhood  $RN_{\nu,G}^+(S)$  of  $S$  is the set of vertices with at least  $\nu n$  inneighbours in  $S$ . We say that  $G$  is a *robust*  $(\nu, \tau)$ -outexpander if

$$|RN_{\nu,G}^+(S)| \geq |S| + \nu n \text{ for all } S \subseteq V(G) \text{ with } \tau n \leq |S| \leq (1 - \tau)n.$$

Our main result states that every sufficiently large robustly outexpanding regular digraph has an approximate Hamilton decomposition.

**Theorem 1.1.** *For every  $\alpha > 0$  there exists  $\tau > 0$  such that for all  $\nu, \eta > 0$  there exists  $n_0 = n_0(\alpha, \nu, \tau, \eta)$  for which the following holds. Suppose that*

- (i)  *$G$  is an  $r$ -regular digraph on  $n \geq n_0$  vertices, where  $r \geq \alpha n$ ;*
- (ii)  *$G$  is a robust  $(\nu, \tau)$ -outexpander.*

*Then  $G$  contains at least  $(1 - \eta)r$  edge-disjoint Hamilton cycles. Moreover, this set of Hamilton cycles can be found in time polynomial in  $n$ .*

As observed in Lemma 12.1 of [13], every oriented graph whose in- and out-degrees are all at least slightly larger than  $3n/8$  is a robust outexpander, so this does generalise the main result of [16]. Moreover, it turns out that one can relax condition (i) to the requirement that  $G$  is ‘almost regular’. This is due to the fact (observed in [13]) that every almost regular robustly expanding digraph contains a spanning regular digraph of similar degree.

**Corollary 1.2.** *For every  $\alpha > 0$  there exists  $\tau > 0$  such that for all  $\nu, \eta > 0$  there exist  $n_0 = n_0(\alpha, \nu, \tau, \eta)$  and  $\gamma = \gamma(\alpha, \nu, \tau, \eta) > 0$  for which the following holds. Suppose that*

- (i)  *$G$  is a digraph on  $n \geq n_0$  vertices with  $(\alpha - \gamma)n \leq d_G^\pm(x) \leq (\alpha + \gamma)n$  for every  $x$  in  $G$ ;*
- (ii)  *$G$  is a robust  $(\nu, \tau)$ -outexpander.*

*Then  $G$  contains at least  $(\alpha - \eta)n$  edge-disjoint Hamilton cycles. Moreover, this set of Hamilton cycles can be found in time polynomial in  $n$ .*

The result in [16] extends to almost regular oriented graphs in the same way, but is inherently non-algorithmic (see Section 2). Since, for dense digraphs, the condition of being a robust outexpander is much weaker than that of being quasirandom, Corollary 1.2 is much more general than the result in [7] mentioned earlier. Moreover, it is best possible in the sense that, for an almost regular digraph, an approximate Hamilton decomposition is obviously the best one can hope for.

Theorem 1.1 is used as an essential tool by Kühn and Osthus in [13] to prove the following result, which (under the same conditions) guarantees not only an approximate decomposition, but a Hamilton decomposition.

**Theorem 1.3.** *For every  $\alpha > 0$  there exists  $\tau > 0$  such that for every  $\nu > 0$  there exists  $n_0 = n_0(\alpha, \nu, \tau)$  for which the following holds. Suppose that*

- (i)  $G$  is an  $r$ -regular digraph on  $n \geq n_0$  vertices, where  $r \geq \alpha n$ ;
- (ii)  $G$  is a robust  $(\nu, \tau)$ -outexpander.

Then  $G$  has a Hamilton decomposition. Moreover, this decomposition can be found in time polynomial in  $n$ .

So as a special case, Theorem 1.3 implies that Kelly's conjecture holds for all sufficiently large regular tournaments. It also implies a conjecture of Erdős on Hamilton decompositions of regular tournaments. However, it turns out that the notion of robust (out)expansion extends far beyond the class of tournaments and many further applications of Theorem 1.1 are explored by Kühn and Osthus in [14]. For example, the notion of robust expansion can be extended to undirected graphs in a natural way and one can deduce a version of Theorem 1.3 for undirected graphs. In [14] this in turn is used to prove an approximate version of a conjecture of Nash-Williams on Hamilton decompositions of dense regular graphs. Random regular graphs of linear degree as well as  $(n, d, \lambda)$ -graphs (for appropriate values of these parameters) are further examples of robustly expanding graphs. In combination with a result of Gutin and Yeo [10], Theorem 1.3 can also be used to solve a problem of Glover and Punnen [9] as well as Alon, Gutin and Krivelevich [1] on TSP tour domination (see [13] for details). For this application, it is crucial that the Hamilton decomposition can be found in polynomial time.

Roughly speaking, the argument leading to Theorem 1.3 uses Theorem 1.1 in the following way: let  $G$  be a robustly expanding digraph. The first step is to remove a 'robustly decomposable' spanning regular digraph  $H$  from  $G$  to obtain  $G'$ .  $H$  will be sparse compared to  $G$  and will have the property that it has a Hamilton decomposition even if we add the edges of a digraph  $H'$ , which is very sparse compared to  $H$  and also regular (on the same vertex set) but otherwise arbitrary. Now  $G'$  is still a robust outexpander, so one can apply Theorem 1.1 to  $G'$  obtain an approximate Hamilton decomposition of  $G'$ . Let  $H'$  denote the set of edges not contained in any of the Hamilton cycles of this approximate decomposition of  $G'$ . Then the fact that  $H$  is robustly decomposable implies that  $H \cup H'$  has a Hamilton decomposition. Together with the approximate decomposition of  $G'$ , this yields a Hamilton decomposition of the entire digraph  $G$ . Note that the above approach means that for Theorem 1.3 to be algorithmic, one needs Theorem 1.1 to be algorithmic too.

This paper is organised as follows. In the next section, we give a brief outline of the argument. We then collect the necessary tools in Section 3 (which is mostly concerned with Szemerédi's regularity lemma) and Section 4 (which mainly collects properties of robust outexpanders). We then prove Theorem 1.1 in Section 5. In Section 6, we deduce Corollary 1.2 from Theorem 1.1.

## 2. SKETCH OF THE PROOF OF THEOREM 1.1

Roughly speaking, the strategy of the proof of Theorem 1.1 is the following. Suppose that a digraph  $G$  satisfies the conditions of Theorem 1.1. First remove the edges of a carefully chosen spanning sparse subdigraph  $H$  from  $G$  and let  $G'$  consist of the remaining edges of  $G$ . Next, find an approximate decomposition of

$G'$  into edge-disjoint 1-factors  $F_i$  (where a 1-factor is a spanning union of vertex-disjoint cycles). Finally, the aim is to transform each  $F_i$  into a Hamilton cycle by removing some of its edges and adding some edges of  $H$ . One immediate obstacle to a naïve implementation of this approach is that the  $F_i$  might consist of many cycles, so turning each of them into a Hamilton cycle might require more edges from  $H$  than one can afford. In [7, 16], this was overcome (loosely speaking) by choosing the 1-factors  $F_i$  randomly. It turns out that this has the advantage that the  $F_i$  will have few cycles, i.e. they are already close to being Hamilton cycles. One disadvantage is that this approach is inherently non-algorithmic (and does not seem derandomisable).

A second problem is how to make sure that  $H$  contains the edges that are required to transform each  $F_i$  into a Hamilton cycle. We overcome this by choosing  $H$  and the 1-factors  $F_i$  according to the vertex partition of  $G$  obtained from Szemerédi's regularity lemma. More precisely, we apply the regularity lemma to partition  $G$  into clusters  $V_1, \dots, V_L$  of vertices such that almost all ordered pairs of clusters induce a pseudorandom subdigraph of  $G$ , together with a small (but typically troublesome) exceptional set  $V_0$ . We define the 'reduced multidigraph'  $R(\beta)$  whose vertices are the clusters  $V_j$  with (multiple) edges from  $V_j$  to  $V_k$  if the corresponding subdigraph of  $G$  is pseudorandom and dense. Here the number of edges from  $V_j$  to  $V_k$  is proportional to the density of  $G[V_j, V_k]$ . So each edge of  $R(\beta)$  corresponds to a bipartite pseudorandom digraph between the corresponding pair of clusters in  $G$  (where all these pseudorandom digraphs have same density  $\beta$ ).  $R(\beta)$  inherits many of the properties of  $G$ , in particular it is an almost regular robust outexpander with large minimum semidegree.

The next step is to use the Max-Flow-Min-Cut Theorem to find a spanning regular subdigraph of  $R(\beta)$  which contains almost all edges of  $R(\beta)$ . We can now (arbitrarily) partition this regular subdigraph into a collection of edge-disjoint 1-factors  $F_i$  of  $R(\beta)$  (see Section 5.1). Each of the  $F_i$  corresponds to a vertex-disjoint collection of 'blown-up' cycles which spans most of  $V(G)$ . We will denote each of these collections by  $G_i$  and call  $G_i$  the  $i$ th slice of  $G$ . Note that the  $G_i$  are all edge-disjoint.

Roughly speaking, the aim is to add a small number of edges (which do not lie in any of the other slices) to each  $G_i$  to transform  $G_i$  into a regular digraph which has an approximate Hamilton decomposition. Together, these approximate Hamilton decompositions of the slices then yield an approximate Hamilton decomposition of  $G$ . In Section 5.2, we put aside three sparse subdigraphs  $H_0, H_1, H_2$  which we will use to add the required edges to each  $G_i$ . So together,  $H_0, H_1$  and  $H_2$  play the role of the digraph  $H$  mentioned earlier.

So far we have ignored the exceptional vertices, but to obtain a regular spanning subdigraph we need to incorporate them into each slice  $G_i$ . For convenience, we call any exceptional vertex  $x \in V_0$  and each edge incident with  $V_0$  'red'. In Sections 5.6 and 5.7, we will add red edges to each  $G_i$  in such a way that the resulting slice  $G_i$  is almost regular and only a small part of each cluster is incident to any red edges. Some of these edges come from  $H_1$  and the others will be edges of  $G$  which are not contained in any of the  $H_j$  or any of the  $G_i$  constructed so far.

Together with these red edges, each  $G_i$  is now an almost regular digraph consisting mainly of a union of blown-up cycles. On the other hand,  $G_i$  may not even be connected. But to guarantee many edge-disjoint Hamilton cycles in  $G_i$ , we clearly need to have sufficiently many edge-disjoint paths between these blown-up cycles. For this, we define an ordering of the cycles  $D_1, \dots, D_\ell$  of  $F_i$  and specify ‘bridge vertices’  $x_{i,j}$  (one for each successive pair of cycles) so that  $x_{i,j}$  has many inneighbours in  $D_j$  and many outneighbours in  $D_{j+1}$ . We find the edges incident to these  $x_{i,j}$  within  $H_0$  (see Section 5.5).

We would now like to find a spanning regular subdigraph in each  $G_i$  whose degree is almost as large as that of  $G_i$ . Trivially, this regular subdigraph would then have a decomposition into 1-factors. However, as we have little control over the red edges added so far, they may prevent us from finding a regular subdigraph (see Section 5.8 for a discussion and an example). For this reason, we add extra (red) edges to  $G_i$  from  $H_2$  to balance out the existing red edges. In this way, we can ensure that for each cluster  $V$  of a blown-up cycle  $D$ , the number of edges leaving  $V$  in  $G_i$  equals the number of edges entering its successor  $V^+$  on  $D$ . This is achieved in Sections 5.8 and 5.9, by considering an auxiliary reduced digraph  $R^*$  which also turns out to be a robust outexpander (the latter property is crucial here).

As indicated above, in Section 5.10, we can now find a spanning  $\kappa$ -regular subdigraph  $G_i^*$  of each  $G_i$  (for a suitable  $\kappa$ ). We now decompose each  $G_i^*$  into 1-factors  $f_{i,1}, \dots, f_{i,\kappa}$ . Our aim is to transform each  $f_{i,j}$  into a Hamilton cycle by adding and removing a few edges. The edges we add will be taken from a very sparse digraph  $H_{3,i}$  which we removed from  $G_i$  earlier (so  $H_{3,i}$  can also be viewed as a union of blown-up cycles). The key point of the proof is that we can achieve this transformation by using a very small number of edges from  $H_{3,i}$  for each  $f_{i,j}$ . The reason for this is that we can guarantee that the red edges added in the course of the proof are ‘localised’ within each  $G_i$ , i.e. on each blown-up cycle of each  $F_i$  there are long intervals of clusters which are not incident to any red edges. This means that for each 1-factor  $f_{i,j}$ , its subdigraph induced by any such interval  $I$  consists of long paths. If some of these paths lie on different cycles of  $f_{i,j}$ , we can merge these into a single cycle by adding and removing edges of  $H_{3,i}$  which are induced by just a single pair of consecutive clusters on  $I$ . Crucially, this enables us to use the bipartite subdigraphs of  $H_{3,i}$  induced by other pairs of consecutive clusters on  $I$  to transform other 1-factors  $f_{i,j'}$  of the slice  $G_i$ . Repeating this process until we have merged all cycles of  $f_{i,j}$  into a single cycle eventually transforms the  $f_{i,j}$  into  $\kappa$  edge-disjoint Hamilton cycles, as required (see Lemma 4.5 and Section 5.11).

An approach based on the regularity lemma was already used in [16]. However, as mentioned earlier, the argument there relied on a random choice of the 1-factors, which did not translate into an algorithm. This problem is overcome by the above ‘localisation’ idea, which automatically produces 1-factors which are ‘well behaved’ with respect to red edges in the sense described above. However, this ‘localisation property’ is quite difficult to achieve and relies on additional ideas such as a refinement of the original regularity partition and a special ‘unwinding’ of blown-up cycles (see Section 5.3 and Lemma 4.4).

## 3. NOTATION AND THE DIREGULARITY LEMMA

**3.1. Notation.** Throughout we will omit floors and ceilings where the argument is unaffected. The constants in the hierarchies used to state our results are chosen from right to left. For example, if we claim that a result holds whenever  $0 < 1/n \ll a \ll b \ll c \leq 1$  (where  $n$  is the order of the graph or digraph), then there are non-decreasing functions  $f : (0, 1] \rightarrow (0, 1]$ ,  $g : (0, 1] \rightarrow (0, 1]$  and  $h : (0, 1] \rightarrow (0, 1]$  such that the result holds for all  $0 < a, b, c \leq 1$  and all  $n \in \mathbb{N}$  with  $b \leq f(c)$ ,  $a \leq g(b)$  and  $1/n \leq h(a)$ . Hierarchies with more constants are defined in a similar way. We write  $a = b \pm \varepsilon$  for  $a \in [b - \varepsilon, b + \varepsilon]$ .

For an undirected graph  $G$  containing a vertex  $x$  we write  $N_G(x)$  for the neighbourhood of  $x$  and  $d_G(x)$  for its degree. For a digraph  $G$  we write  $xy$  for the edge directed from  $x$  to  $y$  and write  $N_G^+(x)$  for the *outneighbourhood*, that is the set of vertices receiving an edge from  $x$ , and write  $d_G^+(x) := |N_G^+(x)|$  for the *outdegree* of  $x$ . We define the *inneighbourhood*  $N_G^-(x)$  and *indegree*  $d_G^-(x)$  similarly. For a collection of vertices  $U \subseteq V(G)$  we write  $d_G^+(U)$  for the total number of edges sent out by the vertices in  $U$ . We define  $d_G^-(U)$  analogously. We will omit the  $G$  subscript in the above and in similar situations elsewhere if this is unambiguous. Denote the minimum outdegree by  $\delta^+(G)$  and the minimum indegree by  $\delta^-(G)$ . Let the *minimum semidegree*  $\delta^0(G)$  be the minimum of  $\delta^+(G)$  and  $\delta^-(G)$ . Denote the maximum outdegree by  $\Delta^+(G)$  and define  $\Delta^-(G)$  and analogously. Let  $\Delta^0(G)$  denote the maximum of  $\Delta^+(G)$  and  $\Delta^-(G)$ . If  $G$  is a multidigraph then neighbourhoods are multisets. For any positive integer  $r$ , an  *$r$ -regular* digraph on  $n$  vertices is such that every vertex has exactly  $r$  outneighbours and  $r$  inneighbours. A *1-factor* of a multidigraph  $G$  is a 1-regular spanning digraph; that is, a collection of vertex-disjoint cycles that together contain all the vertices of  $G$ .

If  $G$  is a multidigraph and  $U \subseteq V(G)$ , we write  $G[U]$  for the sub-multidigraph of  $G$  *induced* by  $U$ . That is, the digraph with vertex set  $U$  and edge set obtained from  $E(G)$  by including only those edges with both endpoints contained in  $U$ . If  $G[U]$  has empty edge set, we say that  $U$  is an *isolated* subset of  $G$ . If  $G$  is a digraph and  $U \subseteq V(G)$  we write  $G \setminus U$  for the digraph with vertex set  $V(G) \setminus U$  and edge set obtained from  $E(G)$  by deleting all edges incident to a vertex of  $U$ .

Given a digraph  $R$  and a positive integer  $r$ , the  *$r$ -fold blow-up*  $r \otimes R$  of  $R$  is the digraph obtained from  $R$  by replacing every vertex  $x$  of  $R$  by  $r$  vertices and replacing every edge  $xy$  of  $R$  by the complete bipartite graph  $K_{r,r}$  between the two sets of  $r$  vertices corresponding to  $x$  and  $y$  such that all the edges of  $K_{r,r}$  are oriented towards the  $r$  vertices corresponding to  $y$ . We say that any edge in this  $K_{r,r}$  is *contained in the blow-up of  $xy$* . Now consider the case when  $V_1, \dots, V_k$  is a partition of some set  $V$  of vertices and  $R$  is a digraph whose vertices are  $V_1, \dots, V_k$ . If  $R$  is a directed cycle, say  $R = C = V_1 \dots V_k$ , and  $G$  is a digraph with  $V(G) \subseteq V = V_1 \cup \dots \cup V_k$ , we say that *(the edges of)  $G$  wind(s) around  $C$*  if, for every edge  $xy$  of  $G$ , there exists an index  $j$  such that  $x \in V_j$  and  $y \in V_{j+1}$ .

**3.2. A Chernoff bound and its derandomisation.** In the proof of Claims 5.3 and 5.4, we will use the following standard Chernoff type bound (see e.g. Corollary 2.3 in [12] and Theorem 2.2 in [19]).

**Proposition 3.1.** *Suppose  $X$  has binomial distribution and  $0 < a < 1$ . Then*

$$\mathbb{P}(X \geq (1+a)\mathbb{E}X) \leq e^{-\frac{a^2}{3}\mathbb{E}X} \text{ and } \mathbb{P}(X \leq (1-a)\mathbb{E}X) \leq e^{-\frac{a^2}{3}\mathbb{E}X}.$$

To obtain an algorithmic version of Theorem 1.1, we need to ‘derandomise’ our applications of Proposition 3.1. This can be done via the well known ‘method of conditional probabilities.’ The following result of Srivastav and Stangier (Theorem 2.10 in [19]) provides a convenient way to apply this method. It implies that any construction based on a polynomial number of applications of Proposition 3.1 can be derandomised to provide a polynomial time algorithm.

Suppose we are given  $N$  independent 0/1 random variables  $X_1, \dots, X_N$  where  $\mathbb{P}(X_j = 1) = p$  and  $\mathbb{P}(X_j = 0) = 1 - p$  for some rational  $0 \leq p \leq 1$ . Suppose that  $1 \leq i \leq m$ . Let  $w_{ij} \in \{0, 1\}$ . Denote by  $\phi_i$  the random variables  $\phi_i := \sum_{j=1}^N w_{ij}X_j$ . Fix  $\beta_i$  with  $0 < \beta_i < 1$ . Now let  $E_i^+$  denote the event that  $\phi_i \geq (1 + \beta_i)\mathbb{E}[\phi_i]$  and let  $E_i^-$  denote the event that  $\phi_i \leq (1 - \beta_i)\mathbb{E}[\phi_i]$ . Let  $E_i$  be either  $E_i^+$  or  $E_i^-$ .

**Theorem 3.2.** [19] *Let  $E_1, \dots, E_m$  be events such that*

$$\sum_{i=1}^m e^{-\beta_i^2 \mathbb{E}(\phi_i)/3} \leq 1/2.$$

*Then*

$$\mathbb{P}\left(\bigcap_{i=1}^m E_i\right) \geq 1/2$$

*and a vector  $x \in \bigcap_{i=1}^m E_i$  can be constructed in time  $O(mN^2 \log(mN))$ .*

In general, it will usually be clear that the proofs can be translated into polynomial time algorithms. We do not prove an explicit bound on the time needed to find the set of edge-disjoint Hamilton cycles guaranteed by Theorem 1.1, apart from the fact that the time is polynomial in  $n$ .

**3.3. The diregularity lemma.** We will use the directed version of Szemerédi’s regularity lemma. To state it we need some definitions. We write  $d_G(A, B)$  for the *density*  $\frac{e_G(A, B)}{|A||B|}$  of an undirected bipartite graph  $G$  with vertex classes  $A$  and  $B$ . Given  $\varepsilon > 0$  we say that  $G$  is  $\varepsilon$ -*regular* if every  $X \subseteq A$  and  $Y \subseteq B$  with  $|X| \geq \varepsilon|A|$  and  $|Y| \geq \varepsilon|B|$  satisfy  $|d(A, B) - d(X, Y)| \leq \varepsilon$ . Given  $\varepsilon, d \in (0, 1)$  we say that  $G$  is  $(\varepsilon, d)$ -*regular* if  $G$  is  $\varepsilon$ -regular and  $d_G(A, B) = d \pm \varepsilon$ . We say that  $G$  is  $(\varepsilon, d)$ -*superregular* if both of the following hold:

- $G$  is  $(\varepsilon, d)$ -regular;
- $d(a) = (d \pm \varepsilon)|B|$ ,  $d(b) = (d \pm \varepsilon)|A|$  for all  $a \in A$ ,  $b \in B$ .

Given disjoint vertex sets  $A$  and  $B$  in a digraph  $G$ , write  $(A, B)_G$  for the oriented bipartite subgraph of  $G$  whose vertex classes are  $A$  and  $B$  and whose edges are all those from  $A$  to  $B$  in  $G$ . We say that  $(A, B)_G$  has any of the regularity properties

above if the requirements hold for the underlying undirected bipartite graph of  $(A, B)_G$ .

The diregularity lemma is a variant of the regularity lemma for digraphs due to Alon and Shapira [2]. We will use the degree form which can be derived from the standard version in the same manner as the undirected degree form. The proof of the diregularity lemma itself is similar to the undirected version.

**Lemma 3.3.** (Degree form of the diregularity lemma) *For every  $\varepsilon \in (0, 1)$  and every integer  $M'$  there are integers  $M$  and  $n_0$  such that if  $G$  is a digraph on  $n \geq n_0$  vertices and  $d \in [0, 1]$  is any real number, then there is a partition of the vertex set of  $G$  into  $V_0, \dots, V_L$  and a spanning subdigraph  $G'$  of  $G$  such that the following holds:*

- $M' \leq L \leq M$ ;
- $|V_0| \leq \varepsilon n$ ;
- $|V_1| = \dots = |V_L| =: m$ ;
- $d_{G'}^\pm(x) > d_G^\pm(x) - (d + \varepsilon)n$  for all vertices  $x \in V(G)$ ;
- For all  $1 \leq i \leq L$  the digraph  $G'[V_i]$  is empty;
- For all  $1 \leq j \leq L$  with  $i \neq j$  the pair  $(V_i, V_j)_{G'}$  is  $\varepsilon$ -regular and has density either 0 or at least  $d$ .

