THE RANK THREE CASE OF THE HANNA NEUMANN CONJECTURE

Warren Dicks* and Edward Formanek**

Abstract. For a free group G, $\mathrm{rk}(G)$ denotes the rank of G, and, for each positive integer n, $\mathrm{rk}_{-n}(G)$ denotes $\max\{\mathrm{rk}(G) - n, 0\}$. Let H and K be finitely generated subgroups of a free group. Hanna Neumann conjectured that $\mathrm{rk}_{-1}(H \cap K) \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K)$. We prove that

$$\mathrm{rk}_{-1}(H \cap K) \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K) + \mathrm{rk}_{-3}(H)\,\mathrm{rk}_{-3}(K).$$

This extends results of Hanna Neumann, R. G. Burns and G. Tardos, and shows that, if H has rank three or less, then the conjectured inequality holds. Our argument consists of proving the corresponding case of the Amalgamated Graph Conjecture, and therefore applies to Walter Neumann's strengthened version of the Hanna Neumann Conjecture.

§0. Intersections of subgroups of free groups

Let G be a free group, and H, K be finitely generated subgroups of G. Let

$$\sum = \sum_{HgK \in H\backslash G/K} \mathrm{rk}_{-1}(H^g \cap K),$$

where the summation is over the set of (H,K) double cosets in G, with each double coset HgK contributing $\mathrm{rk}_{-1}(H^g \cap K)$, a value which does not depend on the choice of representative g of the double coset. Walter Neumann's strengthened version of the Hanna Neumann Conjecture is the conjecture that $\sum \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K)$. Combined work of Hanna Neumann [3] and Walter Neumann [4] shows that $\sum \leq 2\,\mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K)$; in particular, $\sum$ is finite. It follows from work of R. G. Burns [1] that

$$\sum \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K) + \max\{\mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-2}(K), \mathrm{rk}_{-2}(H)\,\mathrm{rk}_{-1}(K)\}.$$

*Partially supported by the DGES (Spain) through grant number PB96-1152.

**Partially supported by the National Science Foundation (USA), and the Ministerio de Educación y Ciencies (Spain) through a grant held at the Centre de Recerca Matemàtica of the Institut d'Estudis Catalans.

G. Tardos [5], [6] recently improved this to

$$\sum \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K) + \max\{\mathrm{rk}_{-2}(H)\,\mathrm{rk}_{-2}(K) - 1, 0\}.$$

We shall further improve this to

$$\sum \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K) + \mathrm{rk}_{-3}(H)\,\mathrm{rk}_{-3}(K).$$

In summary then, our main result is the following.

0.1 Theorem. *If G is a free group, and H and K are finitely generated subgroups of G, then*

$$\sum_{HgK \in H\backslash G/K} \mathrm{rk}_{-1}(H^g \cap K) \leq \mathrm{rk}_{-1}(H)\,\mathrm{rk}_{-1}(K) + \mathrm{rk}_{-3}(H)\,\mathrm{rk}_{-3}(K). \quad \square$$

By [2, Theorem 4.1], this is equivalent to a graph-theoretic result which we will state in the next section and prove in the remainder of the article.

§1. The graph-theoretic setting

We begin by recalling the basic concepts used in [2], and recording some related concepts.

1.1 Definition. A *finite bipartite graph* $X = (V_r, V_y, E, r, y)$ is defined to be a five-tuple consisting of three finite disjoint sets V_r, V_y, E, and two maps $r\colon E \to V_r$, $y\colon E \to V_y$. We understand that, as a set, X is $V_r \vee V_y \vee E$, where the symbol $\vee$ denotes the disjoint union.

If $X = (V_r, V_y, E, r, y)$ and $X' = (V_r', V_y', E', r'y')$ are finite bipartite graphs, then an injective map of sets $X \to X'$ is said to be an *injective morphism* of finite bipartite graphs if it carries V_r (resp. V_y, resp. E) to V_r' (resp. V_y', resp. E'), and forms a commuting triangle with r, r', and a commuting triangle with y, y'. We define *isomorphism* and *isomorphic* in the natural way. $\square$

1.2 Definitions. Let $X = (V_r, V_y, E, r, y)$ be a finite bipartite graph.

The elements of V_r (resp. V_y, resp. E, resp $V_r \vee V_y$) are called the *red vertices* (resp. *yellow vertices*, resp. *edges*, resp. *vertices*). The *incidence map* $(r, y)\colon E \to V_r \times V_y$ associates to each edge the ordered pair consisting of its red and yellow vertices. We depict an edge e as an arc joining $r(e)$ and $y(e)$, and we say that e is *incident* to $r(e)$ and to $y(e)$.

We use the prefix X- or the post-modifier "of X" where we wish to emphasize the ambient graph. Thus we speak of X-edges and edges of X.

We write $EX = E$ and $VX = V_r \vee V_y$.

The *valence* of an X-vertex v is the number of X-edges incident to v.

The *size* of X is the triple of natural numbers

$$\text{size}(X) = (|V_r|, |V_y|, |E|),$$

where $|E|$ denotes the cardinal of the set E.

A subset X' of X is called a *subgraph* of X if $r(X' \cap E) \cup y(X' \cap E) \subseteq X'$, that is, every element of $X' \cap E$ has both its vertices in X'. In this event, X' is a finite bipartite graph in a natural way. If, moreover, every element of E which has both its vertices in X' itself lies in X', then X' is said to be a *full* subgraph of X.

We say that X is *disconnected* if X is the union of two disjoint nonempty subgraphs; otherwise, X is said to be *connected*. The maximal connected subgraphs of X are called the *components* of X.

We say that X *breaks into two equal parts* if the components of X are isomorphic in pairs, that is, there is an even number of elements in each equivalence class of the equivalence relation "is isomorphic to", on the set of components of X. Recall that "isomorphic" always means "isomorphic as bipartite graphs".

If the incidence map of X is injective then we say that X is *simple-edged*; in this case there is at most one X-edge joining any two vertices, and we view E as a subset of $V_r \times V_y$. If the incidence map is bijective then we say that X is *complete*.

If X is simple-edged, then we denote by completion(X) the complete finite bipartite graph having the same red and yellow vertex sets as X, so the edge set is $V_r \times V_y$, and X is a subgraph of completion(X). An edge of completion(X) is called an X-*bond* (or a *bond of* X). If $\mathfrak{X}$ is a set of subsets of VX, an $\mathfrak{X}$-*bond* is an X-bond which is a bond of at least one element of $\mathfrak{X}$.

Let D be a finite bipartite graph. We say that X is a finite bipartite D-*graph* if there is specified an injective morphism $D \to X$. If A and B are finite bipartite D-graphs, then their D-*amalgam*, denoted $A \vee_D B$, is defined to be the finite bipartite D-graph obtained from $A \vee B$ by identifying the two copies of D. $\square$

Our main graph-theoretic result, proved over the remainder of the article, is the following.

1.3 Theorem. *Let D be a finite bipartite graph, and let $(m, n, p) = \mathrm{size}(D)$. If there exist finite bipartite (simple-edged) D-graphs A, B, C, such that (the finite bipartite graph) $(A \vee_D B) \vee (B \vee_D C) \vee (C \vee_D A)$ is simple-edged and breaks into two equal parts, then*

$$\frac{p}{2} \leq \frac{m}{2}\frac{n}{2} + \max\{\frac{m}{2} - 2, 0\} \max\{\frac{n}{2} - 2, 0\}. \quad \square$$

Combining Theorem 1.3 and [2, Theorem 4.1] gives Theorem 0.1.

The Amalgamated Graph Conjecture is the conjecture that if the hypotheses of Theorem 1.3 hold, then $p \leq \frac{1}{2}mn$; by [2, Theorem 2.5], this is equivalent to the strengthened version of the Hanna Neumann Conjecture.

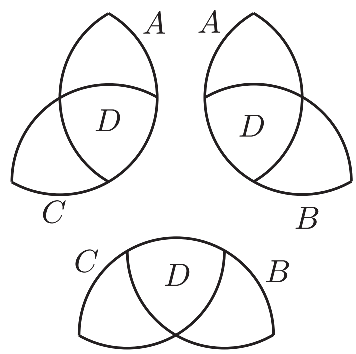

Fig. 1. The "amalgamated graph" $(A \vee_D B) \vee (B \vee_D C) \vee (C \vee_D A)$.

We will argue by contradiction, so we assume the following.

1.4 Hypotheses. Throughout the remainder of the article, suppose there are given natural numbers m, n, p, such that

$$p > \frac{1}{2}(mn + \max\{m - 4, 0\} \max\{n - 4, 0\}).$$

Suppose also that there is given a finite bipartite simple-edged graph D of size (m, n, p), and finite bipartite simple-edged D-graphs A, B, C, such that $AB \vee BC \vee CA$ is simple-edged and breaks into two equal parts, where AB denotes the D-amalgam $A \vee_D B$, and similarly for BC, CA. We set $ABC = A \vee_D B \vee_D C$, and view AB, BC, CA as (not necesssarily full) subgraphs of ABC. $\square$

It will become clear during the course of the proof that D has one very large component and a few small components, and these can be either single yellow vertices, or a single red vertex joined to a few yellow vertices, but there is at most one component of the latter type.

1.5 Lemma. (i) *The natural numbers m, n, and p are even and positive, so* $p \geq 2 + \frac{1}{2}(mn + \max\{m-4, 0\} \max\{n-4, 0\})$.

(ii) *If $m < 4$ then $m = 2$, and $p \geq n + 2$, and the two red vertices have total valence at least 4.*

(iii) *If $n < 4$ then $n = 2$, and $p \geq m + 2$, and the two yellow vertices have total valence at least 4.*

(iv) *If $m \geq 4$ and $n \geq 4$ then $p \geq (m-2)(n-2) + 6$, and either every pair of red vertices of D has total valence at least 3, or every pair of yellow vertices of D has total valence at least 3.*

Proof. (i). By counting edges and red and yellow vertices of A, B, C, and D, it is straightforward to prove that m, n and p are even. It is clear that $p > 0$, so $m > 0$ and $n > 0$. Also, $p - \frac{1}{2}(mn + \max\{m-4, 0\} \max\{n-4, 0\})$ is a positive even integer.

(ii). Suppose $m < 4$. By (i), $m = 2$, $n \geq 2$, and $p \geq 2 + \frac{1}{2}mn = n + 2 \geq 4$. Thus D has two red vertices and at least four edges, so the two red vertices have total valence at least four.

(iii) is similar to (ii).

(iv). Suppose $m \geq 4$ and $n \geq 4$. Then

$$p \geq 2 + \frac{1}{2}(mn + (m-4)(n-4)) = (m-2)(n-2) + 6.$$

(This can be seen as a special case of the identity

$$ab + gh = \frac{1}{2}(a+g)(b+h) + \frac{1}{2}(a-g)(b-h),$$

which we will use again in §6.) Now assume that the statement is false, so D contains a pair of red vertices of total valence at most two, and a pair of yellow vertices of total valence at most two. These four vertices are then incident to some number p' of edges, and $p' \leq 4$. On deleting the four vertices and the p' edges we get a simple-edged subgraph of D of size $(m-2, n-2, p-p')$; thus $p - p' \leq (m-2)(n-2) \leq p - 6$, so $p' \geq 6$, a contradiction. □

Thus, invoking red-yellow symmetry, we may, and shall, assume the following.

1.6 Additional hypotheses. Every pair of red vertice of D has total valence at least three, and if $m \geq 4$ then $n \geq 4$. □

Tardos' result [6] covers the case $m = 2$, and part of our proof will reproduce his proof.

We conclude this section with some useful terminology.

1.7 Definitions. We say that a triple (g, h, i) of natural numbers is *Type I* if $g = i = 0$, *Type II* if $g = 1$, $h = i = 0$, *Type III* if $g = h = i = 1$, *Type IV* if $g = h = 1$, $i = 0$, and *Type V* if $g = 1$, $h = i \geq 2$.

In other words, $(0, h, 0)$, $h \geq 0$, are Type I, $(1, 0, 0)$ is Type II, $(1, 1, 1)$ is Type III, $(1, 1, 0)$ is Type IV, and $(1, h, h)$, $h \geq 2$, are Type V.

We say (g, h, i) is Type I-III if (g, h, i) is a triple of natural numbers of Type I, or II, or III; and similarly for ranges other than I-III. □

1.8 Definitions. For a positive integer N, and a set S, a family

$$(S_i \mid i = 1, \dots, N)$$

of subsets of S is called an *N-partition of S* if the S_i are pairwise disjoint and the union of all the S_i is S. Each S_i is called a *part*, or *cell*, of the N-partition of S. Notice that a part can be empty.

Having an N-partition of S is equivalent to having a function from S to $\{1, \dots, N\}$. It is also equivalent to having an expression of S as the disjoint union of N sets. □

§2. Atomic factors

We now introduce terminology and concepts that will be fundamental to the proof.

Throughout this section let X be a finite bipartite graph, and (g, h, i) be Type I-V.

We will be analyzing graphs by the ways they can be broken up as unions of two proper subgraphs with intersection of a specific type. There are three cases.

2.1 Definitions. By a *break* of X, we mean a pair (V_1, V_2) of subsets of VX such that $V_1 \cup V_2 = VX$.

Let (V_1, V_2) be a break of X, and let X_i denote the full subgraph of X with vertex set V_i, for $i = 1, 2$.

The break is *proper* if both V_1 and V_2 are proper subsets of VX.

The break and an X-bond e are (mutually) *compatible* if e does not join $V_1 - V_2$ to $V_2 - V_1$.

Thus the break is compatible with the X-edges if and only if $X = X_1 \cup X_2$.

We say that X is *(g, h)-trivial* if X has at most g red vertices and at most h yellow vertices; otherwise X is *(g, h)-nontrivial.*

(i). If (g, h, i) is Type I-III, that is, $(0, h, h)$, $h \geq 1$, or $(1, 0, 0)$, or $(1, 1, 1)$, then we say that (V_1, V_2) is a *(g, h, i)-break* of X (or with respect to the X-edges) if it is compatible with the X-edges and $V_1 \cap V_2$ is (g, h)-trivial. See Fig. 2.

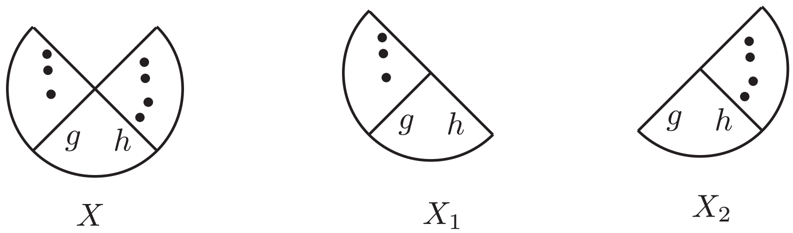

Fig. 2. A (g,h,gh)-break for $gh \leq 1$.

(ii). If (g,h,i) is Type V, that is, $(1,h,h)$, $h \geq 2$, then we say that (V_1,V_2) is a (g,h,i)-*break* of X (or with respect to the X-edges) if it is compatible with the X-edges and $V_1 \cap V_2$ is 2-partitioned, and the first part is (g,h)-trivial, and the second part consists of (any number of) yellow vertices which are joined by at most one X-edge to $V_1 - V_2$, by at most one X-edge to $V_2 - V_1$, and by no X-edges to $V_1 \cap V_2$. See Fig. 3.

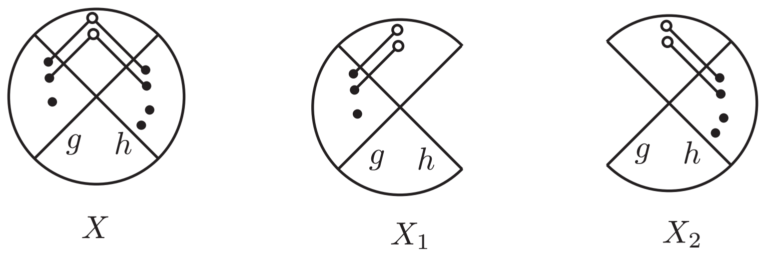

Fig. 3. A (g,h,gh)-break for $g=1$, $h \geq 2$.

(iii). We say that (V_1,V_2) and (V_2,V_1) are $(1,1,0)$-*breaks* of X (or with respect to the X-edges) if they are compatible with the X-edges and $V_1 \cap V_2$ is 3-partitioned, and the third part consists of (any number of) yellow vertices which are joined by at most one edge to $V_1 - V_2$, by at most one edge to $V_2 - V_1$, and by no edges to $V_1 \cap V_2$, while the second part is $(1,1)$-trivial, and the third part is a $(0,1)$-trivial set joined by at most one edge to $V_1 - V_2$. If this $(0,1)$-trivial set is nonempty, it consists of a yellow vertex called the *crossover vertex* of the break, and, if, moreover, it is joined by an edge to $V_1 - V_2$, this edge is called the *crossover edge* of the break. See Fig. 4.

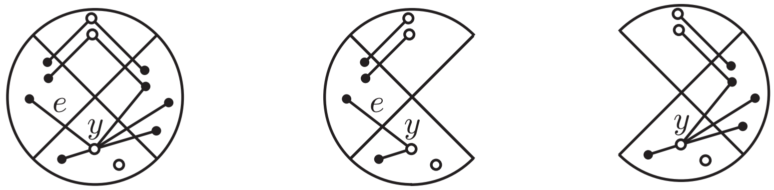

Fig. 4. A $(1, 1, 0)$-break, with crossover vertex y and crossover edge e.

The red-yellow asymmetry in this definition stems from the red-yellow asymmetry of Hypotheses 1.6.

Our collection of types of breaks is partially ordered, corresponding to a lexicographic ordering of triples which uses the usual ordering in the first two coordinates and the reverse ordering in the third coordinate. Thus, if $0 \leq g \leq g' \leq 1$, $0 \leq h \leq h'$ then each (g, h, gh)-break is a $(g', h', g'h')$-break. Each $(1, 1, 1)$-break is a $(1, 1, 0)$-break. Each $(1, 1, 0)$-break is a $(1, 2, 2)$-break.

Given a diagram as in Figs. 2-4, or its left-right reflection, we see that VX is 4-partitioned; reading counter-clockwise, starting at the bottom, we call the four cells the *south, east, north* and *west cells* of the break. The division of $V_1 \cap V_2$ into two cells will be useful for counting purposes. The situation of Fig. 4 is called an *eastern* $(1, 1, 0)$-break, since the crossover vertex is strongly attached to the east. The left-right dual situation is called a *western* $(1, 1, 0)$-break.

Suppose (V_1, V_2) is a (g, h, i)-break of X. Each northern vertex has, among all the bonds incident to it, (at most) two distinguished ones, one joined to the east cell and one to the west cell. Each crossover vertex has, among all the bonds incident to it, (at most) two distinguished ones, the crossover edge (joining it to the east or west cell) and the bond joining it to (the red vertex in) the south cell. These distinguished bonds form an important part of the data of a break. An X-bond e is said to *interfere* with the break if either e is not compatible with the break, so joins east to west, or e is incident to a northern or crossover vertex but is not one of its (at most two) distinguished bonds.

A yellow vertex together with three or less incident bonds will be called a *tripod*. We define *bipod*, *unipod*, and *nullipod* in the corresponding way. Notice that adding bipods does not impede Type IV-V breaks. □

Type I-II breaks were used in [2], [6], while Type III breaks were introduced in [6].

We can now define the (g, h, i)-atomic factors of X.

2.2 Definitions. We say that X is (g, h, i)-*decomposable* if X is (g, h)-nontrivial, and has a proper (g, h, i)-break. For example, $(0, 0, 0)$-decomposable is the same as disconnected.

We say that X is (g, h, i)-*atomic* if X is (g, h)-nontrivial and has no proper (g, h, i)-break. For example, being $(0, 0, 0)$-atomic is the same as being nonempty and connected.

Let $X_0 = X$. Suppose we have defined, for some integer $N \geq 0$, a subgraph X_N of completion(X); we define X_{N+1} to be the subgraph of completion(X) obtained by adding to X_N the bonds of all the (g, h, i)-atomic subgraphs of X_N. This recursively defines an ascending chain of subgraphs of completion(X), and we let X_∞ denote the union of this (eventually constant) chain. By a (g, h, i)-*atomic factor* of X we mean the vertex set of a maximal (g, h, i)-atomic subgraph of X_∞. The (g, h, i)-*atomic factorization* of X is the set of all (g, h, i)-atomic factors of X. $\square$

It will be convenient not to have edges in the (g, h, i)-atomic factors, although it would be natural to take the full subgraphs.

In deducing a contradiction from Hypotheses 1.4 and 1.6, we shall not use the full force of the fact that the graph $AB \vee BC \vee CA$ breaks into two equal parts, but shall use only Lemma 1.5 and the fact that, for each (g, h, i) of Type I-V, $AB \vee BC \vee CA$ has an even number of (g, h, i)-atomic factors.

Notice that if $\mathfrak{X}$ is the (g, h, i)-atomic factorization wiht respect to the X-edges, then it is also the (g, h, i)-atomic factorization with respect to the bonds which are X-edges or $\mathfrak{X}$-bonds.

2.3 Definitions. By a (g, h, i)-*factorization of* X we mean a set $\mathfrak{X}$ of subsets of VX such that $\mathfrak{X}$ is the (g, h, i)-atomic factorization with respect to the bonds which are X-edges or $\mathfrak{X}$-bonds.

Let $\mathfrak{Y}$, $\mathfrak{Z}$ be sets of subsets of VX. For each $Y \in \mathfrak{Y}$, let $\mathfrak{Z}_Y$ be the set of elements of $\mathfrak{Z}$ which lie in Y. We say that $\mathfrak{Z}$ is a (g, h, i)-*factorization of* $\mathfrak{Y}$ if $(\mathfrak{Z}_Y \mid Y \in \mathfrak{Y})$ is a $\mathfrak{Y}$-partition of $\mathfrak{Z}$, and, for each $Y \in \mathfrak{Y}$, $\mathfrak{Z}_Y$ is a (g, h, i)-factorization of Y with respect to the set of all Y-bonds which are X-edges or $\mathfrak{Z}$-bonds.

So far we have looked at factorizations from the constructive viewpoint, but now we take the destructive viewpoint.

2.4 Definitions. We will not write out a formal definition of a *factorization tree of* X, but just describe it informally as a finite process which does the following. It starts with VX. Either it stops here, or it breaks VX into two subsets V_1, V_2 whose union is all of VX. We call V_1 the *west child* and V_2 the *east child*. The same process is now applied to each of the two children. Since the process is finite, each chain of descendents is finite.

Let T be a factorization tree of X.

Let $\mathfrak{T}$ be the set consisting of the final terms of all the chains in T. We say T *terminates in* $\mathfrak{T}$. Let $\mathfrak{X}$ be the set of (g,h)-nontrivial elements of $\mathfrak{T}$. If every break in T is a (g,h,i)-break with respect to the bonds which are X-edges or $\mathfrak{X}$-bonds, we say T is a (g,h,i)-*factorization tree* of X that *converts* VX into $\mathfrak{X}$. If, moreover, $\mathfrak{X}$ is the (g,h,i)-atomic factorization of X, we say that T is a (g,h,i)-*atomic factorization tree* of X. $\square$

We will now see that these latter objects exist.

2.5 Proposition. *If* $\mathfrak{X}$ *is a* (g,h,i)*-factorization of* X*, then there exists a* (g,h,i)*-factorization tree of* X *which converts* VX *into* $\mathfrak{X}$*.*

Proof. Let us alter X by adding in all $\mathfrak{X}$-bonds, so $\mathfrak{X}$ is now the (g,h,i)-atomic factorization of X.

We have the following trichotomy: X is (g,h)-trivial, or X is (g,h,i)-atomic, or X is (g,h,i)-decomposable.

If X is (g,h)-trivial, then $\mathfrak{X}$ is empty and the process stops.

If X is (g,h,i)-atomic, then $\mathfrak{X} = \{VX\}$ and the process stops.

If X is (g,h,i)-decomposable, then it has a proper (g,h,i)-break (V_1,V_2) and the process continues as follows.

We take V_1 and V_2 to be the two children of VX.

Here each element of $\mathfrak{X}$ is the vertex set of a (complete) (g,h,i)-atomic subgraph of X, so lies in exactly one of V_1, V_2. Hence we get a 2-partition of $\mathfrak{X}$, and denote the parts $\mathfrak{X}_1$, $\mathfrak{X}_2$.