We call  $V_1, \dots, V_L$  *clusters*,  $V_0$  the *exceptional set* and the vertices in  $V_0$  *exceptional vertices*. We refer to  $G'$  as the *pure digraph*. The last condition of the lemma says that all pairs of clusters are  $\varepsilon$ -regular in both directions (but possibly with different densities). The *reduced digraph*  $R$  of  $G$  with parameters  $\varepsilon, d$  and  $M'$  is the digraph whose vertices are  $V_1, \dots, V_L$  and in which  $V_i V_j$  is an edge precisely when  $(V_i, V_j)_{G'}$  is  $\varepsilon$ -regular and has density at least  $d$ . For each edge  $V_i V_j$  of  $G$  we write  $d_{ij}$  for the density of  $(V_i, V_j)_{G'}$ . Suppose  $0 < 1/M' \ll \varepsilon \ll \beta \ll d \ll 1$ . The *reduced multidigraph*  $R(\beta)$  of  $G$  with parameters  $\varepsilon, \beta, d, M'$  is obtained from  $R$  by setting  $V(R(\beta)) := V(R)$  and adding  $\lfloor d_{ij}/\beta \rfloor$  directed edges from  $V_i$  to  $V_j$  whenever  $V_i V_j \in E(R)$ . These digraphs inherit some of the key properties of  $G$ , as the next few results show (which are variants of well known observations, see e.g. Lemma 11 in [16] for the next result).

**Lemma 3.4.** *Let  $0 < 1/n_0 \ll 1/M' \ll \varepsilon \ll \beta \ll d \leq d' \ll c_1 \leq c_2 < 1$  and let  $G$  be a digraph of order  $n \geq n_0$  with  $\delta^0(G) \geq c_1 n$  and  $\Delta^0(G) \leq c_2 n$ . Apply Lemma 3.3 with parameters  $\varepsilon, d$  and  $M'$  to obtain a pure digraph  $G'$  and a reduced digraph  $R$  of  $G$  and let  $R'$  denote the subdigraph of  $R$  whose edges correspond to pairs of density at least  $d'$ . Let  $R(\beta)$  denote the reduced multidigraph of  $G$  with parameters  $\varepsilon, \beta, d$  and  $M'$  and let  $R'(\beta)$  be the multidigraph obtained from  $R(\beta)$  by including only those edges which also correspond to an edge of  $R'$ . Let  $L := |R| = |R(\beta)|$ . Then*

- (i)  $\delta^0(R') \geq (c_1 - 3d')L$ .
- (ii)  $\delta^0(R'(\beta)) \geq (c_1 - 4d')\frac{L}{\beta}$  and  $\Delta^0(R'(\beta)) \leq (c_2 + 2\varepsilon)\frac{L}{\beta}$ .

*Proof.* To prove (i), we consider the weighted digraph  $R'_w$  obtained from  $R'$  by giving each edge  $V_i V_j$  of  $R'$  weight  $d_{ij}$ . Given a cluster  $V_i$ , we write  $w^+(V_i)$  for the

sum of the weights of all edges sent out by  $V_i$  in  $R'_w$ . We define  $w^-(V_i)$  similarly and write  $w^0(R'_w)$  for the minimum of  $\min\{w^+(V_i), w^-(V_i)\}$  over all clusters  $V_i$ . Note that  $\delta^0(R') \geq w^0(R'_w)$ . Moreover, Lemma 3.3 implies that  $d_{G' \setminus V_0}^\pm(x) > (c_1 - 2d)n$  for all  $x \in V(G' \setminus V_0)$ . Thus each  $V_i \in V(R')$  satisfies

$$(c_1 - 2d)nm \leq e_{G'}(V_i, V(G') \setminus V_0) \leq m^2 w^+(V_i) + (d'm^2)L$$

and so  $w^+(V_i) \geq (c_1 - 2d - d')L \geq (c_1 - 3d')L$ . Arguing in the same way for inweights gives us  $\delta^0(R') \geq w^0(R'_w) \geq (c_1 - 3d')L$ . We can deduce the first part of (ii) by noting that

$$d_{R'(\beta)}^+(V_i) = \sum_{V_j \in N_{R'}^+(V_i)} \lfloor d_{ij}/\beta \rfloor \geq w^+(V_i)/\beta - L > (c_1 - 4d')\frac{L}{\beta}.$$

Similar arguments can be used to show the remaining bounds.  $\square$

**Lemma 3.5.** *Let  $M', n_0$  be positive integers and let  $\varepsilon, d, \nu, \tau$  be positive constants such that  $1/n_0 \ll 1/M' \ll \varepsilon \ll d \leq d' \leq \nu \leq \tau < 1$  and  $d' \leq \nu/20$ . Let  $G$  be a digraph on  $n \geq n_0$  vertices such that  $G$  is a robust  $(\nu, \tau)$ -outexpander. Let  $R$  be the reduced digraph of  $G$  with parameters  $\varepsilon, d$  and  $M'$  with clusters of size  $m$  and let  $R'$  be the subdigraph of  $R$  whose edges correspond to pairs of density at least  $d'$ . Then  $R'$  is a robust  $(\nu/4, 3\tau)$ -outexpander.*

*Proof.* Let  $G'$  denote the pure digraph,  $L := |V(R)|$ , let  $V_1, \dots, V_L$  be the clusters of  $G$  and  $V_0$  the exceptional set. Let  $m := |V_1| = \dots = |V_L|$ . Suppose  $S \subseteq V(R')$  has  $3\tau L \leq |S| \leq (1 - 3\tau)L$ . Let  $S_G$  denote the union of all vertices in clusters in  $S$ . So  $S_G \subseteq V(G)$  and  $2\tau n \leq |S_G| \leq (1 - 2\tau)n$ . For every  $x \in RN_{\nu, G}^+(S_G)$  we have that  $|N_{G'}^-(x) \cap S_G| \geq |N_G^-(x) \cap S_G| - (d + \varepsilon)n \geq (\nu - d - \varepsilon)n \geq \nu n/2$ . This implies that

$$|RN_{\nu/2, G'}^+(S_G)| \geq |RN_{\nu, G}^+(S_G)| \geq |S_G| + \nu n \geq |S|m + \nu Lm$$

and every vertex  $x \in RN_{\nu/2, G'}^+(S_G)$  has at least  $\nu n/2$  inneighbours in  $S_G$ . Suppose, for a contradiction, that  $|RN_{\nu/4, R'}^+(S)| < |S| + \nu L/4$ . Let  $RN'$  denote the union of all vertices in clusters in  $RN_{\nu/4, R'}^+(S)$  and let  $T := RN_{\nu/2, G'}^+(S_G) \setminus RN'$ ; then  $|T| \geq \nu n/4$ .

Note that by definition, for all  $V$  outside  $RN_{\nu/4, R'}^+(S)$ , there exists a collection  $\mathcal{V}$  of at least  $|S| - \nu L/4$  clusters  $U \in S$  so that there is no edge from  $U$  to  $V$  in  $R'$ . So by assumption such a  $\mathcal{V}$  exists for any  $V$  which has non-empty intersection with  $T$ .

We say that a vertex  $x \in V$  is *bad* if it has indegree at least  $2d'm$  in at least  $\sqrt{\varepsilon}L$  of the clusters in  $\mathcal{V}$ . The final property of Lemma 3.3 implies that there are at most  $\varepsilon m$  vertices in  $V$  that have indegree at least  $2d'm$  in some fixed cluster of  $\mathcal{V}$ . So by double counting the number of such vertex-cluster pairs, we see that any cluster contains at most  $\sqrt{\varepsilon}m$  bad vertices.

Say that a cluster  $V$  is *significant* if  $|V \cap T| \geq \varepsilon^{1/3}m$ . Then there are at least  $\nu L/5$  significant clusters and we write  $V' := V \cap T$ . Consider any  $x \in V'$ , where

$V$  is significant. We say that a cluster  $U$  in  $S$  is *rich for  $x$*  if  $x$  has at least  $\nu m/10$  inneighbours in  $U$ . Since  $x$  has at least  $\nu n/2$  inneighbours in  $S_G$ , there are at least  $\nu L/3$  clusters in  $S$  which are rich for  $x$ . So there are at least  $\nu L/12 \geq \sqrt{\varepsilon}L$  clusters in  $\mathcal{V}$  which are rich for  $x$ . Since  $d' \leq \nu/20$ , this means that every  $x$  in  $V'$  is bad. Thus  $V$  contains at least  $\varepsilon^{1/3}m$  bad vertices, a contradiction.  $\square$

The following simple observation is well known, the version given here is proved as Proposition 4.3(i) and (iii) in [13].

**Proposition 3.6.** *Suppose that  $0 < 1/m \ll \varepsilon \leq d' \leq d \ll 1$ . Let  $G$  be a bipartite graph with vertex classes  $A$  and  $B$  of size  $m$ . Suppose that  $G'$  is obtained from  $G$  by removing at most  $d'm$  vertices from each vertex class and at most  $d'm$  edges incident to each vertex from  $G$ .*

- (i) *If  $G$  is  $(\varepsilon, d)$ -regular then  $G'$  is  $(2\sqrt{d'}, d)$ -regular.*
- (ii) *If  $G$  is  $(\varepsilon, d)$ -superregular then  $G'$  is  $(2\sqrt{d'}, d)$ -superregular.*

The next result shows that we can partition an  $\varepsilon$ -(super)regular pair into edge-disjoint  $\varepsilon'$ -(super)regular spanning subgraphs. The proof is almost identical to that of Lemma 4.10(iii) and (iv) in [13] (which covers the case  $K = 2$ ) so we omit it here.

**Lemma 3.7.** *Let  $K$  be an integer and let  $0 < 1/m \ll \varepsilon \ll \gamma_1, \dots, \gamma_K \ll 1$  such that  $\gamma_1 + \dots + \gamma_K \leq d \leq 1$ .*

- (i) *If  $G$  is an  $(\varepsilon, d)$ -regular bipartite graph with vertex classes  $X, Y$  of size  $m$ , then it contains  $K$  edge-disjoint spanning subgraphs  $J_1, \dots, J_K$  such that for each  $1 \leq k \leq K$  we have that  $J_k$  is  $(\varepsilon^{1/12}, \gamma_k)$ -regular. Moreover, if  $x \in X$  satisfies  $d_G(x) = (d \pm \varepsilon)m$ , then  $d_{J_k}(x) = (\gamma_k \pm \varepsilon^{1/12})m$  for each  $1 \leq k \leq K$ .*
- (ii) *If  $G$  is an  $(\varepsilon, d)$ -superregular bipartite graph with vertex classes of size  $m$ , then it contains  $K$  edge-disjoint spanning subgraphs  $J_1, \dots, J_K$  such that for each  $1 \leq k \leq K$  we have that  $J_k$  is  $(\varepsilon^{1/12}, \gamma_k)$ -superregular.*

Moreover, the spanning subgraphs can be found in time polynomial in  $m$ .

The proof of Theorem 1.1 begins by decomposing our digraph into ‘blown-up’ 1-factors and we will need the following well known and easy fact that allows us to extract almost spanning blown-up 1-factors in which pairs are superregular.

**Lemma 3.8.** *Let  $0 < \varepsilon \leq \gamma \leq 1 \leq m$  and let  $D$  be a digraph with vertex clusters  $V_1, \dots, V_k$  each of size  $m$  such that  $(V_j, V_{j+1})_D$  is  $(\varepsilon, \gamma)$ -regular for  $1 \leq j \leq k$ , where  $V_{k+1} := V_1$ . Then there exists a subdigraph  $D'$  of  $D$  with vertex clusters  $V'_1, \dots, V'_k$  where  $V'_j \subseteq V_j$ ,  $|V'_j| = (1 - 2\varepsilon)m$  and  $(V'_j, V'_{j+1})_{D'}$  is  $(4\varepsilon, \gamma)$ -superregular for  $1 \leq j \leq k$ , where  $V'_{k+1} := V'_1$ .*

*Proof.* For each  $1 \leq j \leq k$ , each  $V_i$  contains at most  $2\varepsilon m$  vertices whose outdegree or indegree in  $D$  is either at most  $(\gamma - 2\varepsilon)m$  or at least  $(\gamma + 2\varepsilon)m$ . Deleting exactly  $2\varepsilon m$  vertices including these from each cluster gives us  $D'$ .  $\square$

We will use the following crude version of the fact that every  $\varepsilon$ -regular pair contains a subgraph of given maximum degree  $\Delta$  whose average degree is close to  $\Delta$ , which is Lemma 13 in [16].

**Lemma 3.9.** *Suppose that  $0 < 1/m \ll \varepsilon', \varepsilon \ll d_0 \leq d_1 \ll 1$  and that  $(A, B)$  is an  $(\varepsilon, d_1)$ -regular pair with  $m$  vertices in each class. Then  $(A, B)$  contains a subgraph  $H$  whose maximum degree is at most  $d_0 m$  and whose average degree is at least  $d_0 m/8$ .*

The proof proceeds by greedily removing matchings and so  $H$  can be found in polynomial time. (ii) of the following observation is proved as Lemma 5.3 in [13]; (i) is immediate from the definition.

**Lemma 3.10.** *Let  $r \geq 3$  and let  $G$  be a robust  $(\nu, \tau)$ -outexpander with  $0 < 3\nu \leq \tau < 1$ . Let  $G'$  be the  $r$ -fold blow-up of  $G$ . Then*

- (i)  $\delta^0(G') = r\delta^0(G)$ .
- (ii)  $G'$  is a robust  $(\nu^3, 2\tau)$ -outexpander.

**3.4. Uniform refinements.** We will also need to partition each vertex cluster into equal parts in such a way that the neighbourhoods of each vertex restricted to each part is roughly the size we expect it to be. This is very similar to Lemma 4.7 in [13]. To state the result, we need the following definitions. Let  $G$  be a digraph and let  $\mathcal{P}$  be a partition of  $V(G)$  into an exceptional set  $V_0$  and clusters of equal size. Suppose that  $\mathcal{P}'$  is another partition of  $V(G)$  into an exceptional set  $V'_0$  and clusters of equal size. We say that  $\mathcal{P}'$  is an  $\ell$ -refinement of  $\mathcal{P}$  if  $V_0 = V'_0$  and if the clusters in  $\mathcal{P}'$  are obtained by partitioning each cluster in  $\mathcal{P}$  into  $\ell$  subclusters of equal size. (So if  $\mathcal{P}$  contains  $k$  clusters then  $\mathcal{P}'$  contains  $k\ell$  clusters.)  $\mathcal{P}'$  is an  $\varepsilon$ -uniform  $\ell$ -refinement of  $\mathcal{P}$  with respect to  $G$  if it is an  $\ell$ -refinement of  $\mathcal{P}$  which satisfies the following condition:

(URef) Whenever  $x$  is a vertex of  $G$ ,  $V$  is a cluster in  $\mathcal{P}$  and  $|N_G^+(x) \cap V| \geq \varepsilon|V|$  then  $|N_G^+(x) \cap V'| = (1 \pm \varepsilon)|N_G^+(x) \cap V|/\ell$  for each cluster  $V' \in \mathcal{P}'$  with  $V' \subseteq V$ . The inneighbourhoods of the vertices of  $G$  satisfy an analogous condition.

Let  $\mathcal{G}$  be a collection of digraphs on the same vertex set. If  $\mathcal{P}$  is a refinement with respect to  $G$  for all  $G \in \mathcal{G}$  then we say that it is a refinement *with respect to*  $\mathcal{G}$ .

**Lemma 3.11.** *Suppose that  $0 < 1/m \ll 1/k, \varepsilon \ll \varepsilon', d, 1/\ell, 1/t \leq 1$  and  $m/\ell \in \mathbb{N}$ . Suppose that  $\mathcal{G}$  is a collection of  $t$  digraphs on the same set  $V^*$  of  $n \leq 2km$  vertices and that  $\mathcal{P}$  is a partition of  $V^*$  into an exceptional set  $V_0$  and  $k$  clusters of size  $m$ . Then there exists an  $\varepsilon$ -uniform  $\ell$ -refinement of  $\mathcal{P}$  with respect to  $\mathcal{G}$ . Moreover, any  $\varepsilon$ -uniform  $\ell$ -refinement  $\mathcal{P}'$  of  $\mathcal{P}$  automatically satisfies the following conditions for all  $G \in \mathcal{G}$ :*

- (i) *Suppose that  $V, W$  are clusters in  $\mathcal{P}$  and  $V', W'$  are clusters in  $\mathcal{P}'$  with  $V' \subseteq V$  and  $W' \subseteq W$ . If  $G[V, W]$  is  $(\varepsilon, d)$ -superregular then  $G[V', W']$  is  $(\varepsilon', d)$ -superregular.*

- (ii) Suppose that  $V, W$  are clusters in  $\mathcal{P}$  and  $V', W'$  are clusters in  $\mathcal{P}'$  with  $V' \subseteq V$  and  $W' \subseteq W$ . If  $G[V, W]$  is  $(\varepsilon, d)$ -regular then  $G[V', W']$  is  $(\varepsilon', d)$ -regular.

The proof proceeds by considering a random partition of  $V^*$  (which can be derandomised by Theorem 3.2). We omit the proof as it is almost the same as Lemma 4.7 in [13].

Let  $\varepsilon > 0$  and let  $\mathcal{P}$  be a partition of  $V(G)$  into an exceptional set  $V_0$  and clusters of size  $m$ . Let  $\mathcal{P}'$  be another partition of  $V(G)$  into an exceptional set  $V'_0$  and clusters of size  $m'$  where  $m \geq m'$  and  $|m - m'| \leq 2\varepsilon m'$ . We say that  $\mathcal{P}$  and  $\mathcal{P}'$  are  $\varepsilon$ -close if  $|V_0 \cap V'_0| \geq (1 - \varepsilon)|V'_0|$  and if for each cluster  $U$  in  $\mathcal{P}'$  there is a cluster  $V$  in  $\mathcal{P}$  such that  $|U \cap V| \geq (1 - \varepsilon)m'$ . In this case we say that  $U$  and  $V$  are *associated*. Note that  $V$  is unique when  $\varepsilon < 1/2$ . Suppose that  $R$  is a multidigraph whose vertices are the clusters of  $\mathcal{P}$ . Let  $R'$  be the multidigraph obtained from  $R$  by relabelling  $V$  by  $V'$  for each  $V \in \mathcal{P}$  associated with  $V' \in \mathcal{P}'$ . So  $R'$  has vertex set consisting precisely of the clusters of  $\mathcal{P}'$ . Moreover, for each edge  $E$  from  $U$  to  $V$  in  $R$ , there is a unique edge  $E'$  from  $U'$  to  $V'$  in  $R'$  which is *associated* with  $E$ . The following lemma states that refinements of  $\varepsilon$ -close partitions are still  $\varepsilon'$ -close with a slightly bigger parameter  $\varepsilon'$ .

**Lemma 3.12.** *Suppose that  $0 < 1/m \ll 1/k, \varepsilon_1, \varepsilon_2 \ll \varepsilon', d, 1/\ell \leq 1$  and that  $m/\ell \in \mathbb{N}$ . Suppose that  $G$  is a digraph on  $n \leq 2km$  vertices and that  $\mathcal{P}$  is a partition of  $V(G)$  into an exceptional set  $V_0$  and  $k$  clusters of size  $m$ . Let  $\mathcal{P}'$  be an  $\varepsilon_1$ -uniform  $\ell$ -refinement of  $\mathcal{P}$ . Suppose that  $\mathcal{R}$  is another partition of  $V(G)$  into an exceptional set  $V'_0$  and clusters of size  $m'$  that is  $\varepsilon_2$ -close to  $\mathcal{P}$ . Then, in time polynomial in  $m$ , one can find an  $\varepsilon'$ -uniform  $\ell$ -refinement  $\mathcal{R}'$  of  $\mathcal{R}$  which is  $\varepsilon'$ -close to  $\mathcal{P}'$ .*

*Proof.* Let  $U$  be a cluster of  $\mathcal{P}$  and let  $V$  be the cluster of  $\mathcal{R}$  associated with  $U$ . Then, for each  $U'$  in  $\mathcal{P}'$  such that  $U' \subseteq U$  we have that  $|U' \cap V| \geq m'/\ell - \varepsilon_2 m'$ , so we can pick a subset  $V'$  of  $U' \cap V$  of size exactly  $(1 - \varepsilon_2 \ell)m'/\ell$ . There are now exactly  $\varepsilon_2 \ell m'$  vertices of  $V$  which do not lie in any subcluster  $V'$ . Distribute these among the  $V'$  so that every subcluster has equal size  $m'/\ell$ . Together with  $V'_0$ , these subclusters form the partition  $\mathcal{R}'$ . Clearly  $U'$  and  $V'$  are associated clusters of  $\mathcal{P}'$  and  $\mathcal{R}'$  respectively and  $|U' \cap V'| \geq (1 - \varepsilon')m'/\ell$ . It is easy to see that  $\mathcal{R}'$  has the required properties.  $\square$

Observe that if  $\varepsilon_1 \leq \varepsilon_2$  then any  $\varepsilon_1$ -uniform refinement is also an  $\varepsilon_2$ -uniform refinement, and two  $\varepsilon_1$ -close partitions are also  $\varepsilon_2$ -close.

Let  $\mathcal{P}_2$  denote the partition obtained by taking an  $\varepsilon$ -uniform  $\ell_1$ -refinement  $\mathcal{P}_1$  of a partition  $\mathcal{P}$  and then taking an  $\varepsilon$ -uniform  $\ell_2$ -refinement of  $\mathcal{P}_1$ . Then  $\mathcal{P}_2$  is a  $3\varepsilon$ -uniform  $\ell_2 \ell_1$ -refinement of  $\mathcal{P}$ . Indeed, whenever  $x$  is a vertex of  $G$ ,  $V$  is a cluster in  $\mathcal{P}$  and  $|N_G^+(x) \cap V| \geq \varepsilon|V|$ , then for each cluster  $V' \in \mathcal{P}_2$  with  $V' \subseteq V$ , we have

$$(1) \quad |N_G^+(x) \cap V'| = (1 \pm \varepsilon)^2 |N_G^+(x) \cap V| / \ell_2 \ell_1 = (1 \pm 3\varepsilon) |N_G^+(x) \cap V| / \ell_2 \ell_1,$$

and similarly for the inneighbourhoods.

## 4. TOOLS FOR FINDING SUBGRAPHS, 1-FACTORS AND HAMILTON CYCLES

**4.1. Almost regular spanning subgraphs.** The following result (which is proved as Lemma 5.2 in [13]) shows that in a robust outexpander, we can guarantee a spanning subdigraph with a given degree sequence (as long as the required degrees are not too large and do not deviate too much from each other). If  $x$  is a vertex of a multidigraph  $Q$ , we write  $d_Q^+(x)$  for the number of edges in  $Q$  whose initial vertex is  $x$  and  $d_Q^-(x)$  for the number of edges in  $Q$  whose final vertex is  $x$ .

**Lemma 4.1.** *Let  $q \in \mathbb{N}$ . Suppose that  $0 < 1/n \ll \varepsilon \ll \nu \leq \tau \ll \alpha < 1$  and that  $1/n \ll \rho \leq q\nu^2/3$ . Let  $G$  be a digraph on  $n$  vertices with  $\delta^0(G) \geq \alpha n$  which is a robust  $(\nu, \tau)$ -outexpander. Suppose that  $Q$  is a multidigraph on  $V(G)$  such that whenever  $xy \in E(G)$  then  $Q$  contains at least  $q$  edges from  $x$  to  $y$ . For every vertex  $x$  of  $G$ , let  $n_x^+, n_x^- \in \mathbb{N}$  be such that  $(1-\varepsilon)\rho n \leq n_x^+, n_x^- \leq (1+\varepsilon)\rho n$  and such that  $\sum_{x \in V(G)} n_x^+ = \sum_{x \in V(G)} n_x^-$ . Then  $Q$  contains a spanning submultidigraph  $Q'$  such that  $d_{Q'}^+(x) = n_x^+$  and  $d_{Q'}^-(x) = n_x^-$  for every  $x \in V(G) = V(Q)$ .*

The next result (Lemma 16 in [16]) is an analogue of the previous one where we consider superregular pairs instead of robust outexpanders. In both cases, the proof is algorithmic (as it is based on the Max-Flow-Min-Cut Theorem).

**Lemma 4.2.** *Let  $0 < 1/n \ll \varepsilon \ll \beta \ll \alpha' \ll \alpha \ll 1$ . Suppose that  $G = (A, B)$  is an  $(\varepsilon, \beta + \varepsilon)$ -superregular pair where  $|A| = |B| = n$ . Define  $\kappa := (1-\alpha)\beta n$ . Suppose we have a non-negative integer  $m_a^+ \leq \alpha'\beta n$  associated with each  $a \in A$  and a non-negative integer  $m_b^- \leq \alpha'\beta n$  associated with each  $b \in B$  such that  $\sum_{a \in A} m_a^+ = \sum_{b \in B} m_b^-$ . Then  $G$  contains a spanning subgraph  $H$  in which  $d_H(a) = n_a^+ := \kappa - m_a^+$  for any  $a \in A$  and  $d_H(b) = n_b^- := \kappa - m_b^-$  for any  $b \in B$ .*

**4.2. Decomposing regular digraphs into 1-factors.** Petersen proved that every regular undirected graph can be decomposed into 1-factors. The corresponding result for directed graphs is well known; for completeness we include the proof (which is algorithmic as perfect matchings can be found in polynomial time).

**Proposition 4.3.** *Any  $r$ -regular multidigraph  $G$  contains  $r$  edge-disjoint 1-factors.*

*Proof.* Define an undirected bipartite graph  $J$  with two vertex classes  $A$  and  $B$ , each of which is a copy of  $V(G)$ , with an edge from  $a \in A$  to  $b \in B$  for each edge from  $a$  to  $b$  in  $G$ .  $J$  is  $r$ -regular so, by Hall's Theorem [11], contains a perfect matching  $M_1$ . Then  $J \setminus M_1$  is  $(r-1)$ -regular so contains a perfect matching  $M_2$ . Repeating this procedure we can decompose  $J$  into  $r$  perfect matchings, each of which corresponds to a 1-factor in  $G$ .  $\square$

**4.3. Unwinding cycles.** At two points in the proof, we will partition a blown-up cycle into several longer, thinner blown-up cycles on subclusters of the original clusters. The following section describes how this process is implemented and describes a special approximate decomposition to be used in Section 5.3.

Suppose that  $D = p \otimes C_n$  is a  $p$ -fold blow-up of a cycle  $C_n$  of length  $n$ . Let  $X_1, \dots, X_n$  be the vertex classes of  $D$ . We call any edge-disjoint collection

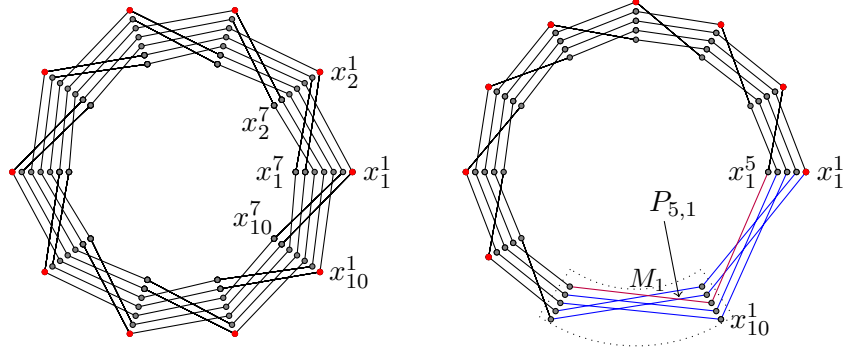


FIGURE 1. Illustrating Lemma 4.4(i) with  $n = 10, p = 7, d = 2$  and Lemma 4.4(ii) with  $n = 10, p = 5, d = 1$ .

$C^1, \dots, C^{p'}$  of  $p'$  Hamilton cycles of  $D$  a  $p'$ -unwinding of  $D$ . The following lemma guarantees a  $(p-1)$ -unwinding in which, for each  $C^d$  and each  $i$ , the  $i$ th vertices of two distinct classes  $X_j$  and  $X_{j'}$  have distance at least  $p$  on  $C^d$ .

**Lemma 4.4.** *Suppose that  $p > 2$  is a prime, suppose  $n \in \mathbb{N}$  and let  $D = p \otimes C_n$  be a  $p$ -fold blow-up of a cycle  $C_n$  of length  $n$ . Denote the vertex classes of  $D$  by  $X_1, \dots, X_n$  where, for all  $j$  with  $1 \leq j \leq n$ , we have  $X_j = \{x_j^1, \dots, x_j^p\}$ . Then  $D$  contains a  $(p-1)$ -unwinding  $C^1, \dots, C^{p-1}$  such that for every  $1 \leq d \leq p-1$  and every  $1 \leq i \leq p$ ,*

- (i) *if  $p$  is coprime to  $n$ , then the vertices  $x_1^i, \dots, x_n^i$  have pairwise distance at least  $p$  on  $C^d$ ;*
- (ii) *if  $p$  is not coprime to  $n$ , then the vertices  $x_1^i, \dots, x_{n-2}^i$  have pairwise distance at least  $p$  on  $C^d$ .*

*Proof.* We first prove (i). Let  $\{a\}$  denote the residue of  $a$  modulo  $p$  and  $[b]$  the residue of  $b$  modulo  $n$  where we adopt the convention that  $\{\ell p\} := p$  and  $[\ell n] := n$  for any  $\ell \in \mathbb{N}$ . For  $1 \leq d \leq p-1$  define the modular arithmetic progression

$$P(d) := (\{1\}, \{1+d\}, \dots, \{1+(np-1)d\})$$

in  $\mathbb{Z}_p$ . For each  $1 \leq k \leq np$  and  $1 \leq d \leq p-1$ , define the edge

$$e_k^d := x_{[k]}^{P(d)_k} x_{[k+1]}^{P(d)_{k+1}},$$

where  $P(d)_k$  denotes the  $k$ th term of  $P(d)$  and  $P(d)_{np+1} := P(d)_1$ . We define  $C^d$  to be the digraph with vertex set  $V(D)$  and edges  $e_1^d, \dots, e_{np}^d$  (see Figure 1). Note that  $C^d$  is clearly a closed walk in  $D$ .

**Claim A.** *For each  $1 \leq d, d' \leq p-1$  and  $1 \leq k, k' \leq np$  the following hold:*

- (a)  *$P(d)$  is periodic with period  $p$ .*
- (b) *Suppose  $P(d)_k = P(d')_{k'}$  and  $P(d)_{k+1} = P(d')_{k'+1}$ . Then  $d = d'$ .*

We first show that the claim implies (i). First note that (a) and the fact that  $n$  is coprime to  $p$  imply that every vertex is visited exactly once in the closed walk  $C^d$ , so  $C^d$  must in fact be a Hamilton cycle. Now suppose  $e_k^d = e_{k'}^{d'}$ . Then (b) implies that also  $d = d'$ . Thus no two  $C^d$  share an edge; thus  $C^1, \dots, C^{p-1}$  is a collection of edge-disjoint Hamilton cycles. (a) implies that, on each  $C^d$ , the distance between  $x_\ell^i$  and  $x_{\ell'}^i$  is a multiple of  $p$  for any  $1 \leq \ell, \ell' \leq n$ . Therefore  $C^1, \dots, C^{p-1}$  have the required property.

To prove (a) of the claim, note that  $P(d)_k = P(d)_{k'}$  if and only if  $1 + kd \equiv 1 + k'd \pmod{p}$  if and only if  $k \equiv k' \pmod{p}$  since  $d$  is coprime to  $p$ . To prove (b), note that  $P(d)_k = P(d')_{k'}$  and  $P(d)_{k+1} = P(d')_{k'+1}$  imply that

$$\begin{aligned} (2) \quad & 1 + kd \equiv 1 + k'd' \pmod{p} \\ (3) \quad & 1 + (k+1)d \equiv 1 + (k'+1)d' \pmod{p}. \end{aligned}$$

Subtracting (2) from (3) gives  $d \equiv d' \pmod{p}$ ; but  $1 \leq d, d' \leq p-1$  so  $d = d'$ . This proves the claim and completes the proof of (i).

We now prove (ii). So suppose instead that  $n$  and  $p$  are not coprime. Then  $n' := n - 2$  is coprime to  $p$  since  $p > 2$ . The idea is to use paths derived from the cycles defined above for the first  $n'$  clusters and extend them into Hamilton cycles via the remaining clusters. To this end, form an auxiliary blown-up cycle  $\tilde{D}$  from  $D$  by identifying  $x_j^i$  with  $x_{j'}^i$ , whenever  $1 \leq i \leq p$  and  $j, j' \in \{n-1, n, 1\}$  and call this vertex  $x_1^i$  in  $\tilde{D}$ . Now remove any resulting loops from  $\tilde{D}$ . So  $\tilde{D} = p \otimes C_{n-2}$ . Next, apply (i) to  $\tilde{D}$  to obtain  $\tilde{C}^1, \dots, \tilde{C}^{p-1}$ . Now, for each  $1 \leq d \leq p-1$ , obtain  $E_1(C^d)$  from  $E(\tilde{C}^d)$  by replacing any edge  $x_{n-2}^i x_1^{i'}$  by  $x_{n-2}^i x_{n-1}^{i'}$ . Note that, in  $D$ ,  $E_1(C^1), \dots, E_1(C^{p-1})$  is an edge-disjoint collection of  $p-1$  paths each of length  $n'$ .

**Claim B.** *The collections  $E_1(C^1), \dots, E_1(C^{p-1})$  of edges can be extended into  $p-1$  edge-disjoint Hamilton cycles  $C^1, \dots, C^{p-1}$  respectively such that  $C^d$  is a subdivision of  $\tilde{C}^d$  for each  $1 \leq d \leq p-1$ .*

To see how this implies the lemma, note that since  $C^d$  is a subdivision of  $\tilde{C}^d$ , the distance between any two vertices in  $C^d$  is at least the distance in  $\tilde{C}^d$ . This immediately gives the required property.

It remains to prove the claim. For each  $1 \leq d \leq p-1$  we will need to find a collection of edge-disjoint paths from  $x_{n-1}^i$  to  $x_1^i$  for  $1 \leq i \leq p$  to extend  $E_1(C^d)$  to  $C^d$ . Moreover, these collections must be pairwise edge-disjoint. By Hall's Theorem, we can find  $p-1$  edge-disjoint perfect matchings  $M_1, \dots, M_{p-1}$  in the complete bipartite subgraph  $(X_{n-1}, X_n)$  of  $D$ . For each  $1 \leq d \leq p-1$  and  $1 \leq i \leq p$ , define

$$P_{i,d} = x_{n-1}^i x_n^{i'} x_1^i$$

whenever  $x_{n-1}^i x_n^{i'}$  is an edge in  $M_d$ . Since the  $M_d$  are edge-disjoint matchings, the  $P_{i,d}$  are edge-disjoint paths with the required property. Thus, for each  $1 \leq d \leq$

$p - 1$ , defining

$$E(C^d) := E_1(C^d) \cup \bigcup_{1 \leq i \leq p} P_{i,d}$$

gives  $p - 1$  edge-disjoint Hamilton cycles  $C^1, \dots, C^{p-1}$ . This proves the claim and completes the proof of (ii).  $\square$

**4.4. Merging 1-factors in blown-up cycles.** In Section 5.10 we will have found an approximate decomposition of a robustly expanding digraph into 1-factors. The following lemma will use the special structure of the 1-factors to merge their cycles into a single Hamilton cycle. It is a special case of Lemma 6.5 in [13], which in turn is based on an idea in [6]. As noted in [13], the cycle guaranteed by the lemma can be found in polynomial time. Roughly speaking, the lemma asserts that if we have a 1-regular digraph  $F$  where most of the edges wind around a ‘blown-up’ cycle  $C = V_1 \dots V_k$ , then under certain circumstances we can turn  $F$  into a (single) cycle by replacing a few edges of  $F$  by edges from a digraph  $G$  whose edges all wind around  $C$ .

**Lemma 4.5.** *Let  $0 < 1/m \ll \varepsilon \ll d < 1$ . Let  $V_1, \dots, V_k$  be pairwise disjoint clusters, each of size  $m$  and let  $C = V_1 \dots V_k$  be a directed cycle on these clusters. Let  $J \subseteq E(C)$ . Let  $G$  be a digraph on  $V_1 \cup \dots \cup V_k$  such that  $G[V_i, V_{i+1}]$  is  $(\varepsilon, d)$ -superregular for every  $V_i V_{i+1} \in J$ . Suppose that  $F$  is a 1-regular digraph with  $V_1 \cup \dots \cup V_k \subseteq V(F)$  such that the following properties hold:*

- (i) *For each edge  $V_i V_{i+1} \in J$  the digraph  $F[V_i, V_{i+1}]$  is a perfect matching.*
- (ii) *For each cycle  $D$  in  $F$  there is some edge  $V_i V_{i+1} \in J$  such that  $D$  contains a vertex in  $V_i$ .*
- (iii) *Whenever  $V_i V_{i+1}, V_j V_{j+1} \in J$  are such that  $J$  avoids all edges in the segment  $V_{i+1} C V_j$  of  $C$  from  $V_{i+1}$  to  $V_j$ , then  $F$  contains a path  $P_{ij}$  joining some vertex  $u_{i+1} \in V_{i+1}$  to some vertex  $u'_j \in V_j$  such that  $P_{ij}$  winds around  $C$ .*

*Then we can obtain a cycle on  $V(F)$  from  $F$  by replacing  $F[V_i, V_{i+1}]$  with a suitable perfect matching in  $G[V_i, V_{i+1}]$  for each edge  $V_i V_{i+1} \in J$ .*

It will also be convenient to use the following result from [15], which guarantees a Hamilton cycle in a robustly expanding digraph. The proof of Lemma 4.5 actually consists of repeated applications of Theorem 4.6 to a suitable auxiliary digraph. The proof of Theorem 4.6 can be made algorithmic but this is not needed here as we only apply it to a ‘reduced’ digraph, obtained from the regularity lemma.

**Theorem 4.6.** *Let  $n_0$  be a positive integer and  $\alpha, \nu, \tau$  be positive constants such that  $1/n_0 \ll \nu \leq \tau \ll \alpha < 1$ . Let  $G$  be a digraph on  $n \geq n_0$  vertices with  $\delta^0(G) \geq \alpha n$  which is a robust  $(\nu, \tau)$ -outexpander. Then  $G$  contains a Hamilton cycle.*

## 5. THE PROOF OF THEOREM 1.1

**5.1. Applying the diregularity lemma.** We choose  $\tau$  so that  $\tau \ll \alpha$ . Without loss of generality we may assume that  $\nu \ll \tau$  as any robust  $(\nu, \tau)$ -outexpander

is also a robust  $(\nu', \tau)$ -outexpander for any  $\nu' \leq \nu$ . We may also assume that  $0 < \eta \ll \nu$  as a collection of  $(1 - \eta')r$  edge-disjoint Hamilton cycles certainly contains a collection of  $(1 - \eta)r$  edge-disjoint Hamilton cycles if  $\eta' \leq \eta$ . Define further constants satisfying

$$(4) \quad \begin{aligned} 0 &< 1/n_0 \ll 1/M \ll 1/M' \ll \tilde{\varepsilon} \ll \varepsilon \ll \varepsilon' \ll \xi \ll 1/p \\ &\ll \beta \ll d \ll 1/s \ll \gamma \ll d' \ll \eta \ll \nu \ll \tau \ll \alpha, \end{aligned}$$

where  $s \in \mathbb{N}$  is even and  $p$  is a prime.

Let  $G$  be a digraph of order  $n \geq n_0$  such that  $G$  is an  $r$ -regular robust  $(\nu, \tau)$ -outexpander with  $r \geq \alpha n$ . Define  $\tilde{\alpha}$  by  $r = \tilde{\alpha}n$ . Apply the diregularity lemma (Lemma 3.3) to  $G$  with parameters  $\tilde{\varepsilon}^{12}, d, M'$  to obtain clusters  $\tilde{V}_1, \dots, \tilde{V}_{\tilde{L}}$  of size  $\tilde{m}$ , an exceptional set  $V_0$ , a pure digraph  $G'$  and a reduced digraph  $\tilde{R}$ . So  $|\tilde{R}| = \tilde{L}$  and  $M' \leq \tilde{L} \leq M$ . We denote the above partition of  $G$  by  $\tilde{\mathcal{P}}$  and call the  $\tilde{V}_j$  the *clusters* of  $\tilde{\mathcal{P}}$ , frequently referred to as *base primary clusters* (to distinguish them from other types of cluster defined later on). Let  $\tilde{R}'$  be the spanning subdigraph of  $\tilde{R}$  whose edges correspond to pairs of density at least  $d'$ . So  $\tilde{V}_i \tilde{V}_j$  is an edge of  $\tilde{R}'$  if  $(\tilde{V}_i, \tilde{V}_j)_{G'}$  has density at least  $d'$ .

When  $\tilde{E}$  is an edge of  $\tilde{R}$  from  $\tilde{V}_i$  to  $\tilde{V}_j$  we write  $G'(\tilde{E})$  for the subdigraph  $(\tilde{V}_i, \tilde{V}_j)_{G'}$  and  $d_{ij}$  for the density of this pair. Then by Lemma 3.3,  $G'(\tilde{E})$  is  $(\tilde{\varepsilon}^{12}, d_{ij})$ -regular. Let  $\tilde{R}(\beta)$  denote the reduced multidigraph of  $G$  (obtained from  $\tilde{R}$ ) with parameters  $\tilde{\varepsilon}^{12}, \beta, d$  and  $M'$ . Let  $\tilde{R}'(\beta)$  be the multidigraph obtained from  $\tilde{R}(\beta)$  by including only those edges which also correspond to an edge of  $\tilde{R}'$ . Roughly speaking, our aim is to find an approximate decomposition of  $\tilde{R}(\beta)$  into edge-disjoint 1-factors  $\tilde{F}$ , and then find an approximate Hamilton decomposition of a subdigraph of  $G$  consisting mainly of edges spanned by pairs in each  $\tilde{F}$ .

For each edge  $\tilde{E}$  of  $\tilde{R}$ , apply Lemma 3.7(i) to  $G'(\tilde{E})$  with parameters  $K := \lfloor d_{ij}/\beta \rfloor$  and  $\gamma_k := \beta$  for each  $1 \leq k \leq K$  to obtain  $K$  edge-disjoint  $(\tilde{\varepsilon}, \beta)$ -regular subdigraphs. We associate each of these with a unique edge  $E$  from  $\tilde{V}_i$  to  $\tilde{V}_j$  of  $\tilde{R}(\beta)$  and call the corresponding digraph  $G'(E)$ .

Let  $A$  be a cluster of  $\tilde{R}$  and let  $E(A)$  denote the set of edges incident to  $A$  in  $\tilde{R}(\beta)$ . For an edge  $E$  in  $E(A)$  and  $x \in A$ , we say that the pair  $(x, E)$  is *good* if

- $A$  is the initial cluster of  $E$  and  $d_{G'(E)}^+(x) = (\beta \pm 2\tilde{\varepsilon})\tilde{m}$ ; or
- $A$  is the final cluster of  $E$  and  $d_{G'(E)}^-(x) = (\beta \pm 2\tilde{\varepsilon})\tilde{m}$

and say it is *bad* otherwise. We say that  $x$  is *good* if  $x$  forms a bad pair with at most  $\xi|E(A)|$  edges in  $E(A)$ . Note that for a fixed edge  $E$  in  $E(A)$ , at most  $\tilde{\varepsilon}\tilde{m}$  vertices  $x \in A$  are bad. So by double counting the number of bad pairs, it is easy to see that the number of bad vertices in  $A$  is at most  $\tilde{\varepsilon}\tilde{m}/\xi$ .

We remove every bad vertex from its cluster as well as possibly some more arbitrary vertices so that exactly  $\tilde{\varepsilon}\tilde{m}/\xi$  vertices have been removed from each cluster. We then remove at most  $2sp$  further vertices from each cluster in order to guarantee that the cluster size is divisible by  $2sp$ . We still denote the cluster size by  $\tilde{m}$  and still call the clusters *base primary*. Each vertex removed here is added

to the exceptional set  $V_0$ , which we now call the *core exceptional set*. So

$$(5) \quad |V_0| \leq (\tilde{\varepsilon}^{12} + \tilde{\varepsilon}/\xi)n + 2sp\tilde{L} \stackrel{(4)}{\leq} \sqrt{\tilde{\varepsilon}}n/2.$$

We still denote the partition of  $V(G)$  into  $V_0$  and these clusters by  $\tilde{\mathcal{P}}$ . Note that for each edge  $E$  of  $\tilde{R}(\beta)$ , the digraph  $G'(E)$  is still  $(\sqrt{\tilde{\varepsilon}}, \beta)$ -regular by Proposition 3.6(i) (at most  $\tilde{\varepsilon}m/4$  vertices were removed from each cluster). Lemma 3.4 implies that

$$(6) \quad \begin{aligned} \delta^0(\tilde{R}') &\geq (\tilde{\alpha} - 3d')\tilde{L} \quad \text{and} \quad \delta^0(\tilde{R}'(\beta)) \geq (\tilde{\alpha} - 4d')\frac{\tilde{L}}{\beta}, \\ \delta^0(\tilde{R}(\beta)) &\geq (\tilde{\alpha} - 4d)\frac{\tilde{L}}{\beta} \quad \text{and} \quad \Delta^0(\tilde{R}(\beta)) \leq (\tilde{\alpha} + 2\tilde{\varepsilon}^{12})\frac{\tilde{L}}{\beta}. \end{aligned}$$

By Lemma 3.5,  $\tilde{R}'$  is a robust  $(\nu/4, 3\tau)$ -outexpander. Note that it is a subdigraph of  $\tilde{R}'(\beta) \subseteq \tilde{R}(\beta)$  and that all of its edges have multiplicity at least  $q := d'/\beta$  in  $\tilde{R}'(\beta)$ . Let

$$(7) \quad \tilde{r} := (\tilde{\alpha} - \gamma)\tilde{L}/\beta.$$

Let  $n_U^\pm := d_{\tilde{R}(\beta)}^\pm(U) - \tilde{r}$  and let  $\rho := \gamma/\beta$ , so  $\rho \leq q\nu^2/3$ . Note that

$$(1 - \frac{4d}{\gamma})\rho\tilde{L} = (\gamma - 4d)\frac{\tilde{L}}{\beta} \leq n_U^\pm \leq (\gamma + 2\tilde{\varepsilon}^{12})\frac{\tilde{L}}{\beta} = (1 + \frac{2\tilde{\varepsilon}^{12}}{\gamma})\rho\tilde{L}.$$

So we can apply Lemma 4.1 to  $(G, Q) := (\tilde{R}', \tilde{R}'(\beta))$  to obtain a sub-multidigraph  $W$  of  $\tilde{R}'(\beta)$  (and hence of  $\tilde{R}(\beta)$ ) such that the in- and outdegrees of each cluster  $U$  are exactly  $n_U^\pm$ . So  $\tilde{R}(\beta) \setminus W$  is a spanning  $\tilde{r}$ -regular sub-multidigraph of  $\tilde{R}(\beta)$ . Apply Proposition 4.3 to decompose  $\tilde{R}(\beta) \setminus W$  into  $\tilde{r}$  1-factors  $\tilde{F}_1, \dots, \tilde{F}_{\tilde{r}}$  of  $\tilde{R}(\beta)$ . So each  $\tilde{F}_t$  corresponds to a collection of blown-up cycles spanning  $V(G) \setminus V_0$ . Note that this step would not work if we only considered  $\tilde{R}$  and  $\tilde{R}(\beta)$  and tried to apply Lemma 4.1 to find  $W$  in  $\tilde{R}(\beta)$  directly.