For $i = 1, 2$, we apply the same trichotomy with the set $\mathfrak{X}_i$ and the full subgraph of X with vertex set V_i. Continuing in this way, we find further descendents and build a factorization tree T of X. Each chain in T eventually terminates in either an element of $\mathfrak{X}$ or a (g,h)-trivial set, so we have the desired (g,h,i)-factorization tree of X. $\square$

2.6 Corollary. *A* (g,h,i)*-factorization of a* (g,h,i)*-factorization of* X *is a* (g,h,i)*-factorization of* X*.* $\square$

2.7 Corollary. *Suppose $\mathfrak{X}$ is a (g,h,i)-factorization of X. Then the elements of $\mathfrak{X}$ are (g,h)-nontrivial, and each (g,h,i)-atomic factor of X lies in a unique element of $\mathfrak{X}$. If (g,h,i) is Type IV (resp. not Type IV), then the intersection of any pair of elements is $(g,2h)$-trivial (resp. (g,h)-trivial).*

Proof. Consider the proof of Proposition 2.5. For Type IV-V, we can separate the north cell of the first break from the two children by a sequence of *unipod deletions*, that is, $(1,0,0)$-breaks which create $(1,1)$-trivial terms which will be discarded. Thus each element of $\mathfrak{X}_1$ has (g,h)-trivial intersection with each element of $\mathfrak{X}_2$, except for Type IV, where it is $(g,2h)$-trivial. Now the result follows. □

§3. Shearing

We now introduce terminology and concepts that simplify the presentation of the proof.

Throughout this section let X be a finite bipartite graph, and let (g,h,i) be Type I-V.

3.1 Definitions. Let $\mathfrak{X}$ be a set of subsets of VX.

By an $\mathfrak{X}$-*edge* with respect to X, we mean an X-bond which is either an X-edge or a bond of at least two elements of $\mathfrak{X}$. Where X is clear from the context, we shall omit reference to it, and speak of $\mathfrak{X}$-edges.

We define $\mathfrak{X}$-edged(X) to be the subgraph of completion(X) obtained by adding all the $\mathfrak{X}$-edges to X (or VX).

For any subset Y of VX we define $\mathfrak{X}$-edged(Y) to be the full subgraph of $\mathfrak{X}$-edged(X) with vertex set Y.

We set $\mathfrak{X}^\bullet = \{\mathfrak{X}\text{-edged}(Y) \mid Y \in \mathfrak{X}\}$.

If X' is a (not necessarily full) subgraph of X, we define the restriction of $\mathfrak{X}$ to X', denoted $\mathfrak{X}|_{X'}$, to be the set $\{Y|_{X'} \mid Y \in \mathfrak{X}\}$, where $Y|_{X'}$ means $Y \cap X' = Y \cap VX'$. It is to be understood that $\mathfrak{X}|_{X'}$-edges are with respect to X' (as opposed to the full subgraph of X with the same vertex set).

Thus writing $\mathfrak{X}|_X$ denotes $\mathfrak{X}$, and emphasizes that $\mathfrak{X}$-edges are with respect to X. □

3.2 Definitions. By a *shearing sequence* (with respect to X) we mean a finite sequence

$$\mathfrak{X}_1, \dots, \mathfrak{X}_M$$

of sets of subsets of VX such that, for each $I \in \{1, \dots, M-1\}$, $\mathfrak{X}_{I+1}$ is obtained from $\mathfrak{X}_I$ by deleting a single yellow vertex y_I from a single element $X(I)$ of $\mathfrak{X}_I$ such that y_I has valence at most two in $\mathfrak{X}_I$-edged$(X(I))$. We say that $\mathfrak{X}_M$ is a *shearing* of $\mathfrak{X}_1$.

Let $\mathfrak{X}$, $\mathfrak{Z}$ be sets of subsets of VX. We say $\mathfrak{Z}$ is a (g, h, i)-*shearing* of $\mathfrak{X}$ if either (g, h, i) is Type IV-V and $\mathfrak{Z}$ is a shearing of $\mathfrak{X}$, or (g, h, i) is Type I-III and $\mathfrak{Z} = \mathfrak{X}$. □

¿From the proof of Proposition 2.5 and the above definition, we see that factorization is a process that must be carried out in parallel, while shearing is a process that must be carried out in series.

3.3 Remarks. Consider the shearing sequence in Definition 3.2.

Let $I \in \{1, \dots, M-1\}$. In passing from $\mathfrak{X}_I$-edged(X) to $\mathfrak{X}_{I+1}$-edged(X), we may lose as many as two X-bonds. In the step from $\mathfrak{X}_I^\bullet$ to $\mathfrak{X}_{I+1}^\bullet$, a bipod gets removed; if e is one of the bipod's bonds, and e is not an X-edge, and e is a bond of exactly one element $X'(I) \in \mathfrak{X}_I - \{X(I)\}$, then e gets deleted from $\mathfrak{X}_I$-edged(X), and from $\mathfrak{X}_I$-edged$(X'(I))$.

We can think of shearing as a process applied to $\mathfrak{X}^\bullet$ which successively removes a bipod and at most one other copy of each of the (at most two) bonds involved, and then only when the bond has only the one other occurrence, and is not an X-edge.

We shall be particularly interested in the case where $\mathfrak{X}$ is a (g, h, i)-factorization of X.

3.4 Proposition. *A (g,h,i)-shearing of a (g,h,i)-factorization of a (g,h,i)-shearing of a (g,h,i)-factorization of X is a (g,h,i)-shearing of a (g,h,i)-factorization of X.*

Proof. We may assume that (g, h, i) is Type IV-V. Let $\mathfrak{W}$ be a (g, h, i)-factorization of X, let $\mathfrak{Y}$ be a shearing of $\mathfrak{W}$, and let $\mathfrak{Z}$ be a (g, h, i)-factorization of $\mathfrak{Y}$. We want to show that $\mathfrak{Z}$ is a shearing of some (g, h, i)-factorization $\mathfrak{X}$ of X.

Let $\mathfrak{W}_1, \dots, \mathfrak{W}_M$ be a shearing sequence with $\mathfrak{W} = \mathfrak{W}_1$ and $\mathfrak{Y} = \mathfrak{W}_M$. We may assume we have consistent enumerations

$$\mathfrak{W}_I = (W_{I,J} \mid J = 1, \dots, N), \text{ for } I = 1, \dots M.$$

For $I = 1, \dots M-1$, denote by J_I the element of $\{1, \dots, N\}$ such that $\mathfrak{W}_{I+1}$ is obtained from $\mathfrak{W}_I$ by deleting a yellow vertex from W_{I,J_I}, and let y_I denote this yellow vertex.

By Proposition 2.5, there is a (g, h, i)-factorization tree T of X which converts VX into $\mathfrak{W}$. Also, for each $J \in \{1, \dots N\}$, there is a (g, h, i)-factorization tree T_J, with respect to the Y_J-bonds which are X-edges or $\mathfrak{Z}$-bonds, which converts $Y_J = W_{M,J}$ into a (possibly empty) subset of $\mathfrak{Z}$.

We shall now construct a sequence $\mathfrak{X}_1, \dots, \mathfrak{X}_M$, which, after we ignore repetitions, will be a shearing sequence, and shall also construct, for each

$I \in \{1, \dots, M-1\}$, families $(T_{I,J} \mid J = 1, \dots, N)$ of (g,h,i)-factorization trees, with respect to the bonds which are X-edges or $\mathfrak{X}_I$-bonds, that convert $\mathfrak{W}_I$ into $\mathfrak{X}_I$.

Let $\mathfrak{X}_M = \mathfrak{Z}$. For each $J \in \{1, \dots, N\}$, let $T_{M,J} = T_J$, so we have a family $(T_{M,J} \mid J = 1, \dots, N)$ of (g,h,i)-factorization trees, with respect to the bonds which are X-edges or $\mathfrak{X}_M$-bonds, that convert $\mathfrak{Y} = \mathfrak{W}_M$ into $\mathfrak{Z} = \mathfrak{X}_M$.

We continue recursively in reverse order. Suppose that $I \in \{1, \dots, M-1\}$ and that we have $\mathfrak{X}_{I+1}$ and $(T_{I+1,J} \mid J = 1, \dots, N)$. Recall that $\mathfrak{W}_{I+1}$ is obtained by deleting y_I from W_{I,J_I}, and y_I is incident to at most two $\mathfrak{W}_I$-edges in W_{I,J_I}. For all $J \in \{1, \dots N\} - \{J_I\}$, we set $T_{I,J} = T_{I+1,J}$. We have to change T_{I+1,J_I} to accommodate the extra yellow vertex y_I. There are two cases, and the first has two subcases.

Case 1. y_I is incident to exactly two $\mathfrak{W}_I$-edges e_1 and e_2 in W_{I,J_I}, and, moreover, each of e_1, e_2 is an $\mathfrak{X}_{I+1}$-bond or an X-edge. Let $x_i = r(e_i)$, $i = 1, 2$.

Case 1.1. x_1 and x_2 occur together in a (unique) term X_0 of $\mathfrak{X}_{I+1}$. This is the subcase where there is no break in T_{I+1,J_I} such that x_1 lies in the east cell and x_2 lies in the west cell. We add y_I to each term in the path in T_{I+1,J_I} from the base vertex W_{I+1,J_I} to X_0, so y_I is added to the east or west cell of each break. It is clear that the new breaks are still (g,h,i)-breaks. To pass from $\mathfrak{X}_I$ to $\mathfrak{X}_{I+1}$ we delete the vertex y_I of valence two from the term $X_0 \cup \{y_I\}$.

Case 1.2. There is a (unique) term X_0 in T_{I+1,J_I} where there is a break such that x_1 lies in the east cell and x_2 lies in the west cell. We add y_I to each term in the path in T_{I+1,J_I} from the base vertex W_{I+1,J_I} to X_0, so y_I is added to the east or west cell of each break up to this stage. At this stage, we add y_I to the north cell, and then delete the resulting unipods from the two children. Here $\mathfrak{X}_I = \mathfrak{X}_{I+1}$.

Case 2. y_I has valence at most one in W_{I,J_I}, with respect to the bonds which are $\mathfrak{X}_{I+1}$-bonds or X-edges. Here we delete the unipod directly. Here again $\mathfrak{X}_I = \mathfrak{X}_{I+1}$.

This completes the recursive definition. Set $\mathfrak{X} = \mathfrak{X}_1$. We have a family of (g,h,i)-factorization trees that convert $\mathfrak{W} = \mathfrak{W}_1$ into $\mathfrak{X}$, and combined with the (g,h,i)-factorization tree T that converts VX into $\mathfrak{W}$, this shows that $\mathfrak{X}$ is a (g,h,i)-factorization of X. We also have a shearing sequence from $\mathfrak{X}$ to $\mathfrak{Z}$. Thus $\mathfrak{Z}$ is a shearing of the (g,h,i)-factorization $\mathfrak{X}$, as desired. $\square$

Later we will need to apply the following result. The procedure described in the proof will be called *accommodating an interfering edge.*

3.5 Proposition. *Let e be an X-bond, and $X^+ = X \cup \{e\}$, the graph formed by adding e to X. Suppose that (g,h,i) is Type V, and let $(g^+, h^+, i^+) = (g,h,i) + (0,1,1)$. Then any (g,h,i)-factorization of X is a shearing of a (g^+, h^+, i^+)-factorization of X^+.*

Proof. Here $h \geq 2$, $(g,h,i) = (1,h,h)$ and $(g^+, h^+, i^+) = (1, h+1, h+1)$.

Let $x = r(e)$, $y = y(e)$.

Write $x_1 = x$, $e_1 = e$. Wherever we have a red vertex denoted x_N, we shall let e_N denote the bond joining y and x_N.

Suppose that $\mathfrak{X}$ is a (g,h,i)-factorization of X.

We will construct a (g^+, h^+, i^+)-factorization $\mathfrak{X}^+$ of X^+, and show that $\mathfrak{X}$ is a shearing of $\mathfrak{X}^+$.

By Proposition 2.5, there is a (g,h,i)-factorization tree T which converts VX into $\mathfrak{X}$. We propose to adjust some of the breaks by putting y in the south cell, which will give rise to (g^+, h^+, i^+)-breaks.