**5.2. Thin auxiliary digraphs  $H$ .** We now define edge-disjoint subdigraphs  $H_0^+$ ,  $H_0^-$ ,  $H_1^+$ ,  $H_1^-$  and  $H_2$  of  $G$ , which are sparse ‘shadows’ of the reduced multidigraph. They act as reservoirs of well-distributed edges which will be used at various stages in the proof. The role of  $H_0^\pm$  is to connect blown-up cycles (in Section 5.5) to ensure that our final merging procedure does indeed yield Hamilton cycles. In Section 5.6 edges will be taken from  $H_1^\pm$  to connect the vertices in the special exceptional sets  $V_{0,i}$  (defined later) to the non-exceptional vertices in each slice  $G_i$  (defined in Section 5.3).  $H_2$  will be used to construct ‘balancing edges’ which will be introduced in Section 5.8. We choose these subdigraphs already at this point because if we remove them later then this might destroy the superregularity of the pairs in the  $G_i$ .

We obtain  $H_0^+, H_0^-, H_1^+, H_1^-, H_2$  as follows. Each has vertex set  $V(G)$  and initially contains no edges. Then, for each edge  $E$  of  $\tilde{R}(\beta)$ ,  $G'(E)$  is a  $(\sqrt{\tilde{\varepsilon}}, \beta)$ -regular pair and we can apply Lemma 3.7(i) to  $G'(E)$  with  $\gamma_1 := \beta_1$  where

$$(8) \quad \beta_1 := (1 - 5\gamma)\beta$$

and  $\gamma_2 := \dots = \gamma_6 := \gamma\beta$ , to obtain six edge-disjoint pairs  $J_1, \dots, J_6$ , where  $J_k$  is  $(\tilde{\varepsilon}^{1/24}, \gamma_k)$ -regular, and we call these digraphs  $G^*(E)$ ,  $H_0^+(E)$ ,  $H_0^-(E)$ ,  $H_1^+(E)$ ,  $H_1^-(E)$  and  $H_2(E)$  respectively. We denote the union of  $H(E)$  over all edges  $E$  of  $\tilde{R}(\beta)$  by  $H$ . We will only use the weaker bounds that the ‘remaining’ subdigraph  $G^*(E)$  of  $G'(E)$  is  $(\varepsilon/8, \beta_1)$ -regular and for each  $H = H_0^+, H_0^-, H_1^+, H_1^-, H_2$  we have that  $H(E)$  is  $(\varepsilon, \gamma\beta)$ -regular. Moreover, Lemma 3.7(i) implies that if  $E$  is an edge from  $A$  to  $B$  and if  $x \in A$  and  $y \in B$  are good for  $E$ , then

$$(9) \quad d_{H(E)}^+(x), d_{H(E)}^-(y) = (\gamma\beta \pm 2\varepsilon)\tilde{m}.$$

Note also that  $V_0$  is isolated in each  $H$ . We now derive some further properties of these digraphs which we will need later. Firstly, we have the following property for  $H_0^+$  and  $H_0^-$ :

(H0) Suppose that  $\tilde{A}\tilde{B}$  is an edge of  $\tilde{R}$ . Then for at least  $(1 - \varepsilon')|\tilde{A}|$  of the vertices  $x \in \tilde{A}$  and  $(1 - \varepsilon')|\tilde{B}|$  of the vertices  $y \in \tilde{B}$  we have

$$|N_{H_0^+}^+(x) \cap \tilde{B}| \geq \gamma d\tilde{m}/2 \quad \text{and} \quad |N_{H_0^-}^-(y) \cap \tilde{A}| \geq \gamma d\tilde{m}/2.$$

To see this, note first that every edge  $E$  of  $\tilde{R}$  has multiplicity at least  $d/\beta$  in  $\tilde{R}(\beta)$ . Let  $E_1, \dots, E_\ell$  be the edges of  $\tilde{R}(\beta)$  corresponding to  $E$ . So  $d/\beta \leq \ell \leq 1/\beta$ . Recall that  $H_0^+(E_i)$  is  $(\varepsilon, \gamma\beta)$ -regular. Let  $A'$  be the set of all vertices  $x \in \tilde{A}$  such that  $x$  has outdegree at least  $(\gamma\beta - 2\varepsilon)\tilde{m}$  in each of  $H_0^+(E_1), \dots, H_0^+(E_\ell)$ . Then  $|A'| \geq (1 - \ell\varepsilon)\tilde{m} \geq (1 - \varepsilon')\tilde{m}$ . Moreover, for all  $x \in A'$ , we have

$$|N_{H_0^+}^+(x) \cap \tilde{B}| \geq \ell(\gamma\beta - 2\varepsilon)\tilde{m} \geq \frac{d}{\beta} \frac{\gamma\beta}{2} \tilde{m} \geq \frac{\gamma d\tilde{m}}{2}.$$

The proof of the second inequality is similar.

We also have the following property of  $H_1^+$  and  $H_1^-$ :

(H1) For all  $x \in V(G) \setminus V_0$ , we have  $\gamma\tilde{\alpha}n/3 \leq d_{H_1^+}^\pm(x), d_{H_1^-}^\pm(x) \leq 2\gamma\tilde{\alpha}n$ .

(H1) follows from the fact that  $V_0$  contains all the bad vertices (in the sense of Section 5.1). Indeed, since any vertex  $x \in V(G) \setminus V_0$  is good we have

$$d_{H_1^+}^+(x) \stackrel{(9)}{\geq} \delta^+(\tilde{R}(\beta))(1 - \xi)(\gamma\beta - 2\varepsilon)\tilde{m} \stackrel{(6)}{\geq} \frac{\tilde{\alpha}\tilde{L}}{2\beta} \gamma\beta\tilde{m} \geq \gamma\tilde{\alpha}n/3.$$

The other bounds in (H1) follow similarly.

**5.3. Unwinding cycles.** For each  $1 \leq t \leq \tilde{r}$  we now apply Lemma 3.8 to each cycle in  $\tilde{F}_t$  to remove vertices from each cluster, so that they now have size exactly  $(1 - \varepsilon/4)\tilde{m}$  and such that each edge  $E$  of  $\tilde{F}_t$  corresponds to an  $(\varepsilon/2, \beta_1)$ -superregular pair  $G^*(E)$ . By removing at most  $2sp$  further vertices from each cluster we obtain clusters of size  $m$  such that  $2sp \mid m$ . We call these *adapted primary clusters* or *adapted primary (t)-clusters* if we wish to emphasise the dependence on  $t$ , and say that each such cluster is *associated* with the base primary cluster from which it was formed. Since  $2sp \leq \varepsilon\tilde{m}/4$  it is easy to see that now each edge  $E$  of  $\tilde{F}_t$

corresponds to an  $(\varepsilon, \beta_1)$ -superregular pair  $G^*(E)$ . Note that

$$(10) \quad \frac{1}{m} \leq \frac{2\tilde{L}}{n} \leq \frac{2M}{n_0} \ll \frac{1}{\tilde{L}} \quad \text{and} \quad (1 - \varepsilon)n \stackrel{(5)}{\leq} m\tilde{L} \leq \tilde{m}\tilde{L} \leq n.$$

Let  $\tilde{V}_{0,t}^{\text{spec}}$  denote the set of all those vertices in  $G$  which were removed from the clusters in this step. We call them the *special exceptional vertices* (for the original slice  $t$ ). So  $|\tilde{V}_{0,t}^{\text{spec}}| \leq \varepsilon n/4 + 2sp\tilde{L} \leq \varepsilon n/2$ . Let  $\tilde{V}_{0,t} = V_0 \cup \tilde{V}_{0,t}^{\text{spec}}$ . Then

$$(11) \quad |\tilde{V}_{0,t}| \stackrel{(5)}{\leq} \frac{2\varepsilon n}{3}.$$

We denote the collection of the adapted primary  $(t)$ -clusters together with the exceptional set  $\tilde{V}_{0,t}$  by  $\mathcal{P}(t)$ . Note that  $\mathcal{P}(t)$  and  $\tilde{\mathcal{P}}$  are  $2\varepsilon/3$ -close for every  $1 \leq t \leq \tilde{r}$  (recall that this notion was defined before Lemma 3.12).

For each cycle  $C$  in a given 1-factorisation, we would like to ensure that the outneighbourhood of an exceptional vertex is well-distributed on each cycle, in the sense that each cluster  $V$  of  $C$  contains only a small fraction of the neighbours of any exceptional vertex. Currently, we cannot guarantee this. But we will be able to achieve this property by considering a refinement of the partition  $\mathcal{P}(t)$  for each  $t$ . As associated clusters in each  $\mathcal{P}(t)$  only differ slightly from one another, we can find this refinement in such a way that the subclusters are also similar by ensuring that all such refinements are close to a refinement of  $\tilde{\mathcal{P}}$ .

Let  $\mathcal{G} = \{G, H_0^+, H_0^-, H_1^+, H_1^-, H_2\}$ . We now apply Lemma 3.11 to our base primary clusters and exceptional set  $V_0$  to obtain an  $\tilde{\varepsilon}$ -uniform  $s$ -refinement  $\mathcal{P}'_s$  of  $\tilde{\mathcal{P}}$  with respect to  $\mathcal{G}$ , and we call the resulting subclusters *base  $s$ -clusters*. So we have  $L_s := s\tilde{L}$  base  $s$ -clusters. Apply Lemma 3.11 to  $\mathcal{P}'_s$  to obtain an  $\tilde{\varepsilon}$ -uniform  $p$ -refinement  $\mathcal{P}'_p$  of  $\mathcal{P}'_s$  with respect to  $\mathcal{G}$ . Let

$$(12) \quad L_p := pL_s = sp\tilde{L}.$$

We call the  $L_p$  subclusters obtained from an  $s$ -cluster *base  $p$ -clusters*. By the remark before (1),  $\mathcal{P}'_p$  is also a  $3\tilde{\varepsilon}$ -uniform  $sp$ -refinement of  $\tilde{\mathcal{P}}$ . Finally apply Lemma 3.11 to  $\mathcal{P}'_p$  to obtain an  $\tilde{\varepsilon}$ -uniform 2-refinement  $\mathcal{P}'_{2p}$  of  $\mathcal{P}'_p$  with respect to  $\mathcal{G}$ . The argument before (1) implies that  $\mathcal{P}'_{2p}$  is a  $4\tilde{\varepsilon}$ -uniform  $2p$ -refinement of  $\mathcal{P}'_s$  and a  $5\tilde{\varepsilon}$ -uniform  $2sp$ -refinement of  $\tilde{\mathcal{P}}$ . We call the subclusters obtained from an  $s$ -cluster *base  $2p$ -clusters*.

Define constants  $\varepsilon_s, \varepsilon_p, \varepsilon_{2p}$  such that  $\varepsilon \ll \varepsilon_s \ll \varepsilon_p \ll \varepsilon_{2p} \ll \varepsilon'$ . Now do the following for each  $t$  with  $1 \leq t \leq \tilde{r}$ . Apply Lemma 3.12 to  $\tilde{\mathcal{P}}$  to obtain an  $\varepsilon_s$ -uniform  $s$ -refinement  $\mathcal{P}_s(t)$  of  $\mathcal{P}(t)$  that is  $\varepsilon_s$ -close to  $\mathcal{P}'_s$ . Next apply Lemma 3.12 to  $\mathcal{P}'_s$  to obtain an  $\varepsilon_p$ -uniform  $p$ -refinement  $\mathcal{P}_p(t)$  of  $\mathcal{P}_s(t)$  that is  $\varepsilon_p$ -close to  $\mathcal{P}'_p$ . By the observation at the end of Section 3.4,  $\mathcal{P}_p(t)$  is also an  $\varepsilon'$ -uniform  $sp$ -refinement of  $\mathcal{P}(t)$ . Finally apply Lemma 3.11 to  $\mathcal{P}'_p$  to obtain an  $\varepsilon_{2p}$ -uniform 2-refinement  $\mathcal{P}_{2p}(t)$  of  $\mathcal{P}_p(t)$  that is  $\varepsilon_{2p}$ -close to  $\mathcal{P}'_{2p}$ . Again,  $\mathcal{P}_{2p}(t)$  is an  $\varepsilon'$ -uniform  $2p$ -refinement of  $\mathcal{P}_s(t)$  and an  $\varepsilon'$ -uniform  $2sp$ -refinement of  $\mathcal{P}(t)$ . For  $j = s, p, 2p$  we call the clusters of  $\mathcal{P}_j(t)$  the (*adapted*)  $j$ -clusters or  $j$ -( $t$ )-clusters if we wish to emphasise the dependence on  $t$ . For each such  $j$  we have that  $\mathcal{P}_j(t)$  is an  $\varepsilon'$ -uniform refinement

of  $\mathcal{P}(t)$  that is  $\varepsilon'$ -close to  $\mathcal{P}'_j$ , so each adapted  $j$ -cluster in  $\mathcal{P}_j(t)$  is associated with a unique base  $j$ -cluster in  $\mathcal{P}'_j$ . Write

$$(13) \quad m_s := m/s \quad \text{and} \quad m_p := m/sp$$

for the respective sizes of the  $s$ - and  $p$ -clusters (which are the same for all  $t$  though the clusters themselves are different). Note that

$$(14) \quad m_p \leq \frac{n}{L_p} \leq 2m_p.$$

By a slight abuse of notation we can consider  $\tilde{R}$  and  $\tilde{R}(\beta)$  as digraphs on either base or adapted  $(t)$ -clusters, depending on the context. For each  $0 \leq t \leq \tilde{r}$ , we now define corresponding reduced digraphs for the refinements defined above, where, for convenience,  $\mathcal{P}_j(0) := \mathcal{P}'_j$ .

Let  $R_s = s \otimes \tilde{R}$  be the  $s$ -fold blow-up of  $\tilde{R}$ , where for a vertex  $W$  of  $\tilde{R}$  (which is an adapted primary  $(t)$ -cluster if  $t \geq 1$ ), the corresponding vertices in  $R_s$  are the subclusters of  $W$  in  $\mathcal{P}_s(t)$ . Define  $R_s(\beta) = s \otimes \tilde{R}(\beta)$  analogously. Also let  $R_p = p \otimes R_s$ , where for a vertex  $U$  of  $R_s$  the corresponding vertices in  $R_p$  are the subclusters of  $U$  in  $\mathcal{P}_p(t)$ . So the vertices of  $R_p$  are precisely the  $p$ -clusters and also  $R_p = sp \otimes \tilde{R}$ . Define  $R_p(\beta) = p \otimes R_s(\beta) = sp \otimes \tilde{R}(\beta)$  analogously. Note that apart from the fact that the clusters which form their vertex sets are slightly different for different values of  $t$ , these digraphs are the same, so there is no need for any dependence on  $t$  in the notation.

Suppose that  $\tilde{E}$  is an edge of  $\tilde{R}(\beta)$  from  $\tilde{U}$  to  $\tilde{W}$  and that  $U$  is an  $s$ -cluster which is a subcluster of  $\tilde{U}$  and  $W$  is an  $s$ -cluster which is a subcluster of  $\tilde{W}$ . Note that there is a unique edge  $E$  in  $R_s(\beta)$  from  $U$  to  $W$  corresponding to  $\tilde{E}$ . Thus to each edge  $E$  of  $R_s(\beta)$  we can associate the digraph

$$(15) \quad G^*(E) := G^*(\tilde{E})[U, W].$$

We make a similar association for each edge  $F$  of  $R_p(\beta)$  by defining  $G^*(F)$  analogously.

We now use the 1-factors  $\tilde{F}_t$  to define edge-disjoint 1-factors  $F'_j$  in the reduced digraph  $R_s(\beta)$  and then use the  $F'_j$  to find edge-disjoint 1-factors  $F_i$  in  $R_p(\beta)$ . Note that each cycle  $C$  of  $\tilde{F}_t$  corresponds to an  $s$ -fold blow-up  $C'$  of  $C$  in  $R_s(\beta)$ . So for each cycle  $C$  in  $\tilde{F}_t$  we can apply Lemma 4.4 to obtain an  $(s-1)$ -unwinding  $C_1, \dots, C_{s-1}$  of  $C'$ . Here we do not need the special properties of the  $(s-1)$ -unwinding which are guaranteed by Lemma 4.4; in fact any unwinding yielding edge-disjoint Hamilton cycles will do. So  $\tilde{F}_t$  corresponds to a set of  $(s-1)$  1-factors  $F'_j$  (with  $(t-1)(s-1) + 1 \leq j \leq t(s-1)$ ) of  $R_s(\beta)$ . We say that such an  $F'_j$  has *original factor type*  $t$  (and that  $t$  is the *original type* of  $j$ ). Note that for each cluster  $W$  of  $\tilde{R}$ , there are  $s$  clusters of  $F'_j$  which are subclusters of  $W$ . Moreover, all of these lie on the same cycle of  $F'_j$ . Let

$$(16) \quad r_s := (s-1)\tilde{r}.$$

Then altogether this gives us a set of  $r_s$  edge-disjoint 1-factors  $F'_1, \dots, F'_{r_s}$  of  $R_s(\beta)$ .

Consider any cycle  $C_\ell$  of a 1-factor  $F'_j$  obtained from a cycle  $C$  of  $\tilde{F}_t$  as above. Let  $K$  be the length of  $C$ ; so  $C_\ell$  has length  $Ks$ . We say that an  $s$ -cluster lying on  $C_\ell$  is *clean for  $F'_j$*  if it belongs to the last  $K$  clusters of  $C_\ell$  (where for each cycle we pick a consistent ordering of its vertices in advance). Note that  $K \geq 2$  and so  $C_\ell$  has at least two clean  $s$ -clusters. Moreover, for each adapted primary cluster  $W$ , exactly one subcluster of  $W$  in  $\mathcal{P}_s(t)$  is clean for  $F'_j$ . Note that for different 1-factors  $F'_j$ , the set of clean clusters will usually be different.

It turns out that we actually need a stronger property than the one described above, namely we need that  $(\star)$  below holds. (This will enable us to ensure that, in the digraphs  $G_i$  that we consider later, only a few clusters will contain vertices sending or receiving an edge from the exceptional set and these will be sufficiently far apart.) For this, we use our refinement  $\mathcal{P}_p(t)$  of each  $s$ -cluster into  $p$  subclusters and unwind the cycles in the above 1-factorisation again.

For every  $V \in \mathcal{P}_s(t)$ , let  $V^1, \dots, V^p$  be the  $p$ -clusters contained in  $V$ . Note that the collection of all  $V^k$  over all  $s$ -clusters  $V$  contained in an adapted primary cluster  $W$  are precisely the  $p$ -clusters refining  $W$ . For each cycle  $D = V_1 \dots V_K$  in  $F'_j$  (where this is the same ordering we specified above) let  $D'$  be the  $p$ -fold blow-up of  $D$  whose vertex classes are the  $p$ -clusters  $V_\ell^k$  contained in  $V_1, \dots, V_K$ . Apply Lemma 4.4 to  $D'$  to find a  $(p-1)$ -unwinding  $D_1, \dots, D_{p-1}$  of  $D'$  with  $V_\ell^k$  playing the role of  $x_\ell^k$ . We have the following property:

- ( $\star$ ) For each  $1 \leq d \leq p-1$  and  $1 \leq k \leq p$ , the  $p$ -clusters  $V_1^k, \dots, V_{K-2}^k$  have pairwise distance at least  $p$  on  $D_d$ .

Note that  $(\star)$  holds only for the  $p$ -clusters in  $V_1, \dots, V_{K-2}$  and not necessarily  $V_{K-1}$  or  $V_K$ . This is the reason for introducing the clean  $s$ -clusters: recall that  $V_{K-1}$  and  $V_K$  are clean. This will mean that we will never introduce any edges between their vertices and the exceptional set (see (b) in Section 5.4).

Moreover, for all  $1 \leq \ell \leq K$ , we have from Lemma 4.4 that  $V_\ell^1, \dots, V_\ell^p$  lie on the same cycle  $D_d$ . Additionally, their successors on  $D_d$  all belong to a single adapted primary cluster  $V_{\ell+1}$ . Also  $F'_j$  corresponds to a set of  $(p-1)$  edge-disjoint 1-factors  $F_i$  (with  $(j-1)(p-1)+1 \leq i \leq j(p-1)$ ) of  $R_p(\beta)$ . We say that such an  $F_i$  has *intermediate factor type  $j$*  and *original factor type  $t$*  where  $t$  is the original type of  $F'_j$ . For each  $i$ , write  $V_{0,i}^{\text{spec}} := \tilde{V}_{0,t}^{\text{spec}}$  for the special exceptional set associated with  $F_i$ , where  $t$  is the original factor type of  $F_i$ . Note that for every  $i$ , every vertex in  $G$  is contained either in a  $p$ -cluster of  $F_i$ , in  $V_0$  or in  $V_{0,i}^{\text{spec}}$ . Note also that for each adapted primary cluster  $W$  of  $\tilde{R}$ , there are  $sp$  clusters of  $F_i$  which are subclusters of  $W$ . Also let

$$(17) \quad r_p := (p-1)r_s.$$

Then altogether this gives us a set of  $r_p$  edge-disjoint 1-factors  $F_1, \dots, F_{r_p}$  of  $R_p(\beta)$ . Note that for each  $t$ , there are exactly  $(s-1)(p-1)$  of the  $F_i$  which have original factor type  $t$ . Furthermore,

$$(18) \quad r_p \stackrel{(7)}{\leq} \frac{\tilde{\alpha} sp \tilde{L}}{\beta} \quad \text{and so} \quad r_p \stackrel{(12)}{\leq} \frac{\tilde{\alpha} L_p}{\beta}; \quad \text{also} \quad 1/m \leq 1/m_p \stackrel{(10)}{\ll} 1/r_p.$$

For each edge  $E \in E(F_i)$  from  $A$  to  $B$ , let  $G_i(E) := G^*(E)$ , where  $G^*(E)$  was defined just after (15). Let  $G_i$  denote the union of the digraphs  $G_i(E)$  over all  $E$  with  $E \in E(F_i)$  and call it the  $i$ th *slice*. Clearly  $G_1, \dots, G_{r_p}$  are edge-disjoint. Given  $E \in E(F_i)$  for some  $1 \leq i \leq r_p$ , let  $\tilde{E} \in E(\tilde{F}_t)$  be the unique edge such that  $E$  is in the blow-up of  $\tilde{E}$ , where  $F_i$  has original factor type  $t$ . As noted directly before (10),  $G^*(\tilde{E})$  is  $(\varepsilon, \beta_1)$ -superregular and hence  $G_i(E) = G^*(E)$  is  $(\varepsilon', \beta_1)$ -superregular by Lemma 3.11(i).

Recall that since  $V_{0,i}^{\text{spec}}$  is different for each  $i$ , the vertex set of a  $p$ -cluster will be slightly different in  $G_i$  and  $G_{i'}$  when  $F_i$  and  $F_{i'}$  have different original factor types. Note that if  $U$  is a base  $2p$ -cluster (of size  $m_p/2$ ), and  $U_{(t)}$  is the associated  $2p(t)$ -cluster, then

$$(19) \quad |U \cap U_{(t)}| \geq (1 - \varepsilon')m_p/2.$$

as the corresponding partitions are  $\varepsilon'$ -close. (On the other hand,  $\bigcap_{1 \leq t \leq \tilde{r}} U_{(t)}$  may be empty.) The same statements hold for  $s$ - and  $p$ -clusters. When adding edges incident to exceptional vertices in Section 5.6 we need to be careful about distinguishing between base  $2p$ -clusters and the  $2p(t)$ -clusters which are actually contained in the clusters of our slices.

**5.4. Red clusters and edges.** The aim of this Section is to lay some groundwork for Sections 5.5, 5.6 and 5.7 by specifying the properties that the edges between the exceptional vertices and the rest of  $V(G)$  need to satisfy. In Section 5.5 our aim is to remove a bounded number of *bridge vertices*  $V_{0,i}^{\text{bridge}}$  from each  $G_i \setminus (V_0 \cup V_{0,i}^{\text{spec}})$  and change their neighbourhoods in such a way that the blown-up cycles in  $G_i$  are connected via bridge vertices. Some additional vertices will also be removed and added to  $V_{0,i}^{\text{spec}}$  to keep the cluster sizes equal. In Section 5.6 we will add edges to  $G_i$  which are incident to  $V_0$ . In Section 5.7 we will do the same for  $V_{0,i}^{\text{spec}}$ . We will then let

$$V_{0,i} := V_0 \cup V_{0,i}^{\text{spec}} \cup V_{0,i}^{\text{bridge}}.$$

$V_{0,i}$  is then the exceptional set for the slice  $G_i$ : each vertex will lie either in a cluster of  $G_i$  or in  $V_{0,i}$ . Any edge incident to a vertex in  $V_{0,i}$  and any vertex in a cluster of  $G_i$  incident to such an edge will be called *i-red* (or *red* if this is unambiguous). Roughly speaking, when adding red edges to  $G_i$ , we will need to ensure that  $G_i$  is a spanning almost-regular digraph, that no non-exceptional vertex has large *i-red* degree and that the set of red vertices is small and well-distributed.

To achieve this, for each  $i$  we will only add red edges incident to some carefully selected  $2p$ -clusters and then apply property  $(\star)$ . More precisely, for fixed  $i$ , let  $j = j(i)$  and  $t = t(i)$  respectively be the intermediate and original factor types of  $G_i$ . For each  $s$ -cluster  $U$  of  $\mathcal{P}_s(t)$ , let  $U_1, \dots, U_p$  denote the  $p$ -clusters of  $\mathcal{P}_p(t)$  which are subclusters of  $U$ . For  $1 \leq \ell \leq p$ , let  $U(\ell)$  and  $U(2\ell)$  be the  $2p$ -clusters contained in  $U_\ell$ . In  $G_i$ , we will add red edges between  $V_{0,i}$  and  $U(k)$  only if

- (a)  $t \equiv k \pmod{2p}$  and
- (b)  $U$  is not a clean cluster in  $F'_j$ .

We call such a  $2p$ -cluster  $U(k)$  *i-red* and we call a  $p$ -cluster *i-red* if it contains an *i-red*  $2p$ -cluster (or simply *red* if this is unambiguous). Note that (a) implies that every  $s$ -cluster  $U$  which is not clean contains exactly one red  $2p$ -cluster. Thus  $U$  contains exactly one red  $p$ -cluster. Moreover, recall that any adapted primary cluster contains exactly one clean  $s$ -cluster; thus it contains exactly  $s - 1$  red  $p$ -clusters.

All red vertices will be contained in red  $2p$ -clusters, but note that we do not require every red cluster to contain a red vertex. Let

$$(20) \quad \kappa := (1 - \gamma)\beta_1 m_p.$$

We would like to find exactly  $\kappa$  edge-disjoint Hamilton cycles in each of the  $G_i$ . For this, we will first need to add edges so that  $G_i$  satisfies the following for all  $i$  with  $1 \leq i \leq r_p$ :

- (Red0) There exists a sequence  $D_1 x_1 D_2 x_2 \dots x_{\ell-1} D_{\ell} x_{\ell} D_1$  with the following properties:
  - Each  $D_j$  is a cycle of  $F_i$  and every cycle of  $F_i$  appears at least once in the sequence;
  - $V_{0,i}^{\text{bridge}} := \{x_1, \dots, x_{\ell}\}$  and each  $x_j$  has exactly  $\kappa$  outneighbours in  $D_{j+1}$  and exactly  $\kappa$  inneighbours in  $D_j$ .
- (Red1)  $d_{G_i}^{\pm}(x) = \kappa$  for all  $x \in V_{0,i}$ ;
- (Red2)  $V_{0,i}$  is an independent set in  $G_i$ ;
- (Red3)  $|N_{G_i}^{\pm}(y) \cap V_{0,i}| \leq \sqrt{\xi}\beta_1 m_p$  for all  $y \notin V_{0,i}$ ;
- (Red4) For every red  $p$ -cluster  $V$ , all red edges of  $G_i$  are incident to a single  $2p$ -cluster contained in  $V$ . In particular,  $|N_{G_i}^{\pm}(V_{0,i}) \cap V| \leq m_p/2$  for all clusters  $V \in R_p$ ;
- (Red5) If  $V, V'$  are red  $p$ -clusters on a cycle  $C$  of  $F_i$ , then they have distance at least  $p$  on  $C$ ;
- (Red6) If a  $p$ -cluster  $V$  contains the final vertex of a red edge in  $G_i$ , then it contains no initial vertices of red edges in  $G_i$ , and vice versa;
- (Red7)  $G_1, \dots, G_{r_p}$  are edge-disjoint and  $G_i(E)$  is  $(2\varepsilon', \beta_1)$ -superregular for all  $E \in E(F_i)$ .

Roughly speaking, given a 1-factor  $f$  of  $G_i$ , (Red0) and (Red1) will ensure that  $f$  has a path between any pair of successive cycles  $D_j$  of  $F_i$ . (Red2)–(Red6) imply that the red edges are well-distributed. This will be crucial when applying Lemma 4.5 to transform  $f$  into a Hamilton cycle in Section 5.11.

Suppose that  $V$  is a red  $p$ -cluster. Since only one of the two  $2p$ -clusters contained in  $V$  is red, it follows that (Red4) will automatically be satisfied for  $V$  if we add red edges according to (a) and (b). It is easy to see that this will also satisfy (Red5). Indeed, recall that every non-clean  $s$ -cluster contains exactly one red  $p$ -cluster. Moreover, if  $U_{\ell}$  and  $U_{\ell'}$  are non-clean  $s$ -clusters then the  $p$ -cluster  $U_{\ell}^k$  is red if and only if  $U_{\ell'}^k$  is red. Suppose a cycle  $C$  in  $F_i$  was obtained by unwinding the blow-up of  $C'$  in  $F_j'$ ; then the last two  $s$ -clusters in  $C'$  (using the same ordering

as in Section 5.3) are clean and hence contain no red  $p$ -clusters by (b). So  $(\star)$  implies that the red clusters on  $C$  will have distance at least  $p$  apart.

Let  $F = r_p/4p$ . For each  $k$  with  $1 \leq k \leq 2p$ , note that the number of all digraphs  $G_i$  whose original type  $t$  satisfies (a) is  $2F$ . For each  $k$ , consider an ordering of all these graphs.

We will now fix which of the (red)  $2p$ -clusters will receive red edges and which of them will send out red edges. For each  $i = 1, \dots, r_p$ , let  $t = t(i)$  be the original type of  $G_i$  and let  $k$  with  $1 \leq k \leq 2p$  satisfy (a). Suppose that  $G_i$  is the  $f$ th graph with original type  $t$  (so  $1 \leq f \leq 2F$ ). For each adapted primary cluster  $W \in \mathcal{P}(t)$ , let  $W_1, \dots, W_s$  denote the set of all  $s$ -clusters contained in  $W$ . Recall that exactly one of the  $W_j$  is clean. Now choose a set  $S_W^+$  of  $s$ -clusters from  $W_1, \dots, W_{s/2}$  so that none of the  $s$ -clusters in  $S_W^+$  is clean and so that  $|S_W^+| = s/2 - 1$  (recall that  $s$  is even). Let  $I_W^+$  denote the set of indices of the  $s$ -clusters in  $S_W^+$ . Similarly choose  $S_W^-$  from  $W_{s/2+1}, \dots, W_s$  with  $|S_W^-| = s/2 - 1$  which avoids the clean cluster and let  $I_W^-$  be the corresponding set of indices.

For each  $s$ -cluster  $W_\ell$  contained in  $W$ , let  $W_\ell(k)$  denote the  $k$ th  $2p$ -cluster contained in  $W_\ell$ , where  $k$  is defined as in the previous paragraph. We call  $W_\ell(k)$  *in-red* (for  $i$ ) if

- $1 \leq f \leq F$  and  $\ell \in I_W^+$ , or  $F < f \leq 2F$  and  $\ell \in I_W^-$ .

We call  $W_\ell(k)$  *out-red* (for  $i$ ) if

- $1 \leq f \leq F$  and  $\ell \in I_W^-$ , or  $F < f \leq 2F$  and  $\ell \in I_W^+$ .

If a  $p$ -cluster  $V$  contains an in-red  $2p$ -cluster, we say that  $V$  is *in-red* (and similarly for *out-red* clusters). So the number of in-red  $p$ -clusters in each adapted primary cluster  $W$  is exactly

$$(21) \quad |I_W^-| = \frac{s}{2} - 1$$

and similarly for out-red clusters.

**5.5. Connecting blown-up cycles.** In the final section of the proof we will successively find 1-factors in each  $G_i$  and then turn each of these into a Hamilton cycle. As mentioned earlier, (Red0) will be used to ensure that each 1-factor  $f$  of  $G_i$  has a path connecting any pair of consecutive cycles of  $F_i$ , which will make it possible to merge the cycles of  $f$  into a Hamilton cycle. In this section we will modify  $G_i$  so that (Red0) holds.

We will join cycles by choosing *bridge vertices*  $x_{i,j}$  in  $V(G) \setminus (V_0 \cup V_{0,i}^{\text{spec}})$  whose neighbourhoods will be chosen from the sparse digraphs  $H_0^\pm$  defined in Section 5.2. In what follows, we write  $A_j^-$  for the predecessor of the  $p$ -cluster  $A_j$  in  $F_i$ .

**Claim 5.1.** *There is a sequence  $A_1 B_1 A_2 B_2 \dots A_{\tilde{L}} B_{\tilde{L}} A_1$  of  $p$ -clusters in  $R_p$  such that, for each  $1 \leq j, j' \leq \tilde{L}$  the following hold:*

- (i) *Let  $\tilde{A}_j$  and  $\tilde{B}_j$  be the adapted primary clusters containing  $A_j$  and  $B_j$  respectively. Then there is an edge from  $\tilde{A}_j$  to  $\tilde{B}_j$  in  $\tilde{R}$ .*
- (ii)  *$A_j^-$  is out-red and  $B_j$  is in-red;*

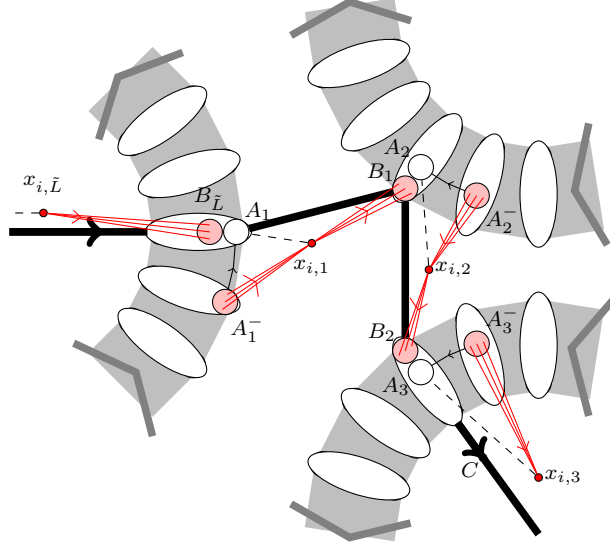


FIGURE 2. Bridge vertices  $x_{i,1}, x_{i,2}, x_{i,3}$  chosen from  $p$ -clusters  $A_1, A_2, A_3$  respectively.

- (iii)  $B_j$  and  $A_{j+1}$  lie in the same adapted primary cluster (where  $A_{\tilde{L}+1} := A_1$ ).
- (iv) Every adapted cluster contains exactly one  $A_j$  and exactly one  $B_{j'}$ .
- (v) All the  $A_j$  and  $B_{j'}$  are distinct.

To prove the claim, observe that, by Lemma 3.5,  $\tilde{R}$  is a robust  $(\nu/4, 3\tau)$ -outexpander, so Theorem 4.6 implies that  $\tilde{R}$  contains a Hamilton cycle  $C = \tilde{A}_1 \dots \tilde{A}_{\tilde{L}}$ . We will choose  $A_j$  in  $\tilde{A}_j$  and  $B_j$  in  $\tilde{A}_{j+1}$ . This will automatically satisfy (i), (iii) and (iv) and ensures that they will be distinct, except possibly  $A_j = B_{j-1}$ . Now recall that, by (21), each adapted primary cluster contains exactly  $s/2 - 1$  in-red and  $s/2 - 1$  out-red  $p$ -clusters. Moreover, as noted after  $(\star)$ , they all lie on the same cycle in  $F_i$  and  $p$ -clusters directly preceding those in  $\tilde{A}_j$  on  $F_i$  all lie in the same adapted primary cluster, which we call  $\tilde{A}_j^-$ . Thus we can always choose an in-red  $B_{j-1}$  in  $\tilde{A}_j$  and an out-red  $A_j^- \in \tilde{A}_j^-$  whose successor  $A_j$  on  $F_i$  lies in  $\tilde{A}_j$ , proving (ii). Moreover, we have  $s/2 - 1 > 1$  choices for  $A_j$  so we may assume that  $A_j$  and  $B_{j-1}$  are distinct. This proves (v) and thus the claim.

We will choose the bridge vertices in the sets  $A_j$ . The next claim guarantees many candidates for these bridge vertices whose neighbourhoods have the required properties.

**Claim 5.2.** *For each  $i$  with  $1 \leq i \leq r_p$ , whenever a  $p$ -cluster  $A_j \in V(F_i)$  is joined by an edge in  $R_p$  to a  $p$ -cluster  $B_j \in V(F_i)$  the following holds. Let  $(A_j^-)'$  and*

$B'_j$  be  $2p$ -clusters contained in  $A_j^-$  and  $B_j$  respectively. Then  $A_j$  contains at least  $m_p/2$  vertices  $x$  such that both  $|N_{H_0}^-(x) \cap (A_j^-)'| > \kappa$  and  $|N_{H_0}^+(x) \cap B'_j| > \kappa$ .

We say that a vertex  $x$  as in Claim 5.2 is  $(i, j)$ -useful. To prove the claim, note that (H0) and (URef) imply that, for at least  $3m_p/4$  of the vertices  $x \in A_j$ , we have

$$|N_{H_0}^+(x) \cap B'_j| \geq \gamma dm_p/5 \stackrel{(20)}{>} \kappa.$$

As  $F_i$  (and thus  $R_p$ ) contains the edge  $A_j^- A_j$ , for at least  $3m_p/4$  of the vertices  $x \in A_j$  we similarly have

$$|N_{H_0}^-(x) \cap (A_j^-)'| > \kappa.$$

So at least  $m_p/2$  the vertices in  $A_j$  satisfy both inequalities, which proves the claim.

Now we choose the set  $V_{0,i}^{\text{bridge}}$  satisfying (Red0). For each  $1 \leq i \leq r_p$ , consider the sequence guaranteed by Claim 5.1 and for each  $1 \leq j \leq \tilde{L}$ , let  $(A_j^-)^{\text{red}}$  and  $B_j^{\text{red}}$  be the unique red  $2p$ -clusters contained in  $A_j^-$  and  $B_j$  respectively. So  $(A_j^-)^{\text{red}}$  is out-red and  $B_j^{\text{red}}$  is in-red. For each  $1 \leq j \leq \tilde{L}$ , apply Claim 5.2 to the pair  $(A_j, B_j)$  with  $(A_j^-)^{\text{red}}, B_j^{\text{red}}$  playing the roles of  $(A_j^-)', B'_j$  respectively, to obtain a vertex  $x_{i,j} \in A_j$  which is  $(i, j)$ -useful and which is distinct from all vertices chosen so far. Note that the latter is possible since Claim 5.1(v) implies that for each  $i$ , we only choose one vertex from  $A_j$ . So altogether, we choose at most  $r_p$  vertices from each  $A_j$ , which is at most  $m_p/3$  by (18). In each  $G_i$ , remove each  $x_{i,j}$  from  $A_j$  and denote the collection of all  $x_{i,j}$  with  $1 \leq j \leq \tilde{L}$  by  $V_{0,i}^{\text{bridge}}$ . This process is illustrated in Figure 2.

Now, for each  $1 \leq i \leq r_p$ , there are exactly  $sp - 1$  of the  $p$ -clusters in each adapted primary  $(t)$ -cluster of  $G_i$  from which no vertices have been removed in this step, where  $t$  is the original type of  $F_i$ . We still need to ensure that  $p$ -clusters of  $G_i$  have equal size, so we choose a further  $(sp - 1)r_p \tilde{L} \leq \varepsilon n/3$  distinct vertices such that exactly one is removed from each untouched  $p$ -cluster in each  $G_i$ . Each such vertex is moved from its cluster into  $V_{0,i}^{\text{spec}}$ . The final inequality in (18) implies that we can assume that each vertex  $x$  is moved into  $V_{0,i}^{\text{spec}}$  for at most one  $1 \leq i \leq r_p$  in this step. Now adapted primary  $(t)$ -clusters become *adapted primary  $[i]$ -clusters* and (adapted)  $s$ -,  $p$ - and  $2p$ -( $t$ )-clusters become *(adapted)  $s$ -,  $p$ - and  $2p$ - $[i]$ -clusters* respectively (or  $s$ -,  $p$ -,  $2p$ -clusters if this is unambiguous). This will not overlap with previous notation as from now on we never refer to  $(t)$ -clusters and only ever refer to base and  $[i]$ -clusters. (19) implies that if  $U$  is a base  $2p$ -cluster and  $U_{[i]}$  is the associated  $2p$ - $[i]$ -cluster, then

$$(22) \quad |U \cap U_{[i]}| \geq (1 - \varepsilon')m_p/2 - 1.$$

We still refer to the cluster sizes  $m, m_s$  and  $m_p$  in the same way since each one has only lost at most  $sp$  vertices (which does not affect any calculations). The  $2p$ -clusters may no longer have exactly the same size, but this also does not affect

any of the calculations. We call  $V_{0,i}^{\text{bridge}}$  the set of *bridge vertices* and say that every edge incident to a bridge vertex is *i*-red. We now have that

$$(23) \quad |V_{0,i}| \stackrel{(11)}{\leq} \varepsilon n.$$

Since in  $G_i$ , we removed exactly one vertex from each  $p$ -cluster, we still have

$$|N_{H_0^+}^+(x_{i,j}) \cap B_j^{\text{red}}| \geq \kappa \quad \text{and} \quad |N_{H_0^-}^-(x_{i,j}) \cap (A_j^-)^{\text{red}}| \geq \kappa.$$

Since the  $x_{i,j}$  are all distinct, it follows that for each  $x_{i,j}$ , we can choose  $\kappa$  of these out-edges from  $H_0^+$  and add them to  $G_i$ . Similarly, we can choose  $\kappa$  of these in-edges from  $H_0^-$  and add them to  $G_i$ , whilst also removing every other edge incident to  $x_{i,j}$  in  $G_i$ . So (Red1) is satisfied for  $V_{0,i}^{\text{bridge}}$ .

It is now easy to verify (Red0). For each  $1 \leq i \leq r_p$ , consider the sequence given by Claim 5.1. Let  $D_j$  be the cycle of  $F_i$  containing the adapted  $p$ -cluster  $A_j$  for each  $1 \leq j \leq \tilde{L}$  and let  $x_j := x_{i,j}$  be the bridge vertex which was removed from  $A_j$  in  $G_i$ . Note that each cycle of  $F_i$  appears several times in the sequence. We claim that  $D_1 x_1 D_2 x_2 \dots x_{\tilde{L}} D_{\tilde{L}} x_{\tilde{L}} D_1$  is a sequence satisfying (Red0). The first property is immediate from Claim 5.1(iv). Each  $x_j$  has inneighbourhood contained in  $A_j^-$  which is in  $D_j$  since  $A_j$  is, and its outneighbourhood is contained in  $B_j$  which lies in the same adapted cluster as  $A_{j+1}$  and thus in the cycle  $D_{j+1}$ . Therefore the second property is also satisfied.