The first breaks to be adjusted are the ones where e_1 must be accommodated. Suppose then that we have a break where either y lies in the north cell, or one of x_1, y lies in the east cell and the other lies in the west cell.

Case 1. y is in the west cell. Here x_1 is in the east cell.

Case 1.1. There is a red vertex x_2 in the south cell, and the corresponding bond e_2 is an X-edge or an $\mathfrak{X}$-bond. We move y into the south cell. This adds a bipod $\{y, e_1, e_2\}$ to the east child. We proceed now as in the proof of Proposition 3.4, with the follwing subsubcases.

Case 1.1.1. x_1 and x_2 lie in some descendent of the east child. We add y to all terms in the corresponding path in T and the bipod gets added to a single term of $\mathfrak{X}$. This will be a bipod that gets sheared in the final stages of the shearing sequence.

Case 1.1.2. x_1 and x_2 separate at some break. We add y to the south cell and get unipods which can be deleted, so $\mathfrak{X}$ does not change.

Case 1.2. Either there is no red vertex in the south cell, or there is a red vertex but the corresponding bond is not an X-edge nor an $\mathfrak{X}$-bond. If we apply the same technique we get a unipod in place of a bipod, and it can be deleted, so $\mathfrak{X}$ does not change. (We think of Case 1.1 as a *generic* case, and Case 1.2 as a *degenerate* version of it, and in the future we will often leave degenerate versions to the reader, as the same techniques apply but with fewer complications.)

Case 2. y is in the north cell. Thus, in this break, y has valence at most 2 with respect to X-edges and $\mathfrak{X}$-bonds.

Case 2.1. x_1 is in the south cell.

Case 2.1.1 (generic). The valence of y is exactly two. Thus, there is a red vertex x_2 in the east cell and a red vertex x_3 in the west cell such that each of the corresponding bonds e_2, e_3 is an X-edge or an $\mathfrak{X}$-bond. We move y

into the south cell, and the $\{y, e_2\}$ unipod in the east child becomes a bipod $\{y, e_2, e_1\}$, and similarly for the west child. We proceed as in the previous case. Thus each bipod may get added to a unique term of $\mathfrak{X}$ in $\mathfrak{X}^+$. These bipods will be sheared in the final stages of the shearing sequence.

Case 2.1.2 (degenerate). The valence of y is zero or one.

Case 2.2. x_1 is in the east cell.

Case 2.2.1 (generic). The valence of y is exactly two, and there is a red vertex x_4 in the south cell. Thus y is incident to a red vertex x_2 in the east cell and a red vertex x_3 in the west cell. Here e_4 is not an X-edge nor an $\mathfrak{X}$-bond, but it may become an $\mathfrak{X}^+$-bond. We move y into the south cell, and the $\{y, e_2\}$ unipod in the east child becomes a tripod $\{y, e_2, e_1, e_4\}$, and the $\{y, e_3\}$ unipod in the west child becomes a bipod $\{y, e_3, e_4\}$. For the descendent of the west child we add the bipod as in the previous case, so this bipod may get added to a unique term of $\mathfrak{X}$. Adding the tripod to the descendents of the east child is more subtle. Here x_2, x_1, x_4 can separate in a subsequent break. If, say x_2 is in the south cell, then the tripod breaks into two bipods and we proceed as in the previous cases. If however, say x_2, x_1 are in the east cell, x_4 in the west cell, and a red vertex x_5 in the south cell, then we get a new bipod and a new tripod, and e_5 is not an X-edge and may become an $\mathfrak{X}^+$-bond. Thus we see that we may add one tripod and many bipods, each one to a different unique term of $\mathfrak{X}$; morever the tripod contains a bond which occurs in only one other term, in one of the added bipods. The final stage of our shearing sequence will remove the bipods first, and then the tripod turns into a bipod which will be sheared.

Case 2.2.2 (degenerate). The valence of y is zero or one, or there is no red vertex in the south cell.

The completes the description of adjustments made to a break to accommodate e_1.

Let $\mathfrak{X}_1$ denote the result of addding the bipods and tripods to $\mathfrak{X}$ arising from applying the preceding process to accommodate e_1 wherever necessary.

Let e_2 be an $\mathfrak{X}_1$-bond which is not an X^+-edge nor an $\mathfrak{X}$-bond. Thus e_2 is incident to y and to a red vertex in one of the terms of $\mathfrak{X}$ which had a bipod or tripod attached.

Now exactly as for e_1, we adjust breaks where necessary to accommodate e_2.

We continue this process until we arrive at a (g^+, h^+, i^+)-factorization tree T^+ which converts VX to a (g^+, h^+, i^+)-factorization $\mathfrak{X}^+$.

To show that $\mathfrak{X}$ is a shearing of $\mathfrak{X}^+$, we consider (yet another) finite simple- edged bipartite graph, Q, with red vertices the red X-vertices, and yellow vertices the elements of $\mathfrak{X}$, and edges indicating membership. The yellow X-vertices can temporarily be ignored.

A path s in Q has the form

$$V_1, z_1, V_2, z_2, \dots, z_{N-1}, V_N$$

where the z_I are red X-vertices, the V_I are elements of $\mathfrak{X}$, and $z_I \in V_I \cap V_{I+1}$. Here s is reduced if we have $V_I \neq V_{I+1}$ and $z_I \neq z_{I+1}$.

Each $\mathfrak{X}^+$-bond which is not an $\mathfrak{X}$-bond is incident to y and a red X-vertex which occurs in the component of x_1 in Q. Notice that the order in which terms of $\mathfrak{X}$ get given bipods and tripods corresponds to reduced paths in Q.

We claim that Q is a forest. Suppose not, so there exists a reduced path s as above with $N \geq 3$ and $V_N = V_1$. We view s as a cycle, with index set $\mathbb{Z}/(N-1)$. We may assume there are no repeated terms in s. Since $g = 1$, two distinct elements of $\mathfrak{X}$ have at most one red vertex in common. It follows that $N \geq 4$. Now consider the subtree of T determined by the $N-1$ different (endpoints of T corresponding to the) V_I. This subtree has $N-3$ branch points, so there exists some branch point. Consider a branch point W which is as far down T as possible, which means that exactly two of W's descendents are among the V_I. Call them V_I and V_J. We create a new graph Q' from Q by replacing all of W's descendents with W, which corresponds to cutting the descendents of W off T. There is a cycle s' in Q' corresponding to s. Here s' has two repeated terms, and need not be reduced if I and J are consecutive, but it does contains a subcycle which is reduced, and now we have reduced N. By induction, we may assume that Q' is a forest, so we have a contradiction. Thus Q is a forest.

This means that we can start the shearing sequence at the terminal vertices of the component of Q which contains x, and carry on shearing towards x. In this order, when a term comes to be sheared, y will be part of a bipod and can be sheared.

Thus $\mathfrak{X}$ is a shearing of $\mathfrak{X}^+$. $\square$

3.6 Remarks. (i). It is clear that, if e is an X-bond or an X-edge, then $\mathfrak{X}$ is already a (g, h, i)-factorization of X^+.

(ii). We will want to apply the above procedure also in the case where two edges e_1, e_2 incident to y are added to X. Here the procedure yields a (g^+, h^+, i^+)-factorization $\mathfrak{X}^+$ as before, and we can perform the same type of shearing sequence to get a set $\mathfrak{Z}$. Here there are three possibilities.

Case (a). $\mathfrak{Z} = \mathfrak{X}$.

Case (b). $\mathfrak{Z}$ is the same as $\mathfrak{X}$ except for one term, which differs in that y is added. In this term y is in a tripod $\{y, e_1, e_2, e_3\}$, where e_3 is an X-bond or an $\mathfrak{X}$-edge.

Case (c). $\mathfrak{Z}$ is the same as $\mathfrak{X}$ except for two terms, which both differ in that y is added, and in both terms y is in a tripod. These tripods have the form $\{y, e_1, e_3, e_4\}$, $\{y, e_2, e_3, e_5\}$, where each of e_4, e_5 is an X-bond or an $\mathfrak{X}$-edge, and e_3 is neither an X-bond nor an $\mathfrak{X}$-edge.

In Cases (b) and (c), the tripods are with reference to $\mathfrak{Z}|_{X^+}$-edges.

Case (b) arises from a (unique) break which has y in the east cell, x_1, x_2 in the west cell, and x_3 in the south cell, or y in the north cell, x_3 in the west cell, and x_1, x_2 in the union of the west and south cells, or the left-right duals. Here neither e_1 nor e_2 is an X-edge nor an $\mathfrak{X}$-bond.

Case (c) arises from a (unique) break which has y in the north cell, x_1, x_4 in the east cell, x_2, x_5 in the west cell, and x_3 in the south cell. Again, neither e_1 nor e_2 is an X-edge nor an $\mathfrak{X}$-bond. □

§4. Admissible factorizations

We devote this section to describing the general situation we plan to consider.

4.1 Definition. Let $\mathfrak{ABC}$ be a set of subsets of $VABC$. We will say that $\mathfrak{ABC}$ *is an admissible factorization* if we are given all the data and conditions described in this section.

First of all, there are given two triples of natural numbers (a, b, c), (g, h, i).

Moreover, (g, h, i) is Type I-V, and, if $m \leq 3$, then (g, h, i) is Type I.

There is given a full subgraph D_0 of D of size (a, b, c), and $a = m - g$. (Thus g is 0 or 1, and $a = m - g$, so D_0 has either all, or all but one, of the red vertices of D.)

The set $\mathfrak{ABC}$ is a (g, h, i)-shearing of a (g, h, i)-factorization of ABC.

There is a distinguished element ABC_0 of $\mathfrak{ABC}$, which is the union of three distinguished subsets A_0, B_0, C_0 whose pairwise intersection is VD_0, so

$$ABC_0 = A_0 \vee_{VD_0} B_0 \vee_{VD_0} C_0.$$

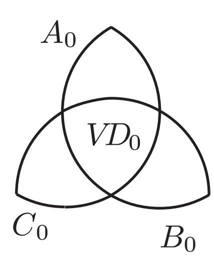

Fig. 5. The 4-partitioned set ABC_0.

The complement $\mathfrak{ABC} - \{ABC_0\}$ is 3-partitioned into subsets $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, so

$$\mathfrak{ABC} = \{ABC_0\} \vee \mathfrak{A} \vee \mathfrak{B} \vee \mathfrak{C}.$$

Moreover, A_0 and the elements of $\mathfrak{A}$ are subsets of VA, and similarly for B, C.

To continue the list of conditions, we need more vocabulary. □

4.2 Definitions. We call D_0 the *nucleus* (with respect to $\mathfrak{ABC}$).

Throughout this section, when we state a definition or a condition about A, B and C, all statements obtained by permuting A, B, and C (and $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ correspondingly) will also be understood. We emphasize this by using the phrase "permutations of ABC apply."

Recall, from Definition 3.1, that an $\mathfrak{ABC}$-edge is either an ABC-edge or a bond of at least two elements of $\mathfrak{ABC}$. Notice that no $\mathfrak{ABC}$-edge joins $A_0 - D$ to $B_0 - D$. Permutations of ABC apply.

We denote $(D - D_0) \cap A_0$ by $D(A_0)$. Permutations of ABC apply.

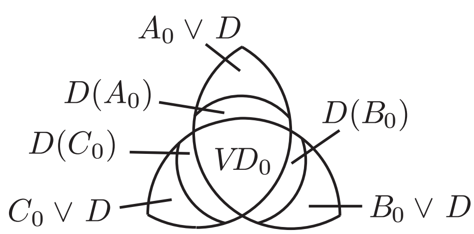

Fig. 6. The 7-partition of ABC_0.

Notice an element of $D(A_0)$ can be joined by an $\mathfrak{ABC}$-edge to any of the seven parts of ABC_0; see Fig. 7. Permutations of ABC apply.

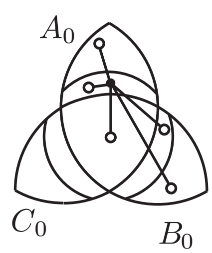

Fig. 7. Some $\mathfrak{ABC}$-edges incident to $D(A_0)$.