(Red2) follows since the in- and outedges incident to bridge vertices were chosen from edge-disjoint subdigraphs  $H_0^-$  and  $H_0^+$  respectively. Furthermore, by Claim 5.1(v), any  $y \notin V_{0,i}$  is incident to at most one *i*-red edge so (Red3) holds. In each red  $p$ -cluster  $V$ , red edges were only added to the unique red  $2p$ -cluster  $W$  contained in  $V$  so (Red4) is satisfied. (Red5) is satisfied by the comments after the statement of (Red7). Moreover every out-red  $p$ -cluster only sends out red edges and every in-red  $p$ -cluster only receives red edges so (Red6) holds. The edge-disjointness in (Red7) is immediate from the construction. Finally, note that any vertex in  $V(G) \setminus V_{0,i}$  lost at most one inneighbour and one outneighbour in  $G_i$ , so for each edge  $E$  of  $F_i$ ,  $G_i(E)$  is certainly still  $(2\varepsilon', \beta_1)$ -superregular. Therefore (Red0) and (Red2)–(Red7) are all satisfied. Note that (Red1) holds for all vertices in  $V_{0,i}^{\text{bridge}}$ . The aim of the next two sections is to maintain these properties while also achieving (Red1) for all vertices in  $V_{0,i}$ .

**5.6. Incorporating the core exceptional set  $V_0$ .** Note that so far,  $G_i$  contains no edges with initial or final vertex in  $V_0 \cup V_{0,i}^{\text{spec}}$ . In this section and the next we will add edges incident to these vertices into the  $G_i$ . Recall that we call such edges and any incident vertices *i*-red or *red* if this is unambiguous. Throughout both sections we will refer to (a) and (b) in Section 5.4. To achieve (Red1), we consider the core exceptional set  $V_0$  and the special exceptional set  $V_{0,i}^{\text{spec}}$  separately. In this section we consider the core exceptional set. Roughly speaking, the set of edges between  $V_0$  and  $G_i \setminus V_0$  will consist of a random subdigraph of  $G$  induced by  $V_0$  and the red  $2p$ -clusters of  $G_i$ . The following claim guarantees the existence of suitable edge-disjoint random subdigraphs. Recall from Section 5.4 that  $F = r_p/4p$ .

**Claim 5.3.** *Let  $X$  be a base  $2p$ -cluster which is a subcluster of a base primary cluster  $W$ . Then in  $G$  we can find  $F$  edge-disjoint bipartite graphs  $E_1^+(X), \dots, E_F^+(X)$  with all edges oriented from  $V_0$  to  $X$  so that for all  $1 \leq f \leq F$  the following hold:*

- (i) *For all  $x \in V_0$  we have  $d_{E_f^+(X)}^+(x) \geq \frac{1-\varepsilon}{2spF} (|N_G^+(x) \cap W| - 5\tilde{\varepsilon}\tilde{m})$ .*
- (ii) *For all  $y \in X$  we have  $d_{E_f^+(X)}^-(y) < \sqrt{\xi}\beta_1 m_p/2$ .*

*We can also find  $E_1^-(X), \dots, E_F^-(X)$  satisfying analogous properties for the in-neighbourhoods.*

To prove the claim, let  $E^+(X)$  denote the digraph induced by the set of edges from  $V_0$  to  $X$  in  $G$ . Now consider a random partition of the edges of  $E^+(X)$  into  $F$  parts  $E_f^+(X)$ . More precisely, assign each edge of  $E^+(X)$  to  $E_f^+(X)$  with probability  $1/F$ , independently of all other edges. There are several cases to consider. Say that  $x \in V_0$  is *prolific* if  $|N_G^+(x) \cap W| > 5\tilde{\varepsilon}\tilde{m}$ . Say that  $V_0$  is *large* if  $|V_0| \geq \sqrt{\xi}\beta_1 m_p/2$  and *small* otherwise. Every  $x \in V_0$  which is not prolific satisfies the condition in (i) with probability 1, and the inequality in (ii) is satisfied with probability 1 if  $V_0$  is small. Suppose that  $x$  is prolific. Then since  $\mathcal{P}'_{2p}$  is a  $5\tilde{\varepsilon}$ -uniform  $2sp$ -refinement of  $\tilde{\mathcal{P}}$ , (URef) implies that  $d_{E^+(X)}^+(x) \geq \frac{1-5\tilde{\varepsilon}}{2sp} |N_G^+(x) \cap W|$ .

Then for each  $1 \leq f \leq F$ , each prolific  $x \in V_0$  and each  $y \in X$ ,

$$\mathbb{E} \left( d_{E_f^+(X)}^+(x) \right) \geq \frac{1-5\tilde{\varepsilon}}{2spF} |N_G^+(x) \cap W| \quad \text{and} \quad \mathbb{E} \left( d_{E_f^+(X)}^-(y) \right) \leq \frac{|V_0|}{F}.$$

By Proposition 3.1 (with  $a := \varepsilon/2$ ) we have that, for fixed  $f$  and prolific  $x \in V_0$ ,

$$\begin{aligned} \mathbb{P} \left( d_{E_f^+(X)}^+(x) \leq \frac{1-\varepsilon}{2spF} |N_G^+(x) \cap W| \right) &\leq \exp \left( -\frac{5\varepsilon^2 \tilde{\varepsilon} \tilde{m} (1-5\tilde{\varepsilon})}{24spF} \right) \\ &\stackrel{(18)}{\leq} \exp \left( -\frac{\tilde{\varepsilon}^2 \beta n}{s^2 p \tilde{L}^2} \right) \stackrel{(4)}{\leq} e^{-\sqrt{n}} \end{aligned}$$

and  $|V_0|F \leq n^2$ . So taking a union bound over all  $f$  and all  $x \in V_0$  we see that the probability that (i) fails for some  $f$  in this partition is at most  $n^2 e^{-\sqrt{n}}$ . Similarly, for large  $V_0$ , fixed  $f$  and  $y \in X$ , Proposition 3.1 implies that

$$\begin{aligned} (24) \quad \mathbb{P} \left( d_{E_f^+(X)}^-(y) > \frac{2|V_0|}{F} \right) &\leq \exp \left( -\frac{\sqrt{\xi}\beta_1 m_p}{6F} \right) \\ &= \exp \left( -\frac{2\sqrt{\xi}\beta_1 m}{3sr_p} \right) \stackrel{(4),(18)}{\leq} e^{-\sqrt{n}}. \end{aligned}$$

Note that  $n \leq 2L_p m_p$  by (14) and

$$r_p \stackrel{(17)}{\geq} sp\tilde{r}/2 \stackrel{(7)}{\geq} sp\tilde{\alpha}\tilde{L}/3 \stackrel{(12)}{=} \tilde{\alpha}L_p/3.$$

Thus

$$(25) \quad \frac{2|V_0|}{F} \stackrel{(5)}{\leq} 4\sqrt{\varepsilon}n \frac{p}{r_p} \leq \frac{24\sqrt{\varepsilon}pm_p}{\tilde{\alpha}} \stackrel{(4)}{<} \frac{\sqrt{\xi}\beta_1 m_p}{2}.$$

Furthermore,  $|X|F \leq n^2$ , so (24) and (25) imply that the probability that (ii) fails for this partition is at most  $n^2 e^{-\sqrt{n}}$ . Therefore the partition satisfies both (i) and (ii) with probability  $1 - 2n^2 e^{-\sqrt{n}} \geq 1/2$ . This proves the claim.

For each  $k$  with  $1 \leq k \leq 2p$ , recall from the end of Section 5.4 that the number of all graphs  $G_i$  whose original type  $t$  satisfies (a) is  $2F$ . For each  $k$ , consider the ordering of all these digraphs as chosen in Section 5.4 and suppose that  $G_i$  is the  $f$ th digraph with original type  $t$ . We now define the edges of  $G_i$  between  $V_0$  and  $V(G) \setminus V_{0,i}$ . For each base  $s$ -cluster  $W_\ell$ , let  $W_\ell(k)$  denote the  $k$ th base  $2p$ -cluster contained in  $W_\ell$ . Apply Claim 5.3 to obtain  $F$  bipartite digraphs  $E_f^+(W_\ell(k))$  and  $F$  bipartite digraphs  $E_f^-(W_\ell(k))$  for each  $W_\ell(k)$ . Now let  $W'_\ell$  denote the  $s$ -[ $i$ ]-cluster associated with  $W_\ell$  and  $W_\ell(k)'$  denote the  $2p$ -[ $i$ ]-cluster associated with  $W_\ell(k)$ . Let  $E_f^+(W_\ell(k)')$  be the subdigraph of  $E_f^+(W_\ell(k))$  consisting of all edges whose final vertex lies in  $W_\ell(k)'$  and let  $E_f^-(W_\ell(k)')$  be the subdigraph of  $E_f^-(W_\ell(k))$  consisting of all edges whose initial vertex lies in  $W_\ell(k)'$ . Then, by (22), for all  $x \in V(G)$  we have

$$(26) \quad d_{E_f^+(W_\ell(k)')}^+(x) \geq d_{E_f^+(W_\ell(k))}^+(x) - \varepsilon' m_p / 2 - 1.$$

An analogous statement is true for the indegrees in  $E_f^-$ . Recall that  $k$  with  $1 \leq k \leq 2p$  is defined by the fact that  $G_i$  has original type  $t$  and  $t \equiv k \pmod{2p}$ , and that  $I_W^+$  and  $I_W^-$  are the indices of the non-clean  $s$ -clusters in  $W$  defined at the end of Section 5.4. If  $1 \leq f \leq F$ , we add the following edges to  $G_i$ :

- all edges lying in the digraphs  $E_f^+(W_\ell(k)')$  with  $\ell \in I_W^+$ ;
- all edges lying in the digraphs  $E_f^-(W_\ell(k)')$  with  $\ell \in I_W^-$ .

If  $F < f \leq 2F$ , we add the following edges to  $G_i$ :

- all edges lying in the digraphs  $E_{f-F}^+(W_\ell(k)')$  with  $\ell \in I_W^-$ ;
- all edges lying in the digraphs  $E_{f-F}^-(W_\ell(k)')$  with  $\ell \in I_W^+$ .

Note this implies that all edges from  $G_i \setminus V_0$  to  $V_0$  have initial vertex in an out-red cluster and similarly for the in-red clusters. Moreover, the sets of edges assigned to  $G_i$  and  $G_{i'}$  are disjoint for  $i \neq i'$ . Indeed, this follows from the fact that, for  $j \neq k$ ,  $E_f^\pm(W_\ell(j))$  and  $E_{f'}^\pm(W_\ell(k))$  are clearly edge-disjoint; that for  $f \neq f'$ ,  $E_f^\pm(W_\ell(k)')$  and  $E_{f'}^\pm(W_\ell(k)')$  are also edge-disjoint; and that each  $E_f^\pm(W_\ell(k))$  is used for at most one of the  $G_i$ .

Therefore Claim 5.3, (26) and (21) imply that for all  $x \in V_0$ , we have that

$$d_{G_i}^+(x) \geq (s/2 - 1) \sum_{W \in \mathcal{P}} \left( \frac{1 - \varepsilon}{2spF} (|N_G^+(x) \cap W| - 5\varepsilon\tilde{m}) - \frac{\varepsilon' m_p}{2} - 1 \right).$$

Note also that

$$(27) \quad 2spF = \frac{sr_p}{2} \stackrel{(18)}{\leq} \frac{s}{2} \frac{\tilde{\alpha} L_p}{\beta} \stackrel{(14)}{\leq} \frac{s}{2} \frac{\tilde{\alpha} n}{\beta m_p}.$$

So

$$\begin{aligned}
 d_{G_i}^+(x) &\geq (s/2 - 1) \frac{(1 - \varepsilon')}{2spF} (\tilde{\alpha}n - |V_0| - 2\varepsilon'n) \\
 (28) \quad &\stackrel{(5),(27)}{\geq} (1 - 4\varepsilon')\beta m_p \stackrel{(8)}{\geq} \beta_1 m_p \stackrel{(20)}{\geq} \kappa,
 \end{aligned}$$

and we have an analogue for indegrees. So we can delete edges from each  $x \in V_0$  so that  $d_{G_i}^\pm(x) = \kappa$  in each slice and hence (Red1) holds for all vertices in  $V_0$ .

**5.7. Incorporating the special exceptional set  $V_{0,i}^{\text{spec}}$ .** We now prove a claim which will be used to achieve (Red1) for the set  $V_{0,i}^{\text{spec}}$  of special exceptional vertices. Before this, we first need to derive a further property (H1') of  $H_1^\pm$  from (H1).

Write  $S_i^+$  for the collection of vertices contained in the out-red  $2p$ -[ $i$ ]-clusters and define  $S_i^-$  analogously. Note that each of  $S_i^\pm$  consists of the vertices in exactly  $s/2 - 1$  of the  $2p$ -[ $i$ ]-clusters in each adapted  $s$ -[ $i$ ]-cluster.

For every  $1 \leq k \leq 2p$  and every base  $s$ -cluster  $U \in \mathcal{P}'_s$ , let  $U(k)$  be the  $k$ th base  $2p$ -cluster of  $U$ , and write  $H_{1,k}^+$  for the spanning subdigraph of  $H_1^+$  consisting of all edges whose final vertex lies in  $\bigcup_{U \in \mathcal{P}'_s} U(k)$ . Also define  $H_{1,k}^-$  to be the spanning subdigraph of  $H_1^-$  consisting of all edges whose initial vertex lies in  $\bigcup_{U \in \mathcal{P}'_s} U(k)$ . We have the following property of  $H_1^\pm$ :

(H1') For all  $x \in V(G) \setminus V_0$ , whenever  $i$  has original type  $t$  and  $k$  satisfies  $1 \leq k \leq 2p$  and (a) we have that

$$\frac{\gamma \tilde{\alpha} n}{20p} \leq |N_{H_{1,k}^+}^+(x) \cap S_i^-|, \quad |N_{H_{1,k}^-}^-(x) \cap S_i^+| \leq \frac{\gamma \tilde{\alpha} n}{p}.$$

To prove (H1'), note that since  $\mathcal{P}'_{2p}$  was a  $5\tilde{\varepsilon}$ -uniform  $2sp$ -refinement of  $\tilde{\mathcal{P}}$ , (URef) implies that, for each  $x \in V(G) \setminus V_0$ , each  $1 \leq k \leq 2p$  and  $U \in \mathcal{P}'_s$ ,

$$|N_{H_{1,k}^+}^+(x) \cap U(k)| \geq \frac{(1 - 5\tilde{\varepsilon})}{2sp} (|N_{H_1^+}^+(x) \cap \tilde{U}| - 5\tilde{\varepsilon}\tilde{m})$$

where  $\tilde{U}$  is the base primary cluster containing  $U$ . If  $U(k)_i$  is the  $2p$ -[ $i$ ]-cluster associated with  $U(k)$ , (22) implies that

$$|N_{H_{1,k}^+}^+(x) \cap U(k)_i| \geq |N_{H_{1,k}^+}^+(x) \cap U(k)| - \varepsilon' m_p / 2 - 1.$$

But whenever  $i$  has original type  $t$  and  $k$  satisfies (a),  $S_i^-$  contains all the vertices from exactly  $s/2 - 1$  of the  $2p$ -[ $i$ ]-clusters  $U(k)_i$  contained in each adapted [ $i$ ]-cluster  $\tilde{U}_i$  associated with  $\tilde{U}$ , so

$$\begin{aligned}
 |N_{H_{1,k}^+}^+(x) \cap S_i^- \cap \tilde{U}_i| &\geq (s/2 - 1) \left( \frac{1 - 5\tilde{\varepsilon}}{2sp} (|N_{H_1^+}^+(x) \cap \tilde{U}| - 5\tilde{\varepsilon}\tilde{m}) - \frac{\varepsilon' m_p}{2} - 1 \right) \\
 &\geq |N_{H_1^+}^+(x) \cap \tilde{U}| / 6p - \varepsilon' s m_p.
 \end{aligned}$$

Therefore, summing over all base primary clusters  $\tilde{U}$  and recalling that  $V_0$  is an isolated set in  $H_1^+$  we have that

$$|N_{H_{1,k}}^+(x) \cap S_i^-| \geq \frac{d_{H_1^+}^+(x)}{6p} - \varepsilon' sm_p \tilde{L} \stackrel{(H1)}{\geq} \frac{\gamma \tilde{\alpha} n}{20p}.$$

The other bounds in (H1') follow similarly.

**Claim 5.4.** *For each  $i$  with  $1 \leq i \leq r_p$ , there are subdigraphs  $Q_i^+$  of  $H_1^+$  and  $Q_i^-$  of  $H_1^-$  each consisting of edges between  $V_{0,i}^{\text{spec}}$  and  $V(G) \setminus V_{0,i}$  so that*

- (i) *for all  $x \in V_{0,i}^{\text{spec}}$  we have  $|N_{Q_i^+}^+(x) \cap S_i^-|, |N_{Q_i^-}^-(x) \cap S_i^+| \geq \kappa$ .*
- (ii) *For all  $y \in V(G) \setminus V_{0,i}$  we have  $d_{Q_i^-}^+(y), d_{Q_i^+}^-(y) \leq \sqrt{\xi} \beta_1 m_p / 3$ .*
- (iii) *all the  $Q_i^\pm$  are pairwise edge-disjoint.*

To prove the claim, for each vertex  $x$  in  $V(G) \setminus V_0$ , we let  $T(x) := \{i : x \in V_{0,i}^{\text{spec}}\}$ . Recall that  $x \in V_{0,i}^{\text{spec}}$  if and only if

- (A)  $x \in \tilde{V}_{0,t}^{\text{spec}}$  and  $i$  has original type  $t$ ; or
- (B)  $x$  was removed to compensate for the removal of a bridge vertex.

Note that  $x$  can satisfy both (A) and (B). Suppose that  $x$  satisfies (A). Let  $\mathcal{L}_x = \{t : x \in \tilde{V}_{0,t}^{\text{spec}}\}$ . Note  $x \notin V_0$ . So  $x$  is good in the sense of Section 5.1, and hence  $|\mathcal{L}_x| \leq \xi \tilde{L} / \beta$ . As observed before (23), any  $x \in V(G) \setminus V_0$  is in at most one set  $V_{0,i}^{\text{spec}}$  due to (B). Therefore

$$|T(x)| \leq |\mathcal{L}_x|(s-1)(p-1) + 1 \leq \xi \tilde{L} sp / \beta \stackrel{(12)}{=} \xi L_p / \beta.$$

For each  $1 \leq i \leq r_p$  and each  $1 \leq k \leq 2p$  we define digraphs  $Q_{i,k}^+$  as follows. For each  $k$ , we randomly assign each edge of  $H_{1,k}^+$  whose initial vertex is  $x$  to one of the digraphs  $Q_{i,k}^+$  with  $i \in T(x)$  with probability  $q := \beta / \xi L_p$  (independently of all other edges, and each edge is assigned to at most one of the  $Q_{i,k}^+$ ). The sum of the probabilities is at most 1. Note that  $V_0$  is an isolated set in  $H_{1,k}^\pm$ . Now define  $Q_i^+ := Q_{i,k}^+$  where  $i$  has original type  $t$  and  $k$  satisfies (a). Then (iii) certainly holds, and for all  $x \in V_{0,i}^{\text{spec}}$ , we have

$$(29) \quad \mathbb{E} \left( |N_{Q_i^+}^+(x) \cap S_i^-| \right) = \frac{\beta |N_{H_{1,k}^+}^+(x) \cap S_i^-|}{\xi L_p} \stackrel{(H1')}{\geq} \frac{\gamma \tilde{\alpha} \beta n}{20p \xi L_p} \stackrel{(14)}{\geq} 2\beta m_p.$$

Proposition 3.1 implies that, for fixed  $1 \leq i \leq r_p$  and fixed  $x \in V_{0,i}^{\text{spec}}$ ,

$$\mathbb{P} \left( |N_{Q_i^+}^+(x) \cap S_i^-| < \beta_1 m_p \right) \leq \exp \left( -\frac{\beta m_p}{6} \right) \stackrel{(12),(14)}{\leq} \exp \left( -\frac{\beta n}{12sp \tilde{L}} \right) \leq e^{-\sqrt{n}}.$$

So a union bound implies that the probability that there exist  $i$  and  $x$  not satisfying this inequality is at most  $n^2 e^{-\sqrt{n}} < 1/4$ . (i) now follows since  $\kappa \leq \beta_1 m_p$  by (20).

For (ii), note that for *any* vertex  $y \in V(G)$  we have

$$(30) \quad \mathbb{E} \left( d_{Q_i^+}^-(y) \right) \leq q |V_{0,i}^{\text{spec}}| \stackrel{(23)}{\leq} \frac{\beta}{\xi L_p} \varepsilon n \stackrel{(14)}{\leq} \frac{2\varepsilon}{\xi} \beta m_p \leq \sqrt{\xi} \beta m_p / 4.$$

Proposition 3.1 shows (as in Claim 5.3) that the probability that the condition in (ii) fails for some  $i$  and some  $y \in V(G)$  is at most  $1/4$ . So there is a choice of  $Q_1^+, \dots, Q_{r_p}^+$  so that all the conditions hold, and similarly for  $Q_1^-, \dots, Q_{r_p}^-$ , which proves Claim 5.4.

It is now easy to obtain the edges of  $G_i$  between  $V_{0,i}^{\text{spec}}$  and  $V(G) \setminus V_{0,i}$ . Apply Claim 5.4 to find edge-disjoint digraphs  $Q_i^\pm$  for each  $1 \leq i \leq r_p$ . Recall that  $S_i^\pm \subseteq V(G) \setminus V_{0,i}$  and so (Red6) will follow if we add  $i$ -red edges with initial vertex in  $S_i^+$  or final vertex in  $S_i^-$ . So for each  $x \in V_{0,i}$  we add exactly  $\kappa$  edges in  $Q_i^+$  going from  $x$  to  $S_i^-$  and exactly  $\kappa$  edges in  $Q_i^-$  going to  $x$  from  $S_i^+$ .

We have now incorporated  $V_{0,i}$  into each  $G_i$ . It remains to verify that (Red0)–(Red7) hold. Recall that we partially verified these properties for the red vertices incident to bridge vertices at the end of Section 5.5. In particular, (Red0) was achieved in Section 5.5 and the edges we have added here do not affect it. The previous paragraph shows that (Red1) holds for all vertices in  $V_{0,i}^{\text{spec}}$ . Since we already verified it for the bridge vertices  $V_{0,i}^{\text{bridge}}$  in Section 5.5 and for  $V_0$  in Section 5.6, it now holds for all vertices in  $V_{0,i}$ . Clearly, our construction satisfies (Red2). (Red3) follows from Claims 5.3(ii) and 5.4(ii) and the fact that each non-exceptional vertex is incident to at most one bridge vertex in each slice. Recall that, in Section 5.4, we showed how (Red4) and (Red5) follow from (a) and (b) of the construction. (Red6) follows from the fact that in constructions including  $V_0$  and  $V_{0,i}^{\text{spec}}$  and  $V_{0,i}^{\text{bridge}}$ , the outedges from  $V_{0,i}$  always went to in-red clusters and the inedges to  $V_{0,i}$  came from out-red clusters. (Red7) follows immediately from the edge-disjointness of the digraphs in Claims 5.3 and 5.4 and the observation in the final paragraph of Section 5.5.

Note that Theorem 3.2 implies that the proofs of Claims 5.3 and 5.4 can be ‘de-randomised’ and so red edges satisfying (Red0)–(Red7) can be found in polynomial time.

**5.8. Finding shadow balancing sequences.** We have now incorporated all the exceptional vertices to form  $r_p$  edge-disjoint slices  $G_i$  of  $G$ , together containing almost all edges, such that each slice is a spanning almost-regular subdigraph of  $G$ . The main aim of this section is to add further red edges to each slice  $G_i$  so that the number of red edges sent out by vertices in each cluster  $V$  equals the number received by its successor  $V^+$  on the cycle of  $F_i$  containing  $V$ .

This ‘balancing property’ is necessary for the following reason. Suppose that  $V$  is out-red and suppose that we have a 1-factor  $f$  containing a red edge sent out to  $V_{0,i}$  by a vertex  $x \in V$ . If  $V^+$  is not red, any edge of  $f$  to  $V^+$  must have its initial vertex in  $V$ . So  $f[V, V^+]$  must be a perfect matching, which is impossible since there can be no edge in  $f$  from  $x$  to a vertex in  $V^+$ . Note that the absence of red edges incident to  $V^-$  does not give rise to the above problem. But we observe a

similar problem for  $U, U^-$  when  $U$  is in-red. So the above ‘balancing property’ is certainly necessary to obtain even a single 1-factor. We will see in Section 5.10 that, combined with our other properties, it is also sufficient.

We will add ‘balancing edges’ between non-exceptional vertices to achieve the above property while also ensuring that no vertex is incident to many red edges. As indicated above, it will turn out to be sufficient to only add such edges to either the predecessor or successor of existing red clusters. By the end of Section 5.9 our new red clusters will consist of consecutive pairs, well-spaced around each blown-up cycle.

We will first find ‘shadow balancing edges’ in the reduced digraph between suitable cluster pairs. For this, we will use the fact that  $R_p$  is a robust outexpander. Then we will choose the required number of edges from the sparse pre-reserved subdigraph  $H_2$  induced on these pairs. When doing this, we need to be careful to maintain (Red6) with  $p$  replaced by  $p - 1$ .

Given  $G_i$ , we denote the set of red  $p$ -clusters by  $T$  (so we suppress the dependence on  $i$  here). Let  $T_{\text{in}}$  denote the set of in-red clusters and define  $T_{\text{out}}$  similarly, so  $T = T_{\text{in}} \cup T_{\text{out}}$ . For a set  $S \subseteq T$  of  $p$ -clusters, we let  $S^-$  denote the predecessors of  $S$  on  $T$  and define  $S^+$  similarly.

Now, for each  $1 \leq i \leq r_p$  and each  $p$ -cluster  $V$ , let

$$(31) \quad s_i^\pm(V) := \sum_{y \in V} |N_{G_i}^\pm(y) \cap V_{0,i}|$$

be the number of red edges entering/leaving  $V$ . So  $s_i^+(V) \neq 0$  only if  $V \in T_{\text{out}}$  and  $s_i^-(V) \neq 0$  only if  $V \in T_{\text{in}}$ . Note that (Red1) implies that

$$(32) \quad \sum_{V \in R_p} s_i^+(V) = \sum_{V \in R_p} s_i^-(V).$$

Let

$$(33) \quad b := \frac{\xi^{1/6} \beta_1 m_p^2}{L_p} \quad \text{and} \quad c := \xi^{1/5} \beta_1 m_p^2.$$

A *balancing sequence*  $B_i$  with respect to  $G_i$  is a spanning subdigraph of  $H_2$  with the following properties:

- (B1)  $d_{B_i}^\pm(y) \leq 8\xi^{1/6} \beta_1 m_p$  for every  $y \notin V_{0,i}$ ;
- (B2) We have the following degree conditions:

$$d_{B_i}^+(V) = \begin{cases} s_i^-(V^+) + c & \text{if } V \in T_{\text{in}}^- \\ c & \text{if } V \in T_{\text{out}} \\ 0 & \text{otherwise} \end{cases}$$

$$d_{B_i}^-(V) = \begin{cases} c & \text{if } V \in T_{\text{in}} \\ s_i^+(V^-) + c & \text{if } V \in T_{\text{out}}^+ \\ 0 & \text{otherwise} \end{cases}$$

We will use so called ‘shadow balancing sequences’ as a framework to find balancing sequences. For this, define an auxiliary digraph  $R^*$  with  $V(R^*) = T$  as follows. Let

$$(34) \quad N_{R^*}^+(V) = \begin{cases} \left( N_{R_p}^+(V^-) \cap T_{\text{in}} \right) \cup \left( N_{R_p}^+(V^-) \cap T_{\text{out}}^+ \right)^- & \text{if } V \in T_{\text{in}} \\ \left( N_{R_p}^+(V) \cap T_{\text{in}} \right) \cup \left( N_{R_p}^+(V) \cap T_{\text{out}}^+ \right)^- & \text{if } V \in T_{\text{out}} \end{cases}$$

This definition reflects the fact that red edges entering  $V \in T_{\text{in}}$  will be balanced by edges leaving  $V^-$  (and entering either  $T_{\text{in}}$  or the successor  $W^+$  of some  $W \in T_{\text{out}}$ ). Similarly an edge leaving  $V \in T_{\text{out}}$  will be balanced by an edge entering  $V^+$ . Note that  $R^*$  depends on  $i$ . If we need to emphasise this, we write  $R_i^*$ .

Define a *shadow balancing sequence*  $B'_i$  to be a multidigraph with vertex set  $V(R^*)$  whose edges are copies of edges of  $R^*$  as follows. Let

$$n_V^+ := \begin{cases} s_i^-(V) + c & \text{if } V \in T_{\text{in}} \\ c & \text{if } V \in T_{\text{out}} \end{cases} \quad \text{and} \quad n_V^- := \begin{cases} c & \text{if } V \in T_{\text{in}} \\ s_i^+(V) + c & \text{if } V \in T_{\text{out}} \end{cases}$$

Then  $B'_i$  has the following properties:

- (B1') no edge of  $R^*$  appears more than  $b$  times in  $B'_i$ .
- (B2') For every  $V \in V(R^*)$ , we have  $d_{B'_i}^+(V) = n_V^+$  and  $d_{B'_i}^-(V) = n_V^-$ .

Note that (32) implies that

$$(35) \quad \sum_{V \in R^*} n_V^+ = \sum_{V \in R^*} n_V^-.$$

To find these shadow balancing sequences, we will need that  $R^*$  is a robust out-expander with sufficiently large minimum semidegree.

**Claim 5.5.** *Let  $\nu' = \nu^3/64$ . Then*

- (i)  $R^*$  is a robust  $(\nu', 12\tau)$ -outexpander.
- (ii)  $\delta^0(R^*) \geq \tilde{\alpha}|R^*|/4$ .

To prove part (i) of the claim, we will use the fact that an  $(s/2-1)$ -fold blow-up of a robust  $(\nu/4, 3\tau)$ -outexpander is a  $(\nu', 6\tau)$ -robust outexpander (see Lemma 3.10). Let  $R_p^{\text{in}} = R_p[T_{\text{in}}]$  and  $R_p^{\text{out}} = R_p[T_{\text{out}}^+]$ . Since every adapted primary cluster contains exactly  $s/2 - 1$  out-red  $p$ -clusters, it follows that  $R_p^{\text{in}}$  is an  $(s/2 - 1)$ -fold blow-up of  $\tilde{R}$ . So it is a robust  $(\nu', 6\tau)$ -outexpander. Similarly,  $R_p^{\text{out}}$  is a robust  $(\nu', 6\tau)$ -outexpander.

Consider any  $S \subseteq T_{\text{in}}$  with  $6\tau|T_{\text{in}}| \leq |S| \leq (1 - 6\tau)|T_{\text{in}}|$ . Note that  $T_{\text{in}}$  and  $T_{\text{out}}$  are disjoint (see e.g. (Red6)). So  $T_{\text{in}}$  and  $(T_{\text{out}}^+)^-$  are disjoint and hence (34) implies that

$$(36) \quad |RN_{\nu', R^*}^+(S)| = |RN_{\nu', R_p}^+(S^-) \cap T_{\text{in}}| + |(RN_{\nu', R_p}^+(S^-) \cap T_{\text{out}}^+)^-|.$$

Now let  $S_{\text{in}}^-$  be obtained from  $S^-$  by replacing each  $p$ -cluster  $V \in S^-$  by an arbitrary (but distinct)  $p$ -cluster  $V_{\text{in}} \in T_{\text{in}}$  which lies in the same adapted primary cluster as  $S^-$ . Note this is possible as  $S \subseteq T_{\text{in}}$  implies that  $S$  (and thus  $S^-$ ) contains at most  $s/2 - 1$  of the  $p$ -clusters from each adapted  $s$ -cluster. Note that

in  $R_p$ , each cluster receives an edge from  $V_{\text{in}}$  if and only if it receives an edge from  $V$ . So (34) implies that

$$\begin{aligned} |RN_{\nu', R_p}^+(S^-) \cap T_{\text{in}}| &= |RN_{\nu', R_p}^+(S_{\text{in}}^-) \cap T_{\text{in}}| = |RN_{\nu', R_p^{\text{in}}}^+(S_{\text{in}}^-)| \\ &\geq |S_{\text{in}}^-| + \nu' |R_p^{\text{in}}| = |S| + \nu' |R^*|/2. \end{aligned}$$

Similarly, let  $S_{\text{out}}^-$  be obtained from  $S^-$  by replacing each  $p$ -cluster  $V \in S^-$  by an arbitrary (but distinct) cluster  $V_{\text{out}} \in T_{\text{out}}^+$  which lies in the same adapted  $s$ -cluster as  $V$ . Then we have

$$\begin{aligned} |(RN_{\nu', R_p}^+(S^-) \cap T_{\text{out}}^+)^-| &= |RN_{\nu', R_p}^+(S^-) \cap T_{\text{out}}^+| \\ &= |RN_{\nu', R_p}^+(S_{\text{out}}^-) \cap T_{\text{out}}^+| = |RN_{\nu', R_p^{\text{out}}}^+(S_{\text{out}}^-)| \\ &\geq |S_{\text{out}}^-| + \nu' |R_p^{\text{out}}| = |S| + \nu' |R^*|/2. \end{aligned}$$

So altogether, we have  $|RN_{\nu', R^*}^+(S)| \geq 2|S| + \nu' |R^*|$ .

Now suppose that  $S \subseteq T_{\text{out}}$  with  $6\tau |T_{\text{out}}| \leq |S| \leq (1 - 6\tau) |T_{\text{out}}|$ . Similarly as above, (34) implies that

$$(37) \quad |RN_{\nu', R^*}^+(S)| = |RN_{\nu', R_p}^+(S) \cap T_{\text{in}}| + |(RN_{\nu', R_p}^+(S) \cap T_{\text{out}}^+)^-| \geq 2|S| + \nu' |R^*|.$$

Now consider any  $S \subseteq V(R^*)$  with  $6\tau |R^*| \leq |S| \leq (1 - 6\tau) |R^*|$ . Then either  $|S \cap T_{\text{in}}| \geq |S|/2$  or  $|S \cap T_{\text{out}}| \geq |S|/2$ . In either case, we get  $|RN_{\nu', R^*}^+(S)| \geq |S| + \nu' |R^*|$ . This proves part (i) of the claim.

To prove part (ii), suppose that  $V \in T_{\text{in}}$ . Note that  $R_p^{\text{in}}$  satisfies  $\delta^0(R_p^{\text{in}}) \geq \tilde{\alpha} |R_p^{\text{in}}|/2$  by Lemma 3.10(i). Choose any  $V_{\text{in}}^- \in T_{\text{in}}$  which lies in the same adapted primary cluster as  $V^-$ . Then, similarly as observed above,  $V_{\text{in}}^-$  has the same outneighbours within the set  $T_{\text{in}}$  as  $V^-$  (both in the digraph  $R_p$ ). So the degree bound follows for  $V$ . The case when  $V \in T_{\text{out}}$  is similar. This proves Claim 5.5.

It is now easy to find shadow balancing sequences  $B'_i$  satisfying (B1') and (B2'). Indeed, note that  $c \leq n_V^\pm \leq c + \sqrt{\xi} \beta_1 m_p^2$  by (Red3). In particular, (33) implies that  $n_V^+ = c(1 \pm \xi^{3/10})$  and similarly for  $n_V^-$ . Let  $R'$  be obtained from  $R^*$  by replacing each of the edges of  $R^*$  by  $b$  copies of this edge and let  $n' := |R^*| = (s - 2)\tilde{L}$ . We will apply Lemma 4.1 as follows:

	$R^*$	$R'$	$n'$	$b$	$\xi^{3/10}$	$\nu'$	$c/n'$
playing the role of	$G$	$Q$	$n$	$q$	$\varepsilon$	$\nu$	$\rho$

Then

$$\rho := \frac{c}{n'} \stackrel{(33)}{=} \frac{\xi^{1/5} \beta_1 m_p^2}{(s - 2)\tilde{L}} \stackrel{(4)}{\leq} \frac{\xi^{1/6} \beta_1 m_p^2 \nu'^2}{3sp\tilde{L}} \stackrel{(12), (33)}{=} \frac{b\nu'^2}{3}$$

as required by Lemma 4.1, and we obtain a spanning subdigraph  $B'_i$  of  $R'$  with  $d_{B'_i}^\pm(V) = n_V^\pm$  for each  $V \in V(R') = V(R^*)$ .

**5.9. Adding balancing sequences.** Note that for each edge  $E'$  of  $R_i^*$ , there is a unique edge  $E$  of  $R_p$  (from a  $p$ -cluster  $A$  to a  $p$ -cluster  $B$ ) which corresponds to  $E'$ . More precisely, (34) shows that if  $E' = VW \in E(R_i^*)$  then

$$(38) \quad E = \begin{cases} V^-W & \text{if } V \in T_{\text{in}}, W \in T_{\text{in}} \\ V^-W^+ & \text{if } V \in T_{\text{in}}, W \in T_{\text{out}} \\ VW & \text{if } V \in T_{\text{out}}, W \in T_{\text{in}} \\ VW^+ & \text{if } V \in T_{\text{out}}, W \in T_{\text{out}}. \end{cases}$$

(As before,  $V^-$  denotes the predecessor of  $V$  on  $F_i$ .) So for each edge of  $B'_i$ , we can choose the corresponding edge of  $R_p$ . For each  $i$  and each edge  $E$  of  $R_p$ , let  $c_i(E)$  denote the number of times that the edge  $E$  is chosen due to  $B'_i$ . So  $c_i(E) \leq b$  by (B1'). If we now replace the chosen edges  $E$  of  $R_p$  with  $c_i(E)$  edges in  $H_2(E)$ , this will give the required balancing sequence  $B_i$ . However, we need to be careful to ensure that we can do this for every  $i$  with  $1 \leq i \leq r_p$  so that all edges are disjoint. We also wish to maintain (Red4) and (Red6).

We now need to consider the dependence on  $i$  again, as clusters in different slices are not quite the same. Given a base  $p$ -cluster  $A$  in  $R_p$ , let  $A[i]$  be the associated  $p$ -[ $i$ ]-cluster. Each  $p$ -[ $i$ ]-cluster  $A[i]$  contains at most one red  $2p$ -[ $i$ ]-cluster by (Red4). If there is such a subcluster, denote it by  $A^*[i]$ . If there is no such subcluster, let  $A^*[i]$  be an arbitrary subcluster of  $A[i]$ . We will only add balancing edges incident to  $A^*[i]$ . Let  $A^*$  be the base  $2p$ -cluster associated with  $A^*[i]$ . Suppose that  $E$  is an edge of  $R_p$  from  $A$  to  $B$ . Let  $\tilde{E} \in E(\tilde{R}(\beta))$  be one of the edges whose blow-up contains  $E$ ; then  $H_2(\tilde{E})$  is  $(\varepsilon, \gamma\beta)$ -regular as observed in Section 5.2. Write  $H_2(E^*)$  for the subdigraph of  $H_2(\tilde{E})$  induced on  $(A^*, B^*)$ ; then by Lemma 3.11(i) we have that  $H_2(E^*)$  is  $(\varepsilon', \gamma\beta)$ -regular.

Write  $H_2(E^*[i])$  for the subdigraph of  $H_2(E^*)$  induced on  $(A^*[i], B^*[i])$ . Whenever  $E$  is chosen due to  $B'_i$ , we will add balancing edges to  $G_i$  from  $H_2(E^*[i])$ . By (22) we have that, for all  $i$  with  $1 \leq i \leq r_p$ ,  $H_2(E^*[i])$  is a subdigraph of  $H_2(E^*)$  obtained by removing at most  $\varepsilon' m_p/2 + 1$  vertices from each vertex class.

**Claim 5.6.** *Let  $d_0 := 8b/m_p^2$  where  $b$  is defined in (33). Suppose that  $H$  is a subdigraph of  $H_2(E^*)$  obtained by removing at most  $\varepsilon' m_p/2 + 1$  vertices from each of  $A^*$  and  $B^*$  and at most  $r_p d_0 m_p$  edges at every vertex. Then  $H$  is  $(\xi^{1/15}, \gamma\beta)$ -regular.*

To prove the claim, note first that

$$(39) \quad d_0 = \frac{8b}{m_p^2} = \frac{8\xi^{1/6}\beta_1}{L_p}.$$

So

$$2r_p d_0 \stackrel{(18)}{\leq} \frac{16\xi^{1/6}\beta_1}{L_p} \frac{\tilde{\alpha}L_p}{\beta} \leq 16\xi^{1/6}\tilde{\alpha} \leq \xi^{1/7}.$$

Also  $\varepsilon' \ll \xi^{1/7}$ . So Proposition 3.6(i) with  $\xi^{1/7}$  playing the role of  $d'$  implies the claim.

Now for each  $i$  in succession we aim to apply Lemma 3.9 to find a set  $C_i(E)$  of  $c_i(E)$  edges in  $H_2(E^*)$ , and remove the edges of  $C_i(E)$  from further consideration. Suppose we have found  $C_1(E), \dots, C_{i-1}(E)$  in  $H_2(E^*)$ . Suppose further that each of these has maximum degree at most  $d_0 m_p$  and that the edges are from  $A^*$  to  $B^*$ . We now wish to find  $C_i(E)$ .

Let  $H_2^{i-1}(E^*)$  denote the subdigraph of  $H_2(E^*)$  obtained by removing the edges of  $C_1(E), \dots, C_{i-1}(E)$  and removing any vertex not present in  $H_2(E^*[i])$ . So  $H_2^{i-1}(E^*)$  is also a subdigraph of  $H_2(E^*[i])$ . By (22), the number of vertices in each vertex class of  $H_2^{i-1}(E^*)$  is at most  $\varepsilon' m_p / 2 + 1$  less than that in  $H_2(E^*)$ . Moreover, at most  $r_p d_0 m_p$  edges have been removed from each vertex. Then Claim 5.6 implies that  $H_2^{i-1}(E^*)$  is  $(\xi^{1/15}, \gamma\beta)$ -regular. So we can apply Lemma 3.9 to find  $C_i(E)$ , with a maximum degree of at most  $8c_i(E)/m_p \leq 8b/m_p = d_0 m_p$ . We continue inductively until we have found  $C_1(E), \dots, C_{r_p}(E)$ .

Now let  $B_i$  be the union of all  $C_i(E)$  over all edges  $E$  of  $R_p$ . Note that the  $B_i$  are edge-disjoint by construction. To verify (B1), note that for all  $y \in V(G) \setminus V_{0,i}$ ,

$$d_{B_i}^\pm(y) \leq L_p d_0 m_p \stackrel{(39)}{=} 8\xi^{1/6} \beta_1 m_p,$$

as required. (38) implies that the clusters that send out shadow balancing edges are precisely  $T_{\text{in}}^- \cup T_{\text{out}}$  and the clusters that receive shadow balancing edges are precisely  $T_{\text{in}} \cup T_{\text{out}}^+$ . Suppose that  $V \in T_{\text{in}}^-$ . Then we have that

$$d_{B_i}^+(V) \stackrel{(38)}{=} d_{B_i}^+(V^+) \stackrel{(B2')}{=} n_{V^+}^+ = s_i^-(V^+) + c$$

so (B2) holds in this case. The other cases follow similarly. Therefore  $B_i$  satisfies (B1) and (B2). Note that only vertices in a single  $2p$ -subcluster of each  $p$ -cluster (which is the red subcluster if one of them is red) are incident to a balancing edge.

For each  $1 \leq i \leq r_p$  we add the edges of  $B_i$  to  $G_i$ . So now  $E(G_i)$  consists of edges from each cluster to its unique successor on  $F_i$  together with the  $i$ -red edges incident to  $V_{0,i}$  and the balancing edges  $B_i$ .

**5.10. Almost decomposing into 1-factors.** Our aim now is to use Lemma 4.2 to find a  $\kappa$ -regular spanning subdigraph of each  $G_i$ . For this, the ‘balancing property’ achieved in Section 5.9 will be crucial.

Before this, for each  $i$ , we first remove a subdigraph  $H_{3,i}$  of  $G_i$ , which will be needed in Section 5.11. We do this as follows. For each edge  $E$  of  $F_i$ , recall that  $G_i(E)$  is  $(2\varepsilon', \beta_1)$ -superregular by (Red7). Apply Lemma 3.7(ii) to  $G_i(E)$  with parameters  $K := 2$  and  $\gamma_1 := \gamma^2 \beta_1, \gamma_2 := \beta_2$  where

$$(40) \quad \beta_2 := (1 - \gamma^2) \beta_1$$

to obtain two edge-disjoint subdigraphs of  $G_i(E)$ : a  $(2\varepsilon'^{1/12}, \gamma^2 \beta_1)$ -superregular digraph  $H_{3,i}(E)$  and a  $(2\varepsilon'^{1/12}, \beta_2)$ -superregular ‘remainder’ subdigraph which we still denote by  $G_i(E)$ . We let  $H_{3,i}$  have vertex set  $V(G)$  and edge set given by the union of  $H_{3,i}(E)$  over all edges  $E$  of  $F_i$ .

We now continue with finding a  $\kappa$ -regular spanning subdigraph of each  $G_i$ . Denote the collection of  $i$ -red edges incident to  $V_{0,i}$  by  $\mathcal{T}_i$ . For each  $1 \leq i \leq r_p$

we call the edges in  $\mathcal{T}_i \cup B_i$  and any  $p$ -cluster containing a vertex incident to such an edge  $i$ -red or red (so balancing edges are also regarded as red now). Write  $d_i^\pm(x) := d_{\mathcal{T}_i}^\pm(x) + d_{B_i}^\pm(x)$  for each  $x \in V(G_i)$  and define  $d_i^\pm(V) = \sum_{x \in V} d_i^\pm(x)$  for  $V \in V(F_i)$ . So by (31) we have that, for each  $V \in V(F_i)$ ,

$$(41) \quad d_i^\pm(V) = s_i^\pm(V) + d_{B_i}^\pm(V).$$

For each  $1 \leq i \leq r_p$  we now have the following properties:

- (Red0') There exists a sequence  $D_1 x_1 D_2 x_2 \dots x_{\ell-1} D_\ell x_\ell D_1$  with the following properties:
  - Each  $D_j$  is a cycle of  $F_i$  and every cycle of  $F_i$  appears at least once in the sequence;
  - $V_{0,i}^{\text{bridge}} = \{x_1, \dots, x_\ell\}$  and each  $x_j$  has exactly  $\kappa$  outneighbours in  $D_{j+1}$  and exactly  $\kappa$  inneighbours in  $D_j$ ;
- (Red1')  $d_i^\pm(x) = \kappa$  for each  $x \in V_{0,i}$ ;
- (Red2')  $V_{0,i}$  is an independent set in  $G_i$ ;
- (Red3')  $d_i^\pm(y) \leq \xi^{1/7} \beta_2 m_p$  for each  $y \in G_i \setminus V_{0,i}$ ;
- (Red4') For every red cluster  $V \in R_p$ , all  $i$ -red edges are incident to a single  $2p$ -cluster contained in  $V$ . In particular, at most  $m_p/2$  vertices in  $V$  are incident to an  $i$ -red edge;
- (Red5') In  $F_i$  any out-red  $p$ -cluster  $V$  is preceded by  $p-3$   $p$ -clusters which are neither out-red nor in-red, and is succeeded by an in-red  $p$ -cluster. Any in-red  $p$ -cluster  $V$  is succeeded by  $p-3$   $p$ -clusters which are neither out-red nor in-red, and is preceded by an out-red  $p$ -cluster;
- (Red6') Each  $p$ -cluster is either out-red, in-red or contains no vertices incident to a red edge;
- (Red7')  $G_1, \dots, G_{r_p}$  are edge-disjoint and  $G_i(E)$  is  $(2\varepsilon^{1/12}, \beta_2)$ -superregular for all  $E \in E(F_i)$ ;
- (B2'')  $d_i^+(V) = d_i^-(V^+)$  for all  $p$ -clusters  $V \in V(F_i)$ .