Let $BC_0 = B_0 \cup C_0$ $(= B_0 \vee_{VD_0} C_0)$, and let $\mathfrak{BC} = \{BC_0\} \vee \mathfrak{B} \vee \mathfrak{C}$, viewed as $\mathfrak{BC} = \mathfrak{BC}|_{BC}$. Thus, for example, a $\mathfrak{BC}$-edge is either a BC-edge or a bond of at least two terms of $\mathfrak{BC}$. Permutations of ABC apply. $\square$

4.3 Definitions. Let e be an $\mathfrak{ABC}$-edge.

We say that e is *D-like* (with respect to $\mathfrak{ABC}$) if either e is a D_0-bond or e is a D-edge which is not an ABC_0-edge.

We say e is *A-like* (with respect to $\mathfrak{ABC}$) if e is not a $\mathfrak{BC}$-bond nor a BC-edge. Permutations of ABC apply.

For example, if e has a vertex in $A - D$ then e is automatically A-like.

We say e is *exceptional* (with respect to $\mathfrak{ABC}$) if e is not A-, B-, C- or D-like.

Consider a term Y of $\mathfrak{ABC}$ which contains the vertices of e. We say that *e is A-like in Y* if e is not a BC-edge, and, for each term Z of $\mathfrak{ABC} - \{Y\}$ which contains the vertices of e, either $Z \in \mathfrak{A}$, or $Z = ABC_0$ and e has a vertex in $A_0 - D_0$. Informally, this can be thought of as saying that if Y is deleted from $\mathfrak{ABC}$ then e is not a $\mathfrak{BC}$-bond nor a BC-edge. When we depict $\mathfrak{ABC}$-edged(Y), if we wish to emphasize that e is A-like in Y, we will give e the label A. Permutations of ABC apply.

We say that e is *A-exceptional* if e is an ABC_0-bond which is incident to an element of $D(A_0)$ but is not A-like in ABC_0. Thus the A-exceptional bonds are the ABC_0-bonds incident to $D(A_0)$ which are BC-edges or $\mathfrak{B} \cup \mathfrak{C}$-bonds; see Fig. 8. Permutations of ABC apply.

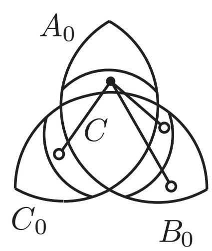

Fig. 8. Some A-exceptional edges.

We say that e is *D-exceptional* if e is not an ABC_0-bond, nor a D-edge, and is a bond of (at least) two of $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$. In particular, e must be a D-bond but not a D_0-bond.

One can show that e is exceptional if and only if it is A- or B- or C- or D-exceptional.

We say that e is *doubly* exceptional if if is A- and B-, or B- and C-, or C- and A-exceptional. In particular e must then be a D-bond but not a D_0-bond. A doubly exceptional bond will always be counted as two exceptional bonds. □

4.4 Remark. We will eventually require that $\mathfrak{BC}$ be a (g, h, i)-shearing of a (g, h, i)-factorization of BC. This will be a consequence of the other conditions, except for Type III where it will be explicitly hypothesized. Clearly $\mathfrak{ABC}|_{BC}$ is a (g, h, i)-shearing of a (g, h, i)-factorization of BC. Notice that each element of $\mathfrak{A}|_{BC}$ lies in D. In each case, we shall successively discard those elements of $\mathfrak{A}|_{BC}$ which can be broken up. When all of $\mathfrak{A}|_{BC}$ has been discarded, we are able to break $D(A_0)$ off $ABC_0|_{BC}$, since there are relatively few $(\mathfrak{ABC} - \mathfrak{A})|_{BC}$-edges involved, and discard it. □

We now consider the five different types separately.

4.5 Conditions for Type I. Here, $(g, h, i) = (0, h, 0)$, $h \geq 0$. We require that $(a, b, c) = (m, n-h, p)$, and that there be no exceptional bonds. □

4.6 Remarks about Type I. We have $(a, b, c) + (g, h, i) = (m, n, p)$.

Since D_0 is a subgraph of D of size $(m, n-h, p)$, D_0 is obtained from D by removing h nullipods, that is yellow vertices of valence zero, so D_0 contains all the red vertices of D.

By Hypotheses 1.6, every pair of red vertices of D_0 has total valence at least three.

Since $g = 0$, the only bonds of interest are ABC-edges, and (g, h, i)-shearing changes nothing.

For any $y \in D(B_0)$, y is a yellow vertex, and every edge incident to y in

$$\mathfrak{ABC}\text{-edged}(ABC_0)$$

is B-like, which means it is not a CA-edge and not a $\mathfrak{C} \cup \mathfrak{A}$-bond; we say that y has $\mathfrak{ABC}$-*valence at most* $B + 0$ *in* ABC_0. Permutations of ABC apply.

Here $\mathfrak{BC}$ is a (g, h, i)-factorization of BC for the following reason. If $h = 0$, then $D_0 = D$ and both $\mathfrak{A}|_{BC}$ and $D(A_0)$ are empty, so $\mathfrak{BC} = \mathfrak{ABC}|_{BC}$, and this is a (g, h, i)-factorization of BC. If $h \geq 1$, then $\mathfrak{A}|_{BC}$ is $(0, \infty)$-trivial, so can be broken up into (g, h)-trivial subsets, which can deleted. On restriction to BC, $D(A_0)$ consists of nullipods, which can be broken off and deleted. Thus $\mathfrak{BC}$ is a (g, h, i)-factorization of BC. Permutations of ABC apply. □

4.7 Definitions. For Type II-V, $g = 1$ and, by hypothesis we will have $m \geq 4$. By Hypotheses 1.6 and Lemma 1.5, $n \geq 4$ and $p \geq (m-2)(n-2)+6$.

Notice $a = m - 1 \geq 3$. The unique red vertex which lies in D but not in D_0 will be denoted x, throughout.

If x lies in ABC_0, we say we are in the *inner case*, and if x does not lie in ABC_0, then we say we are in the *outer case*. This subdivision does not apply for Type I. In the inner case x lies in $D(A_0)$ or $D(B_0)$ or $D(C_0)$. By interchanging A and B if necessary, we may assume that x does not lie in $D(B_0)$.

4.8 Conditions for Type II. Here, $(g, h, i) = (1, 0, 0)$. We require that $(a, b, c) = (m - 1, n, p)$, and that there be no exceptional bonds.

4.9 Remarks about Type II. We have $(a, b, c) + (g, h, i) = (m, n, p)$.

Since D_0 is a subgraph of D of size $(m-1, n, p)$, it is obtained by removing x, and x has valence zero in D.

By Hypotheses 1.6, in D_0, every red vertex has valence at least three.

Since $h = 0$, the only bonds of interest are ABC-edges, and (g, h, i)-shearing changes nothing.

If $x \in D(C_0)$ then x has $\mathfrak{A}\mathfrak{B}\mathfrak{C}$-valence at most $C + 0$ in ABC_0. Permutations of ABC apply.

As for Type I, $\mathfrak{B}\mathfrak{C}$ is a (g, h, i)-factorization of BC. Permutations of ABC apply. □

4.10 Conditions for Type III. Here, $(g, h, i) = (1, 1, 1)$. We require $b = n - 1$, $c = p$ or $p - 1$, and that there be at most one exceptional bond. If $c = p-1$ and there is an exceptional bond, we require that it be a D-edge. We also require that, in the outer case, $\mathfrak{A}\mathfrak{B}$ is a (g, h, i)-factorization of AB; permutations of ABC apply. □

4.11 Remarks about Type III. We have

$$(a, b, c) + (g, h, i) \in \{(m, n, p), (m, n, p + 1)\}.$$

Here D_0 is a subgraph of D of size $(m-1, n-1, p)$ or $(m-1, n-1, p-1)$, so is obtained by removing the red vertex, x, one yellow vertex, denoted y, and zero or one edges, so the total valence of x and y in D is at most one.

There is no doubly exceptional bond, since they count as two exceptional bonds, and this case does not allow two.

By Hypotheses 1.6, in D_0, every red vertex has valence at least two.

In the inner case, $\mathfrak{B}\mathfrak{C}$ is a (g, h, i)-factorization of BC for the following reason. The elements of $\mathfrak{A}|_{BC}$ are $(1, \infty)$-trivial, so can be broken up into (g, h)-trivial subsets, and deleted; recall that $(g, h) = (1, 1)$. Also, on restriction to BC, $D(A_0)$ consists of unipods, which can be removed and deleted. Permutations of ABC apply.

In the outer case, it is clear that the elements of $\mathfrak{A}|_{BC}$ are $(2,2)$-trivial, but we will have to ensure they have less than four $\mathfrak{ABC}|_{BC}$-edges in order to be able to break them up into (g,h)-trivial sets. □

4.12 Conditions for Type IV. Here $(g,h,i) = (1,1,0)$. We require that $(a,b,c) = (m-1, n-1, p)$, and that there be no exceptional bonds. □

4.13 Remarks about Type IV. Here $(a,b,c) + (g,h,i) = (m,n,p)$.

Since D_0 is a subgraph of D of size $(m-1, n-1, p)$, it is obtained by removing the red vertex, x, one yellow vertex, denoted y, and no edges, so the total valence of x and y in D is zero.

By Hypotheses 1.6, in D_0, every red vertex has valence at least three.

Here $\mathfrak{BC}$ is a shearing of a (g,h,i)-factorization of BC by Proposition 3.4 and the following argument. The elements of $\mathfrak{A}|_{BC}$ are $(2,\infty)$-trivial, so can be broken up into $(1,1)$-trivial sets, and discarded. Also, on restriction to BC, $D(A_0)$ consists of (at most one) yellow and (at most one) red vertices of valence zero, which can be broken off and discarded. Permutations of ABC apply. □

4.14 Definition. Let e be an ABC-bond.

We define the *weight* of e (with respect to $\mathfrak{ABC}$), $\mathrm{wt}(e)$, which is a value in $\{0, \frac{1}{2}, 1, 2\}$, as follows.

If e is not an $\mathfrak{ABC}$-edge we set $\mathrm{wt}(e) = 0$. Now we may assume that e is an $\mathfrak{ABC}$-edge.

If e is not exceptional we set $\mathrm{wt}(e) = 0$. Now we may assume that e is exceptional.

If e is doubly exceptional and is not a D-edge then $\mathrm{wt}(e) = 2$. If e is doubly exceptional and is a D-edge then $\mathrm{wt}(e) = 1$. Thus we may assume further that e is not doubly exceptional.

If e is a D-edge we set $\mathrm{wt}(e) = 0$. Now we may assume that e is an exceptional $\mathfrak{ABC}$-edge, but not doubly exceptional, and not a D-edge.

Sometimes we will have a yellow vertex y of $D - D_0$ with a special configuration associated to it as follows.

There is a specified term Y of $\mathfrak{ABC}$ in which y occurs, and either Y lies in $\mathfrak{C}$ or $Y = ABC_0$ and y lies in $D(C_0)$, and, in each case, y is incident to exactly three edges e_1, e_2 and e_3 in $\mathfrak{ABC}$-edged(Y), and e_1, e_2 are exceptional and A-like in Y, and e_3 is C-like in Y. Here e_1 and e_2 are both given weight $\frac{1}{2}$. In $\mathfrak{ABC}$-edged(Y), y is incident to two A-like edges and one C-like edge, and we think of the whole tripod as having weight 1, and call it an *A-tripod in the C-part*. Permutations of ABC apply.

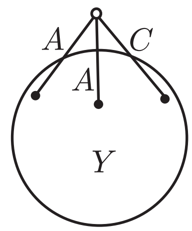

Fig. 9. An A-tripod in the C-part.

There are specified two distinct terms Y_1, Y_2, of $\mathfrak{ABC}$ in which y occurs, and for each $i \in \{1, 2\}$, either Y_i lies in $\mathfrak{C}$ or $Y_i = ABC_0$ and y lies in $D(C_0)$, and, in each case, y is incident to exactly three edges e_{1i}, e_{2i} and e_{3i} in $\mathfrak{ABC}$-edged(Y_i), and e_{1i} is exceptional and A-like in Y_i. Further $e_{31} = e_{32}$ as ABC-bonds, and this bond is a bond of no other term of $\mathfrak{ABC}$. Here e_{11} and e_{12} are both given weight $\frac{1}{2}$. We think of the pair of tripods as having weight one and call it an *A-tripod pair in the C-part.* Permutations of ABC apply.

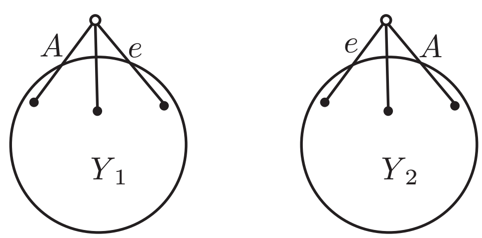

Fig. 10. An A-tripod pair.

In all remaining cases, $\mathrm{wt}(e) = 1$.

The sum of the weights of the ABC-bonds is called the *weight sum* (with respect to $\mathfrak{ABC}$) and denoted d; notice that d is a natural number since the bonds of weight $\frac{1}{2}$ occur in pairs. □

4.15 Conditions for Type V. Here $(g, h, i) = (1, h, h)$, $h \geq 2$. We require $b \leq n - 2$, and $c + 2(n - b) \geq p + h + d$.

If $x \in D_0(C)$, we require that x have $\mathfrak{ABC}$-valence at most $C{+}h$ in ABC_0, that is, x is incident to at most h edges of $\mathfrak{ABC}$-edged(ABC_0) which are AB-edges or $\mathfrak{A} \cup \mathfrak{B}$-bonds (so are not C-like). Permutations of ABC apply.

For each yellow vertex y in $D - D_0$, we require that at most 2 of the ABC-bonds incident to y be exceptional $\mathfrak{ABC}$-edges.

We require that in the case where some $y \in D_0(A)$ is incident to two exceptional $\mathfrak{ABC}$-edges, and one of them, e, is incident to $x \in D_0(C)$, then e is C-like. Permutations of ABC apply. □

4.16 Remarks about Type V. Here D_0 is obtained from D by removing the red vertex x, and $n-b\,(\geq 2)$ yellow vertices, and $p-c\,(\leq 2n-2b-h-d)$ edges.

Here $\mathfrak{BC}$ is a shearing of a (g,h,i)-factorization of BC by Proposition 3.4 and the following argument. The elements of $\mathfrak{A}|_{BC}$ are $(2,\infty)$-trivial, so can be broken up into $(1,1)$-trivial sets, and deleted. Also, on restriction to BC, $D(A_0)$ consists of bipods, and at most one red vertex of valence at most h. The red vertex can be broken off and discarded, and the yellow vertices can be sheared. Permutations of ABC apply.

For each yellow vertex y in $D - D_0$, if y lies in a term Y of $\mathfrak{B}$ then y has valence at most $B+2$ in Y. Also, if $y \in D_0(B)$, then y has valence at most $B+2$ in ABC_0. Permutations of ABC apply.

For each yellow vertex y in $ABC - D$, the bond joining y to x is the only ABC-bond incident to y that can be an exceptional $\mathfrak{ABC}$-edge. □

§5. The structure of the proof of Theorem 1.3

5.1 The first step. We have Hypotheses 1.4 and 1.6.

If we take $(g,h,i)=(0,0,0)$, $(a,b,c)=(m,n,p)$, $D_0 = D$, $A_0 = VA$, $B_0 = VB$, $C_0 = VC$ and $\mathfrak{ABC} = \{VABC\}$, then we have an admissible factorization. □

5.2 The inductive hypothesis. We have an admissible factorization, with the notation of §4 applying. □

5.3 The inductive procedure. We devote the remainder of the article to showing that there exists an admissible factorization $\mathfrak{ABC}'$, with the same notation as before but with primes added, such that $\mathfrak{ABC}'$ is a proper refinement of $\mathfrak{ABC}$. Here *refinement* means that $\mathfrak{ABC}'$ is given by replacing each element of $\mathfrak{ABC}$ with a (possibly empty) set of subsets of itself, and *proper* means that at least one term of $\mathfrak{ABC}$ gets replaced with a (possibly empty) set of proper subsets of itself.

Since $VABC$ is finite, the König tree lemma implies that this cannot continue indefinitely. Hence we will have a contradiction. □

5.4 Notation. We know that $\mathfrak{AB} \vee \mathfrak{BC} \vee \mathfrak{CA}$ is a (g,h,i)-shearing of a (g,h,i)-factorization of $AB \vee BC \vee CA$.

Each (g,h,i)-atomic factor of $AB \vee BC \vee CA$ has a unique (g,h,i)-shearing which lies in a (unique) term of $\mathfrak{AB} \vee \mathfrak{BC} \vee \mathfrak{CA}$.

By Hypotheses 1.4, $AB \vee BC \vee CA$ has an even number of (g,h,i)-atomic factors. But $\mathfrak{AB}$ has $1 + |\mathfrak{A}| + |\mathfrak{B}|$ terms, and permutations of ABC apply. Hence $\mathfrak{AB} \vee \mathfrak{BC} \vee \mathfrak{CA}$ has $3 + 2|\mathfrak{A}| + 2|\mathfrak{B}| + 2|\mathfrak{C}|$ terms, which is an odd number. Thus, for some term Y of $\mathfrak{AB} \vee \mathfrak{BC} \vee \mathfrak{CA}$, the number of (g,h,i)-sheared (g,h,i)-atomic factors of $AB \vee BC \vee CA$ lying in Y is even. It is straightforward to see these must therefore be proper subsets of Y.

By symmetry we may assume that Y is a term of $\mathfrak{AB}$. It then uniquely determines an element of $\mathfrak{ABC}$ denoted YC; that is, if $Y \in \mathfrak{A} \cup \mathfrak{B}$ then $YC = Y$, and if $Y = AB_0$ then $YC = ABC_0$. Permutations of ABC do not apply here.

Let $\mathfrak{T}$ be the (g,h,i)-atomic factorization of AB, and T be a (g,h,i)-atomic factorization tree of AB, as in Proposition 2.5. By Proposition 3.4, restricting $\mathfrak{T}$ to each term of $\mathfrak{AB}$, and omitting all resulting (g,h)-trivial sets, gives a (g,h,i)-shearing $\mathfrak{AB}^\dagger$ of a (g,h,i)-atomic factorization of AB. The distinguished term Y gets replaced with proper subsets of itself, so $\mathfrak{AB}^\dagger$ is a proper refinement of $\mathfrak{AB}$. The 3-partition $\mathfrak{AB} = \{AB_0\} \vee \mathfrak{A} \vee \mathfrak{B}$ determines a corresponding 3-partition $\mathfrak{AB}^\dagger = \mathfrak{AB}_0^\dagger \vee \mathfrak{A}^\dagger \vee \mathfrak{B}^\dagger$.

Notice that for Type III-V, the proper subsets of Y cannot contain all the red vertices of Y; for, if one contains all the red vertices, the others contain at most one red vertex so can be broken up and discarded, so the even number involved must be zero, so Y has no red vertices, so is not AB_0, so is in $\mathfrak{A}$ or $\mathfrak{B}$ and can be discarded. $\square$

All of this notation will be maintained throughout.

We now have a basis for constructing an admissible factorization in which YC is replaced with a (possibly empty) set of proper subsets of itself. In the next section we will examine how T breaks up the nucleus, and in the subsequent sections we will consider various cases separately and show how T can be used to construct a new admissible factorization in each case.

§6. Breaking up the nucleus

We begin by recording two general facts which do not refer specifically to the notation of the previous sections.

6.1 Lemma. *Suppose* a, b, g, h, a', b', g', h' *are real numbers such that* $a+g = a'+g'$, $b+h = b'+h'$. *If* $ab+gh > a'b'+g'h'$, *then*

$$(a-g')(h-h') + (g-g')(b-h') < 0.$$

Proof. By hypothesis $(a+g-g')(b+h-h') = a'b'$ and

$$a'b' + g'h' - (ab+gh) < 0,$$

and these combine to say that $(a+g-g')(b+h-h') + g'h' - (ab+gh) < 0$. We can expand this expression to

$$(a-g')(h-h') + g(h-h') + (a-g')b + gb + g'h' - ab - gh,$$

and then to $(a-g')(h-h') - gh' - g'b + gb + g'h'$, which is

$$(a-g')(h-h') + (g-g')(b-h'). \quad \square$$

We note the special case where $(a',b') = (g',h') = (\frac{a+g}{2}, \frac{b+h}{2})$, although this case follows directly from the identity

$$ab + gh = \frac{1}{2}(a+g)(b+h) + \frac{1}{2}(a-g)(b-h).$$

6.2 Lemma. *If* a, b, g, h *are real numbers such that*

$$ab + gh > \frac{1}{2}(a+g)(b+h),$$

then $(a-g)(b-h) > 0$. $\square$

We now return to the notation of the previous sections.

6.3 Lemma. (i) *For Type I-V,* $a+g = m$, $a-g \geq 2$, $b-h \geq 1$, $1 \geq g$, *and* $c+i > \frac{1}{2}(a+g)(b+h)$.

(ii) *For Type I-IV,* $b+h = n$, $c+i \geq p$.

(iii) *For Type V,* $c+i > (a+g-2)b + 2h$.

Proof. If (g,h,i) is Type I-IV, then, g is 0 or 1, and, by §4, $a+g = m$, $b+h = n$, $c+i \geq p$, $a \geq 2+g$, and $b \geq 2+h$.

This leaves Type V. Here $a = m-1 \geq 3 = 2+g$, $b \leq n-2$, and

$$c + 2(n-b) \geq p + h > (m-2)(n-2) + 4 + h,$$

by Lemma 1.5. Thus,

$$c + 2b - h - mb > (m-2)(n-2) + 4 + 4b - 2n - mb = (m-4)(n-2-b) \geq 0,$$

so $c > mb + h - 2b$. Hence $c + i = c + h > mb + 2h - 2b = (m-2)b + 2h$. Also, $(m-1)b = ab \geq c > mb + h - 2b = (m-1)b + (h-b)$, so $b > h$. Finally,

$$c + i > (m-2)b + 2h = \frac{1}{2}m(b+h) + \frac{1}{2}(m-4)(b-h) \geq \frac{1}{2}m(b+h).$$

Since $m = a+g$, we have proved (i) and (iii). $\square$

6.4 The beginning of a recursive definition. Consider the (g,h,i)-atomic factorization tree T of AB. We orient each break in T so that the west cell has at least as many red vertices of D_0 as the east cell. (Recall that the south cell has at most $g \le 1$ red vertices, and the north cell has no red vertices.) Let $AB(0) = AB_0, AB(1), \ldots, AB(q)$ denote the descending chain of subsets of AB_0 obtained by taking the westmost chain in this tree, and let $(AB(r+1), AB'(r+1))$ denote the (g,h,i)-break of $AB(r)$. Thus we have a distinguished chain in T, called the *main line*, which follows the majority of remaining red vertices of D_0 at each stage, and if there is a tie, we choose one side arbitrarily; we will see that there is never a tie.

We now construct subgraphs $D(0), \ldots, D(q)$ of D_0 recursively.

We take $D(0) = D_0$.

Let $(a_0, b_0, c_0) = (a, b, c)$, $(g_0, h_0, i_0) = (g, h, i)$. Let $E(0)$ be any finite simple-edged bipartite graph of size (g_0, h_0, i_0); such an $E(0)$ exists since $i \le gh$. Here $a_0 - g_0 \ge 2$ and $1 \ge g_0$, by Lemma 6.3(i).

Now suppose that $0 \le r \le q-1$, and that we have defined a full subgraph $D(r)$ of D_0 with vertex set in $AB(r)$, and a finite simple-edged bipartite graph $E(r)$ containing $E(0)$, such that, setting

$$(a_r, b_r, c_r) = \text{size}(D(r)), \quad (g_r, h_r, i_r) = \text{size}(E(r)),$$

we have

$$(a_r, b_r, c_r) + (g_r, h_r, i_r) = (a, b, c) + (g, h, i), \tag{1}$$

$$a_r - g_r \ge 2, \tag{2}$$

$$1 \ge g_r, \tag{3}$$

(4) if $(g,h,i) = (1,1,0)$ and $(a_r, b_r, c_r) \ne (a,b,c)$ then $(g_r, h_r) \ge (1,2)$.

The elements of D which lie in one $D(r)$ but not the next, are said to *escape from the nucleus.*

Now $(AB(r+1), AB'(r+1))$ is a (g,h,i)-break of $AB(r)$, and we can restrict it to $D(r)$. If necessary we re-orient the break so that the west cell contains at least as many red vertices of $D(r)$ as the east cell; we shall see that it is not necessary to re-orient.