(Red0'), (Red1') and (Red2') follow immediately from (Red0), (Red1) and (Red2) respectively. (Red3') follows from summing the degrees given by (Red3) and (B1) and using (40). (Red4') is a consequence of (Red4) and our choice of edges in Section 5.9. (Red5') follows from (Red5) and (B2): indeed, the (red) clusters in  $T = T_{\text{in}} \cup T_{\text{out}}$  are separated by exactly  $p-1$  non-red clusters by (Red5), and by (B2), the only other red clusters are precisely those in  $T_{\text{in}}^- \cup T_{\text{out}}^+$ . (Red6') and edge-disjointness in (Red7') follow from (Red6) and edge-disjointness in (Red7), as well as the construction of  $B_i$  in Sections 5.8 and 5.9. The second part of (Red7') was verified directly after (40). (B2'') is a direct consequence of (B2) and (41). So for example, if  $V \in T_{\text{out}}$  then

$$d_i^+(V) = s_i^+(V) + c = d_{B_i}^-(V^+) = d_i^-(V^+).$$

Consider any edge  $E$  from  $V$  to  $V^+$  in  $F_i$ . We wish to find a subdigraph  $G_i(E)^*$  of  $G_i(E)$  such that, together with the red edges incident to  $V$  and  $V^+$ , every vertex in  $V$  has outdegree  $\kappa$  and every vertex in  $V^+$  has indegree  $\kappa$ . The union of these subdigraphs over all edges  $E \in E(F_i)$ , together with the red edges  $B_i \cup \mathcal{T}_i$ ,

will form a  $\kappa$ -regular spanning subdigraph  $G_i^*$  of  $G_i$ . (Recall that  $\kappa$  was defined in (20).)

Given any  $x \in V$ , let  $m_x^+ = d_i^+(x)$  and given any  $y \in V^+$ , let  $m_y^- = d_i^-(y)$ . By (Red3') we have that  $m_x^+, m_y^- \leq \xi^{1/7} \beta_2 m_p$  and by (B2'') we have that

$$\sum_{x \in V} m_x^+ = \sum_{y \in V^+} m_y^-.$$

Let  $\hat{\varepsilon} := 2\varepsilon^{1/12}$  and  $\hat{\beta} := \beta_2 - \hat{\varepsilon}$ . So (Red7') implies that  $G_i(E)$  is  $(\hat{\varepsilon}, \hat{\beta} + \hat{\varepsilon})$ -superregular for every  $E \in E(F_i)$ . Let

$$\hat{\alpha} := 1 - \frac{(1 - \gamma)\beta_1}{\beta_2 - \hat{\varepsilon}}.$$

So  $\kappa = (1 - \hat{\alpha})\hat{\beta}m_p$ , and it is easy to see that  $\gamma/2 \leq \hat{\alpha} \leq 2\gamma$ , so that  $\hat{\beta} \ll \hat{\alpha} \ll 1$ . Thus we can apply Lemma 4.2 to  $G_i(E)$  with  $\hat{\varepsilon}$  playing the role of  $\varepsilon$ ,  $\hat{\beta}$  playing the role of  $\beta$  and  $\hat{\alpha}$  playing the role of  $\alpha$ . Then we obtain a spanning subdigraph  $G_i(E)^*$  of  $G_i(E)$  in which each  $x \in V$  has outdegree  $\kappa - m_x^+$  and each  $y \in V^+$  has indegree  $\kappa - m_y^-$ . Then

$$G_i^* := \bigcup_{E \in E(F_i)} G_i(E)^* \cup B_i \cup \mathcal{T}_i$$

is a  $\kappa$ -regular spanning subdigraph of  $G_i$  as required. Moreover  $G_1^*, \dots, G_{r_p}^*$  are edge-disjoint subdigraphs of  $G$  by (Red7'). Now apply Proposition 4.3 to each  $G_i^*$  to obtain  $\kappa$  edge-disjoint 1-factors  $f_{i,1}, \dots, f_{i,\kappa}$  of each  $G_i$ .

**5.11. Merging 1-factors into Hamilton cycles.** The final step is to use edges disjoint from our collection of 1-factors to merge cycles such that each 1-factor is transformed into a Hamilton cycle. Then we will have found an approximate decomposition into edge-disjoint Hamilton cycles. The argument will be exactly the same for each  $G_i$ . So since we will work within a fixed  $G_i$ , we will label the  $\kappa$  factors obtained from  $G_i$  as  $f_1, \dots, f_\kappa$ . We wish to use Lemma 4.5 and edges from our pre-reserved digraph  $H_{3,i}$  to merge the cycles in each  $f_j$ .

We say that a non-red cluster is *black* and we say that an edge of  $F_i$  is *black* if both the initial cluster and final cluster are black. So for all black edges  $VV^+$  in  $F_i$  we have that  $f_j[V, V^+]$  is a perfect matching for each  $f_j$ , since in  $G_i$  every edge from a vertex in  $V$  goes to a vertex in  $V^+$ . (Red5') implies that every pair  $U_{\text{out}}U_{\text{in}}$  of consecutive red clusters on any cycle of  $F_i$  is followed by  $p - 3$  consecutive black clusters. Denote the path of length  $p - 4$  from the first of these black clusters to the last by  $I_U$ , so every edge in  $I_U$  is black. So we can choose  $p - 4$  disjoint sets of edges  $J_1, \dots, J_{p-4}$  of  $F_i$  so that for each pair of consecutive red clusters  $U_{\text{out}}U_{\text{in}}$ ,  $J_q$  contains exactly one edge of  $I_U$ . So each  $J_q$  consists of exactly  $|T| = |T_{\text{in}}| + |T_{\text{out}}| = (s - 2)\tilde{L}$  edges of  $F_i$  and has non-empty intersection with any cycle of  $F_i$ .

The idea is to apply Lemma 4.5 repeatedly to transform each of the  $f_j$  into a Hamilton cycle. Each time  $H_{3,i}$  will play the role of  $G$ , and each  $J_q$  will play

the role of  $J$  roughly  $\kappa/p$  times. If  $\mathcal{E}$  is a set of edges in  $F_i$ , we write  $H_{3,i}(\mathcal{E}) := \bigcup_{E \in \mathcal{E}} H_{3,i}(E)$ .

We now describe the merging procedure for  $f_1$ . Denote the cycles of  $F_i$  by  $D_1, \dots, D_\ell$ . Let  $K_1$  be the 1-regular digraph consisting of all cycles of  $f_1$  which contain a vertex in a cluster of  $D_1$ . Now apply Lemma 4.5 as follows:  $D_1$  plays the role of  $C$ ,  $J_1 \cap E(D_1)$  plays the role of  $J$ ,  $K_1$  plays the role of  $F$  and  $H_{3,i}(J_1)$  plays the role of  $G$ .

Condition (i) in Lemma 4.5 is clearly satisfied since every edge of  $J_1$  is black. To verify condition (ii), let  $D$  be any cycle of  $K_1$ . We claim that  $D$  contains a vertex  $x$  from a black cluster  $B$ . To see this, suppose that  $D$  contains a vertex  $y$  which lies in an in-red cluster. Then the next vertex of  $D$  lies in a black cluster. Similarly, if  $y$  lies in an out-red cluster, then the vertex preceding  $y$  on  $D$  lies in a black cluster, which proves the claim. Now let  $I_U$  be the black interval containing  $B$ ; then there is a path in  $D$  (containing  $x$ ) which contains at least one vertex from each cluster in  $I_U$ . But  $J_1 \cap E(D_1)$  contains an edge of  $I_U$ , as required.

To verify (iii), let  $VV^+$  and  $WW^+$  be edges of  $J_1 \cap E(D_1)$  such that  $J_1$  avoids all edges in the segment  $V^+D_1W$ . Then there is exactly one pair of successive red clusters  $U_{\text{out}}U_{\text{in}}$  in this segment. So for each  $v_a \in V^+$  there is a path  $P_a$  in  $f_1$  from  $v_a$  to a distinct vertex  $u_a^{\text{out}}$  in  $U_{\text{out}}$  which winds around  $D_1$ . Similarly, for each  $u_{a'}^{\text{in}} \in U_{\text{in}}$  there is a path  $P_{a'}'$  in  $f_1$  from  $u_{a'}^{\text{in}}$  to a distinct vertex  $w_{a'} \in W$  which winds around  $D_1$ . But by (Red4'), for at least half of the vertices  $u_a^{\text{out}} \in U_{\text{out}}$ , there is an edge in  $f_1$  to some  $u_{a'}^{\text{in}} \in U_{\text{in}}$ . So  $f_1$  contains at least one path  $v_a P_a u_a^{\text{out}} u_{a'}^{\text{in}} P_{a'}' w_{a'}$  from  $v_a \in V^+$  to  $w_{a'} \in W$  which winds around  $D_1$ , as required.

So we can find a matching  $M_1$  in  $H_{3,i}(J_1)$  and a cycle  $C_1$  with  $V(C_1) = V(K_1)$  and  $E(C_1) \subseteq K_1 \cup M_1$ . We replace the 1-regular subdigraph  $K_1$  of  $f_1$  by  $C_1$ . We call the resulting 1-factor  $f_1(1)$  and we denote  $H_{3,i} \setminus M_1$  by  $H_{3,i}^2$ . Note that all cycles of  $f_1$  which contained a vertex in  $D_1$  have now been merged into a single cycle of  $f_1(1)$ .

For  $2 \leq k \leq \ell$  we define  $f_1(k)$  inductively as follows. Let  $K_k$  be the 1-regular digraph consisting of all cycles of  $f_1(k-1)$  which contain a vertex in a cluster of  $D_k$ . Now let  $D_k$  play the role of  $C$ ,  $J_1 \cap E(D_k)$  play the role of  $J$ ,  $K_k$  play the role of  $F$  and  $H_{3,i}(J_1)$  play the role of  $G$ . Note that the  $k$  choices  $J_1 \cap E(D_{k'})$  with  $1 \leq k' \leq k$  playing the role of  $J$  so far are pairwise vertex-disjoint. Exactly as above, the conditions (i)-(iii) are satisfied and we can apply Lemma 4.5 to obtain a 1-factor  $f_1(k)$  in which all cycles containing a vertex in  $D_k$  have been merged. Moreover if two vertices  $x$  and  $y$  lie on a common cycle of  $f_1(k-1)$  they lie on a common cycle of  $f_1(k)$ . We repeat this for all  $1 \leq k \leq \ell$  to obtain  $f_1' := f_1(\ell)$ . We will see below that  $f_1'$  is a Hamilton cycle.

We now aim to carry out a similar procedure for  $f_2, \dots, f_\kappa$  to obtain  $f_2', \dots, f_\kappa'$ . The approach will be to use  $J_1$  for  $f_1, \dots, f_{\kappa'}$  where  $\kappa' := \kappa/(p-4)$  and more generally to use  $J_q$  for  $f_{(q-1)\kappa'+1}, \dots, f_{q\kappa'}$ . Note that, to obtain  $f_1'$ , we removed exactly one perfect matching from each  $H_{3,i}(E)$  for each edge  $E$  of  $J_1$ . To reuse  $J_1$  we need only check that, at each step and for each edge  $E$  of  $J_1$ , the remainder of the sparse digraph  $H_{3,i}(E)$  satisfies the conditions required of  $G$  in Lemma 4.5.

For this, let  $H_{3,i}^t(J_q)$  denote a subdigraph of  $H_{3,i}(J_q)$  obtained by removing  $t$  arbitrary perfect matchings from  $H_{3,i}(E)$  for each  $E \in J_q$ .

**Claim 5.7.** *Let  $\kappa'$  be defined as above and let  $\varepsilon^* := 2\sqrt{\beta_1/p}$ . Then  $H_{3,i}^{\kappa'}(E)$  is  $(\varepsilon^*, \gamma^2\beta_1)$ -superregular whenever  $E$  is an edge in  $J_q$ , where  $1 \leq q \leq p-4$ .*

To see this, it suffices to consider a single edge  $E = XY$  in  $J_1$ . Write  $H := H_{3,i}^{\kappa'}(E)$ . Then, since at each stage we removed a perfect matching, in total we removed  $\kappa'$  edges incident to each vertex in  $X \cup Y$ , which is at most  $\beta_1 m_p/p$  by (20). Since  $H_{3,i}(E)$  is  $(2\varepsilon^{1/12}, \gamma^2\beta_1)$ -superregular (see directly after (40)), we can apply Proposition 3.6(ii) with  $H_{3,i}(E)$  playing the role of  $G$ ,  $H$  playing the role of  $G'$  and  $d' := \beta_1/p$  to find that  $H$  is  $(\varepsilon^*, \gamma^2\beta_1)$ -superregular. Note that  $\varepsilon^* \ll \gamma^2\beta_1$  by (4). This proves the claim.

Suppose that we have constructed  $f'_1, \dots, f'_t$  with  $t < \kappa'$  in the same way as  $f'_1$ . Then we will have used  $t$  perfect matchings in  $H_{3,i}(E)$  for each  $E \in J_1$ . Let  $H_{3,i}^t(J_1)$  denote the subdigraph of  $H_{3,i}(J_1)$  consisting of the remaining edges. Then Claim 5.7 implies that  $H_{3,i}^t(J_1)$  can still play the role of  $G$  in Lemma 4.5. So we can construct  $f'_{t+1}$  in the same way as  $f'_1$ . Thus we can obtain  $f'_1, \dots, f'_{\kappa'}$  as described above.

Now for each  $2 \leq q \leq p-4$  and each  $1 \leq t \leq \kappa'$  we can use  $J_q$  to obtain  $f'_{(q-1)\kappa'+t}$  from  $f'_{(q-1)\kappa'+t}$  in exactly the same way (except that we use edges from  $H_{3,i}(J_q)$  and so  $J_q \cap E(D_k)$  now plays the role of  $J$  for  $1 \leq k \leq \ell$ ).

We have now obtained  $f'_1, \dots, f'_{\kappa'}$ . They are clearly edge-disjoint 1-factors. We claim that  $f'_j$  is a Hamilton cycle for each  $1 \leq j \leq \kappa'$ . Indeed, suppose not. It suffices to consider  $f'_1$ . Let  $C$  and  $C'$  be cycles in  $f'_1$  where  $C$  contains a vertex  $x$  in some cycle  $D$  of  $F_i$  and  $C'$  contains a vertex  $x'$  in some cycle  $D'$  of  $F_i$ . Recall that, by our construction, for all cycles  $D_k$  in  $F_i$ , every vertex in (a cluster of)  $D_k$  is contained in a single cycle in  $f'_1$ . Consider the sequence given by (Red0') as a cyclic sequence and pick an interval

$$D_g x_g D_{g+1} x_{g+1} \dots x_{g'-1} D_{g'} x_{g'}$$

such that  $D = D_g$  and  $D' = D_{g'}$ . By (Red0') and (Red1'), the inneighbour of  $x_g$  in  $f'_1$  is contained in  $D$ , so  $x_g \in V(C)$ . But similarly the outneighbour of  $x_g$  in  $f'_1$  is contained in  $D_{g+1}$ , so all vertices lying in a cluster of  $D_{g+1}$  are contained in  $V(C)$  and thus  $x_{g+1} \in V(C)$ . Continuing along the subsequence we conclude that every vertex lying in a cluster of  $D'$  lies on  $C$ . So  $x'$  lies on both  $C'$  and  $C$ ; so since  $f'_1$  is a 1-factor we must have  $C = C'$ . Thus  $f'_1$  is a Hamilton cycle, and the same holds for  $f'_2, \dots, f'_{\kappa'}$ .

Finally, we can bound the total number of Hamilton cycles as follows. Note that

$$\begin{aligned} \kappa &\stackrel{(8),(13),(20)}{=} (1-\gamma)(1-5\gamma)\beta \frac{m}{sp}. \\ r_p &\stackrel{(7),(16),(17)}{=} (s-1)(p-1)(\tilde{\alpha}-\gamma) \frac{\tilde{L}}{\beta} \geq (1-\sqrt{\gamma})sp \frac{\tilde{\alpha}\tilde{L}}{\beta}. \end{aligned}$$

## APPROXIMATE HAMILTON DECOMPOSITIONS OF REGULAR EXPANDERS

So altogether, after repeating the procedure for every  $1 \leq i \leq r_p$ , we have found

$$\begin{aligned} r_p \kappa &\geq (1 - \gamma)(1 - 5\gamma)(1 - \sqrt{\gamma})\tilde{\alpha}\tilde{L}\tilde{m} \\ &\stackrel{(10)}{\geq} (1 - \sqrt{\gamma})^3(1 - \varepsilon)\tilde{\alpha}n \\ &\stackrel{(4)}{\geq} (1 - \eta)r \end{aligned}$$

edge-disjoint Hamilton cycles, as required. This completes the proof of Theorem 1.1.

## 6. THE PROOF OF COROLLARY 1.2

We now use Theorem 1.1 and Lemma 4.1 to prove Corollary 1.2.

*Proof.* As in the proof of Theorem 1.1, we may assume without loss of generality that  $0 < \eta \ll \nu \ll \tau \ll \alpha$ . Choose  $n_0$  and  $\gamma$  so that  $0 < 1/n_0 \ll \gamma \ll \eta$ . Suppose that  $G$  is a digraph on  $n \geq n_0$  vertices satisfying (i) and (ii). Let

$$n_x^\pm := d_G^\pm(x) - (\alpha - \sqrt{\gamma})n$$

for each  $x \in V(G)$ . We apply Lemma 4.1 to  $G$  with  $\rho = \varepsilon = \sqrt{\gamma}$  and with  $Q = G$  (so  $q = 1$ ) to obtain a subdigraph  $H$  of  $G$  such that  $\tilde{G} := G \setminus H$  is an  $(\alpha - \sqrt{\gamma})n$ -regular digraph on  $n$  vertices. Note that for all  $x \in V(G)$  we have  $d_{\tilde{G}}^-(x) \geq d_G^-(x) - (\sqrt{\gamma} - \gamma)n \geq d_G^-(x) - \nu n/2$ . So for all sets  $S$  of vertices,

$$RN_{\nu/2, \tilde{G}}^+(S) \supseteq RN_{\nu, G}^+(S).$$

Thus  $\tilde{G}$  is a robust  $(\nu/2, \tau)$ -outexpander. Therefore we can apply Theorem 1.1 to  $\tilde{G}$  with parameter  $\eta' := \eta/2\alpha$  to find  $(1 - \eta')(\alpha - \sqrt{\gamma})n > (\alpha - \eta)n$  edge-disjoint Hamilton cycles in  $\tilde{G}$  and hence in  $G$ .  $\square$

## 7. ACKNOWLEDGEMENTS

We are extremely grateful to Daniela Kühn for helpful discussions throughout the project.

## REFERENCES

- [1] N. Alon, G. Gutin and M. Krivelevich, Algorithms with large domination ratio, *J. of Algorithms* **50** (2004), 118–131.
- [2] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *J. of Computer and System Sciences* **69** (2004), 354–382.
- [3] B. Alspach, The wonderful Walecki construction, *Bull. Inst. Combin. Appl.* **52** (2008), 7–20.
- [4] B. Alspach, J.C. Bermond and D. Sotteau, Decompositions into cycles. I. Hamilton decompositions, *Cycles and rays (Montreal, PQ, 1987)*, Kluwer Acad. Publ., Dordrecht, 1990, 9–18.
- [5] D. Bal and A. Frieze, Packing tight Hamilton cycles in uniform hypergraphs, *SIAM J. Discrete Mathematics* **26** (2012), 435–451.
- [6] D. Christofides, P. Keevash, D. Kühn and D. Osthus, A semi-exact degree condition for Hamilton cycles in digraphs, *SIAM J. Discrete Mathematics* **24** (2010), 709–756.
- [7] A. Frieze and M. Krivelevich, On packing Hamilton cycles in  $\varepsilon$ -regular graphs, *J. Combin. Theory B* **94** (2005), 159–172.

- [8] A. Frieze, M. Krivelevich and P.-S. Loh, Packing tight Hamilton cycles in 3-uniform hypergraphs, *Random Structures Algorithms* **40** (2012), 269–300.
- [9] F. Glover and A.P. Punnen, The Travelling Salesman Problem: new solvable cases and linkages with the development of approximation algorithms, *J. of the Operational Research Society* **48** (1997), 502–510.
- [10] G. Gutin and A. Yeo, TSP tour domination and Hamilton cycle decompositions of regular graphs, *Oper. Res. Letters* **28** (2001) 107–111.
- [11] P. Hall, On representatives of subsets, *J. London Math. Soc.* **10** (1935), 26–30.
- [12] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley-Interscience, 2000.
- [13] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: a proof of Kelly’s conjecture for large tournaments, preprint.
- [14] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: applications, preprint.
- [15] D. Kühn, D. Osthus and A. Treglown, Hamiltonian degree sequences in digraphs, *J. Combin. Theory B* **100** (2010), 367–380.
- [16] D. Kühn, D. Osthus and A. Treglown, Hamilton decompositions of regular tournaments, *Proc. London Math. Soc.* **101** (2010), 303–335.
- [17] E. Lucas, *Récréations Mathématiques*, Vol. 2, Gauthier-Villars, 1892.
- [18] J.W. Moon, *Topics on tournaments*, Holt, Rinehart and Winston, New York, 1968.
- [19] A. Strivastav and P. Stangier, Algorithmic Chernoff-Hoeffding inequalities in integer programming, *Random Structures Algorithms* **8** (1996), 27–58.
- [20] T.W. Tillson, A Hamiltonian decomposition of  $K_{2m}^*$ ,  $2m \geq 8$ , *J. Combin. Theory B* **29** (1980), 68–74.

Deryk Osthus, Katherine Staden  
 School of Mathematics  
 University of Birmingham  
 Edgbaston  
 Birmingham  
 B15 2TT  
 UK

*E-mail addresses:* {d.osthus, kls103}@bham.ac.uk