We define $\epsilon = 1$ if $(g_r, h_r, i_r) = (g,h,i) = (1,1,0)$ (so $(a_r, b_r, c_r) = (a,b,c)$, by (1)) and there is a crossover vertex in $D(r)$; otherwise we define $\epsilon = 0$. Notice that

$$i_r \le g_r h_r - \epsilon. \tag{5}$$

For Type IV, when $\epsilon = 1$, so there there is a crossover vertex y which lies in $D(r)$, and $(g_r, h_r, i_r) = (1,1,0)$, we alter the break of $D(r)$ by moving

y into the east (resp. west) cell if the break of $AB(r)$ is western (resp. eastern). Thus, if $\epsilon = 1$ we can have an edge joining the east and west cells of $D(r)$. In this event the edge will notionally get moved to $E(r)$.

We extend the 4-partition to the edge set of $D(r)$ as follows. Edges incident to a northern vertex lie in the north cell. Edges incident to an eastern vertex and not to any northern or western vertices lie in the east cell. Similarly for the west cell. Edges incident to two southern vertices lie in the south cell. Now each cell consists of red vertices, yellow vertices, and edges, so we can define its size. Let (j_s, k_s, l_s) (resp. (j_e, k_e, l_e), (j_n, k_n, l_n) , (j_w, k_w, l_w)) denote the sizes of the resulting south (resp. east, north, west) cells. (The subscript n for 'north' should not be confused with the n given in Hypothesis 1.4.) See Fig. 11.

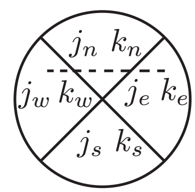

Fig. 11. The break of $D(r)$.

By our choice of (re)orientation,

$$j_w \geq j_e. \tag{6}$$

By the definition of a (g, h, i)-break, $j_n = 0$ and each of the k_n yellow vertices is joined by an edge to at most one (red) vertex in the west cell, and at most one (red) vertex in the east cell, and to no vertices in the south cell. There is a crossover edge in the north cell if and only if $\epsilon = 1$. Also, it follows from (4) and the definition of a break that

$$(j_s, k_s, l_s) \leq (g_r, h_r, g_r h_r). \tag{7}$$

We define $D(r+1)$ to be the graph consisting of the union of the (new) south and west cells of $D(r)$. Let $(a_{r+1}, b_{r+1}, c_{r+1}) = \text{size}(D(r+1))$.

We form a graph $E(r+1)$ by attaching the north and east cells of $D(r)$ to $E(r)$ in such a way that simple-edgedness is maintained; this can be done as follows. If there is a crossover edge, then $(g_r, h_r, i_r) = (1, 1, 0)$

so we can attach the crossover edge to $E(r)$. Now it remains to attach the rest of the east and north cells, which have detached edges previously attached to the south and west cells, respectively. Using (7), we see that each red eastern vertex is incident to at most $k_s \leq h_r$ detached edges, and each yellow eastern vertex is incident to at most $j_s \leq g_r$ detached edges, and each northern vertex is yellow and is incident to at most $g \leq g_r \leq 1$ detached edges. Thus the attachments can be performed in a simple-edged manner. Let $(g_{r+1}, h_{r+1}, h_{r+1})$ denote the size of $E(r+1)$.

Notice that

$$\begin{aligned}(a_{r+1}, b_{r+1}, c_{r+1}) + (g_{r+1}, h_{r+1}, i_{r+1}) &= (a_r, b_r, c_r) + (g_r, h_r, i_r) \\ &= (a, b, c) + (g, h, i) \text{ by (1).}\end{aligned} \tag{8}$$

Moreover,

$$a_{r+1}b_{r+1} + g_{r+1}h_{r+1} \geq c_{r+1} + i_{r+1} > \frac{1}{2}(a_{r+1} + g_{r+1})(b_{r+1} + h_{r+1}),$$

by (8) and Lemma 6.3(i). By Lemma 6.2,

$$(a_{r+1} - g_{r+1})(b_{r+1} - h_{r+1}) > 0. \tag{9}$$

Notice that $a_{r+1} + g_{r+1} = a + g = m$ is even, by (8), Lemma 6.3(i), and Lemma 1.5(i).

If $a_{r+1} < g_{r+1}$, then, by (6), we must have fewer than g red vertices in the south cell, and it follows that $g = 1$ and $a_{r+1} = g_{r+1} - 1$. This contradicts the fact that $a_{r+1} + g_{r+1}$ is even. Thus $a_{r+1} \geq g_{r+1}$. Hence, by (9),

$$a_{r+1} > g_{r+1}, \text{ and } b_{r+1} > h_{r+1}. \tag{10}$$

Since $a_{r+1} + g_{r+1} = m$ is even, we see that $a_{r+1} - g_{r+1} \geq 2$, and that if $m = 2$ then $g_{r+1} = 0$. □

6.5 Lemma. *$g_{r+1} \leq 1$.*

Proof. Suppose not, so

$$g_{r+1} \geq 2. \tag{11}$$

We saw that $m \geq 4$ so, by Hypotheses 1.6, $n \geq 4$, and, by Lemma 1.5(iv), $p > (m-2)(n-2) + 4$.

Case 1. Type I-IV.

By Lemma 6.3(ii), $c + i > (a + g - 2)(b + h - 2) + 4$, so, by (8),

$$\begin{aligned} a_{r+1}b_{r+1} + g_{r+1}h_{r+1} &\geq c_{r+1} + i_{r+1} \\ &> (a_{r+1} + g_{r+1} - 2)(b_{r+1} + h_{r+1} - 2) + 4. \end{aligned}$$

By Lemma 6.1, $(a_{r+1}-2)(h_{r+1}-2)+(g_{r+1}-2)(b_{r+1}-2) < 0$, and, by (10) and (11), we see that $h_{r+1} \geq 2$ is impossible, so $h_{r+1} \leq 1$. Thus $E(r + 1)$ has at least two red vertices and at most one yellow vertex.

For Type I, when a red vertex escapes, it takes with it all edges of D incident to it, so each of the (at least two) red vertices in $D_0 - D(r+1)$ has valence at most one in D_0. Thus D_0 has two red vertices of total valence at most two, which contradicts §4.6.

For Type II, when a yellow vertex escapes, it takes with it all edges of D incident to it, and when a red vertex escapes, it takes all the remaining edges of D incident to it. Hence, each of the (at least one) red vertices in $D_0 - D(r + 1)$ has valence at most one in D_0. Thus D_0 has a red vertex of valence at most one, which contradicts §4.9.

For Type III, the north cell is empty and the south cell has size at most $(1, 1, 1)$. Here $D(r) = D_0$ and no yellow vertex lies in the east cell, and at least one red vertex lies in the east cell, so this red vertex has valence at most one in D_0. This contradicts §4.11.

For Type IV, since $h_r \leq h_{r+1} \leq 1$, we see by (4) that $(g_r, h_r, i_r) = (1, 1, 0)$, and $D(r) = D_0$. There are no yellow vertices in the north and east cells, and there is at least one red vertex in the east cell which must have valence at most two. This contradicts §4.13.

Case 2. Type V.

$$\begin{aligned} a_{r+1}b_{r+1} + g_{r+1}h_{r+1} \geq c_{r+1} + i_{r+1} &= c + i \text{ by (8)} \\ &> (a + g - 2)b + 2h \text{ by Lemma 6.3(iii)}. \end{aligned}$$

By (8) again and Lemma 6.1,

$$0 > (a_{r+1} - 2)(h_{r+1} - h) + (g_{r+1} - 2)(b_{r+1} - h),$$

but all four of these differences are nonnegative, by (10), (11), and the fact that $E(r + 1)$ contains $E(0)$.

Thus in all events we have a contradiction, so (11) is false. □

6.6 The conclusion of the recursive definition. In the following we refer to the breaks of D_0.

Notice that either $g = 0$ and the north cell is necessarily empty, or $g = 1$ and, by the foregoing result, the east cell has no red vertex. Hence, the vertices in the north cell are not attached to the east cell, so have valence at most one. If there is a crossover edge and it lies in D_0, then the break is eastern and the crossover vertex has valence at most two in D_0.

We have seen that $g_{r+1} \leq 1$, and $a_{r+1} > g_{r+1}$, and, if

$$(g_{r+1}, h_{r+1}, i_{r+1}) \neq (g_r, h_r, i_r) = (g, h, i) = (1, 1, 0),$$

then the east cell contains a yellow vertex, which gets transferred to $E(r{+}1)$, so $(g_{r+1}, h_{r+1}) \geq (1, 2)$. This completes the recursive definition of the $D(r)$ and $E(r)$, for $1 \leq r \leq q$.

We set $D'_0 = D(q)$ and $(a', b', c') = (a_q, b_q, c_q)$, the size of D'_0. We will take D'_0 to be the new nucleus. $\square$

To summarize the results obtained in this section, we observe that we have seen that the (g, h, i)-atomic factorization tree T of AB has a main line, which determines a descending chain of subgraphs

$$D(0) = D_0, D(1), \ldots, D(q) = D'_0,$$

and D'_0 will be the new nucleus.

If $m < 4$, then only nullipods escape. If $m \geq 4$, then $n \geq 4$, and nullipods, unipods, and perhaps one bipod escape, as well as at most $1 - g$ (which is one or zero) red vertices. All the escaping vertices and edges can be attached to a finite simple-edged bipartite graph of size (g, h, i) to form a new finite simple-edged bipartite graph. For Type I-III, the north cells of the breaks are empty, and each $D(r)$ consists of the south and west cells of the previous break. For Type IV, the north cells contain vertices of valence at most one, and each $D(r)$ consists of the south and west cells of the previous break, but omitting any crossover vertex at the first proper break. For Type V, again the north cells contain vertices of valence at most one, and each $D(r)$ consists of the south and west cells of the previous break. For Type II-III and V, only unipods escape. For Type IV, unipods and up to one bipod escape.

§7 A first approximation to a new admissible factorization

We shall use the notation of the previous sections, particularly §5.4, and §6.

7.1 Definitions. From §6, we have the new nucleus D'_0, of size (a', b', c').

¿From §5.4, we have $\mathfrak{AB}^\dagger = \mathfrak{AB}_0^\dagger \vee \mathfrak{A}^\dagger \vee \mathfrak{B}^\dagger$, a (g,h,i)-factorization of $\mathfrak{AB}$, obtained by applying a (g,h,i)-atomic factorization of AB to each term of $\mathfrak{AB}$.

We know that $\mathfrak{AB}^\dagger$ is a (g,h,i)-shearing of a (g,h,i)-factorization of AB which properly refines $\mathfrak{AB}$.

Throughout the remainder of the article, let us tacitly assume that, if we are dealing with (g,h,i)-breaks, then (g,h)-trivial elements are to be automatically deleted from any set of subsets of $VABC$ we define, and that corresponding conventions apply as we increase the break size.

Recall that $\mathfrak{AB}_0^\dagger$ is the set of descendents of AB_0, and in §6 we found a distinguished term $AB(q)$. Let AB_0^+ denote $AB(q)$ with the, at most one, and then only for Type IV, escaping bipod sheared off. Set $A_0^+ = AB_0^+ \cap A_0$ and $B_0^+ = AB_0^+ \cap B_0$. It can be seen that $AB_0^+ = A_0^+ \vee_{VD'_0} B_0^+$. Let

$$\begin{aligned}
\mathfrak{A}_0^\dagger &= (\mathfrak{AB}_0^\dagger - \{AB_0^+\})|_{A_0},\\
\mathfrak{B}_0^\dagger &= (\mathfrak{AB}_0^\dagger - \{AB_0^+\})|_{B_0},\\
\mathfrak{A}^+ &= \mathfrak{A}_0^\dagger \vee \mathfrak{A}^\dagger,\\
\mathfrak{B}^+ &= \mathfrak{B}_0^\dagger \vee \mathfrak{B}^\dagger,\\
\mathfrak{AB}^+ &= \{AB_0^+\} \vee \mathfrak{A}^+ \vee \mathfrak{B}^+.
\end{aligned}$$

Thus $\mathfrak{AB}^+$ is obtained from $\mathfrak{AB}^\dagger$ by shearing off the at most one escaping bipod from the distinguished descendent of AB_0, and then breaking up each undistinguished descendent of AB_0 into the part that lies in A_0 and the part that lies in B_0. In each of these latter breaks we take the south cell to be the intersection with D_0. (This will entail increasing the break size.)

We define $C_0^+ = C_0$, $\mathfrak{C}^+ = \mathfrak{C}$, $ABC_0^+ = AB_0^+ \cup C_0^+$, and

$$\mathfrak{ABC}^+ = \{ABC_0^+\} \vee \mathfrak{A}^+ \vee \mathfrak{B}^+ \vee \mathfrak{C}^+.$$

Set $D(A_0^+) = A_0^+ - D'_0$, $D(B_0^+) = B_0^+ - D'_0$, $D(C_0^+) = D(C_0) \cup VD_0 - D'_0$.

Thus we have added back the terms in $\mathfrak{C}$, and added back C_0 to the distinguished term, including the at most one escaping bipod. □

There exists a (g,h,i)-factorization tree and a (g,h,i)-shearing sequence which converts $VABC$ into $\mathfrak{ABC}$, and we have a family of factorization trees which convert $\mathfrak{ABC}$ into $\mathfrak{ABC}^+$, and there is a main line from ABC_0 to $ABBC_0^+$, and all of C_0 travels down it.

This is a first approximation to an admissible factorization, but we have not even analyzed what sort of breaks occur in the new family of factorization trees.

We now consider the five types separately, in descending order, starting with Type V, and then dealing with the other cases by varying the argument appropriately.

§8. Type V

8.1 Hypotheses. Throughout this section, we suppose $(g, h, i) = (1, h, h)$, $h \geq 2$, so §4.15 applies. □

For Type V, D_0' is obtained from D_0 by removing unipods.

8.2 Definitions. Set $(g^+, h^+, i^+) = (g_q, h_q, g_q h_q)$. Thus (g^+, h^+, i^+) is Type V, and $(g^+, h^+) = (g, h) + (a, b) - (a', b')$. □

We shall be temporarily interested in (g^+, h^+, i^+)-breaks.

There are no escaping bipods, so $\mathfrak{AB}^+$ is obtained from $\mathfrak{AB}^\dagger$ by breaking up each undistinguished descendent of AB_0 into the part that lies in A_0 and the part that lies in B_0. In the factorization tree, this gives a break in which the south cell is the intersection with D_0, so this south cell is 2-partitioned into, first, the intersection with D_0' and, second, the intersection with $D_0 - D_0'$. These are (g, h)- and $(g^+ - g, h^+ - h)$-trivial, respectively, so the south cell is (g^+, h^+)-trivial. Hence $\mathfrak{AB}^+$ is a (g^+, h^+, i^+)-factorization of $\mathfrak{AB}$, and hence is a shearing of a (g^+, h^+, i^+)-factorization of AB which properly refines $\mathfrak{AB}$.

We now describe the changes that have to made in $\mathfrak{ABC}^+$.

8.3 Adjustments. Let us say that, in our context, an ABC-bond e is *unbalanced* if e is an ABC-edge or an $\mathfrak{ABC}^+$-bond such that e interferes in the factorization tree of $\{ABC_0\} \vee \mathfrak{A} \vee \mathfrak{B}$, that is, there is some break where e interferes, as in Definition 2.1.

Each unbalanced bond will be accommodated by increasing the break size by $(0, 1, 1)$ and moving the yellow vertex to the south cell. By Proposition 3.5, we can shear to get back the set that we started with before the bond was taken into account, unless we have to accommodate two unbalanced bonds with the same yellow vertex, in which case tripods or tripod pairs may get added.

Let us describe the procedure in detail by considering various cases, subcases, etc.

Case 1. The factorization of ABC_0.

Here we have to consider the effect of adding C_0 and any C-edges and $\mathfrak{C}$-bonds to each term in the main line of the factorization of AB_0. There are two sorts of unbalanced bonds to take into account. Firstly, the $\mathfrak{ABC}^+$-bonds or ABC-edges joining $D_0 - D_0'$ to AB_0, and secondly, the $\mathfrak{ABC}$-edges joining $C_0 - D_0$ to $AB_0 - AB_0^+$. Once we have accommodated such bonds, there is no obstacle to adding C_0 and any C-edges and $\mathfrak{C}$-bonds to each term in the main line.

Case 1.1. $\mathfrak{ABC}^+$-bonds and ABC-edges joining $D_0 - D_0'$ to AB_0.

Let y be a vertex in $D_0 - D_0'$, so y is yellow.

We increased the break size by $(0, h^+ - h, h^+ - h)$ when we passed from (g, h, i) to (g^+, h^+, i^+), so, for each of the $h^+ - h$ (yellow) vertices in $D_0 - D_0'$, we have already increased the break size by $(0, 1, 1)$. This means that we are free to move y into the south cell of any break without increasing the break size further. We investigate the consequences.

Recall, from §6, that there is some $r \in \{0, \dots, q-1\}$ such that y lies in $D(r)$ but not in $D(r+1)$, and not even in $AB(r+1)$. At this break of $AB(r)$, in the generic case, the south cell has a red vertex x_y, and we let e_y denote the bond joining y and x_y. (As usual, the degenerate case where x_y does not exist can be handled in the same way, and we do not discuss it.) We have already added y (and all of C_0) to the main line, so we move y from the east cell to the south cell at this juncture. The descendents of the west child of $AB(r)$ do not contain y, so y gets added to the west cell at each subsequent step of the main line.

Recall that, if an undistinguished term of $\mathfrak{AB}_0^\dagger$ contains y, then y is in the south cell of the break into the parts that lie in A_0 and B_0.

Let f_y denote the bond joining y to the red vertex x of $D - D_0'$. It is possible that, f_y is an unbalanced exceptional bond. It is also possible that $e_y = f_y$, in which case $x \in D(A_0)$.

We now have the breaks arranged so that e_y and f_y are the only unbalanced bonds incident to y, and e_y now lies in both children of $AB(r) \cup C_0$. We proceed as in the proof of Proposition 3.5, moving y into the south cell of each break where e_y has to be accommodated. This creates bipods, which we then shear. If e_y does not have to be accommodated, it becomes a C-exceptional bond incident to y. If $e_y = f_y$, then e_y can become doubly exceptional. If $e_y \neq f_y$, and we are in the inner case, then either $x \in D(C_0)$ or e is C-like in ABC_0^+.

We remark that all bonds which join y to D_0', other than e_y, have become C-like.

Case 1.2. $\mathfrak{ABC}$-edges joining $C_0 - D_0$ to $AB_0 - AB_0^+$.

Let e_1 be such a bond, let $y = y(e_1)$, $x_1 = r(e_1)$. Notice that e_1 is A-, B- or C-exceptional.

Case 1.2.1. y occurs in a C-tripod in ABC_0.

Here we may assume that, in the (g, h, i)-atomic factorization of AB, y is removed as part of a unipod in a $(1,0,0)$-break, which is then discarded. We add y to every term of the main line, and move it into the south cell of the unipod break, which becomes a $(1,1,1)$-break. We attempt to restore the C-tripod to ABC_0^+. This succeeds unless one of the three red vertices involved lies in $AB_0 - AB_0^+$. In the latter case, there is a corresponding break on the main line, and we increase the break size by $(0,1,1)$, and add y to the south cell and the east child. We then shear a sequence of bipods to eliminate all that is left of the tripod or tripod pair.

Case 1.2.2. y does not occur in a C-tripod in ABC_0.

Here we accommodate e_1 as in Proposition 3.5 by increasing the break size by $(0,1,1)$. If there is no other edge incident to y to be accommodated, we have created bipods which we then shear. Thus e_1 becomes unexceptional, unless it was doubly exceptional and becomes D-exceptional. This leaves the case where there is another edge e_2 incident to y to be accommmodated. By the conditions for Type V, neither e_1 nor e_2 is doubly exceptional.

Case 1.2.2.1. e_1 and e_2 occur in a C-tripod.

Here, after the above-mentioned bipod is sheared off ABC_0^+, the C-tripod becomes a bipod, and we shear it, and if there is a C-tripod pair, the other tripod becomes a bipod and gets sheared. In this case e_1 will not be an $\mathfrak{A}\mathfrak{B}\mathfrak{C}'$-bond.

Case 1.2.2.2. e_1 and e_2 are ABC_0-bonds.

Let $x_2 = r(e_2)$. Here y must lie in $D - D_0$, and we may assume it lies in $D(A_0)$ or $D(C_0)$. We may also assume that, in the inner case, if $x \in \{x_1, x_2\}$ then $x_1 = x$.

Case 1.2.2.2.1. y lies in $D(A_0)$.

Here y lies in $D(A_0) - D(A_0^+)$, and we may end up adding a C-tripod involving y, e_1 and e_2 to ABC_0^+. Suppose that e_1 has to be accommodated. If $x_1 \in AB_0$ then e_1 is not an AB-edge nor a bond of $\mathfrak{A}^+ \cup \mathfrak{B}^+$, so e_1 must have become C-like. If $x_1 \in C_0 - D_0'$, then either $x_1 = x \in C(D_0)$ or $x_1 \in C_0 - C(D_0)$. In the latter case, e_1 is clearly C-like. In the former case, by the conditions for Type V, e_1 is C-like. Thus, in all cases, e_1 is C-like, and similarly for e_2.

There is a least r such that in the break of $AB(r)$, y lies in the north or east cell, and it is moved into the south cell, following the procedure of Proposition 3.5. If y lies in the north cell, then, after shearing, nothing is added, or a C-tripod is added to ABC_0^+, or a C-tripod pair is added to ABC_0^+ and another term. If y lies in the east cell, then, after shearing, nothing is added, or a C-tripod is added to ABC_0^+.

This completes the list of adjustments to ABC_0^+, and we now have the vertex sets ABC_0', A_0', B_0', C_0', $D(A_0')$, $D(B_0')$, $D(C_0')$, AB_0', BC_0', CA_0'.

Case 1.2.2.2.2. y lies in $D(C_0)$.

For $j = 1, 2$, x_j lies in $AB_0 - AB_0^+$, and for some r_j, x_j lies in the east cell of the break of $AB(r_j)$, and y is moved into the south cell, following the procedure of Proposition 3.5. If $x_j = x$ then $x \in D(A_0)$, and, by the conditions for Type V, e_j is A-like. Here we pass from the inner to the outer case. In the most complicated generic case, $r_1 = r_2$ and there is a red vertex x_3 in the south cell of the break of $AB(r_1) = AB(r_2)$. We let e_3 denote the bond joining y and x_3. Here, after shearing, either we have added nothing, or we have added a tripod $\{y, e_1, e_2, e_3\}$ to a unique term of $\mathfrak{A}^+$. Thus e_1, e_2 have become A-like and unexceptional, while e_3 has become an A-exceptional bond incident to y.

Case 2. Factorizations of $\mathfrak{A}$ and $\mathfrak{B}$.

It remains to consider the impact of adding in $\mathfrak{C}$-bonds, C-edges, and bonds joining $C_0 - D_0$ to AB_0^+, to the factorizations of terms of $\mathfrak{A}$ and $\mathfrak{B}$. Here we have factorizations of $\mathfrak{A}$, $\mathfrak{B}$ and descendents of AB_0, arising from a factorization of AB.

Any $\mathfrak{C}$-bond or C-edge that needs to be accommodated in a break off the main line must be a bond of $\{ABC_0\} \cup \mathfrak{A} \cup \mathfrak{B}$, but not an AB-edge nor a bond of $\{ABC_0'\} \vee \mathfrak{A}^+ \vee \mathfrak{B}^+$. Thus it must be a D-bond, and not a D-edge nor a D_0'-bond. We have already dealt with the D_0-bonds, so we are left with a D-bond which is not a D-edge and has a vertex in $D - D_0$. This is an exceptional bond.

Any imbalanced bond joining $C_0 - D_0$ to AB_0' that needs to be accommodated off the main line must be a C-exceptional ABC_0-bond which is an ABC_0'-bond, but not an AB-edge nor a bond of $\mathfrak{A}^+ \cup \mathfrak{B}^+$.

To accommodate such a bond, we increase the break size by $(0, 1, 1)$, proceed as in Proposition 3.5, moving the yellow vertex in the breaks involved from one of the three other cells to the south cell. We then shear. Bases of C-tripods or C-tripod pairs are handled as in Case 1.2.1. If there are two exceptional bonds incident to the same yellow vertex, it may transpire that we add a C-tripod to a term of $\mathfrak{A}^+$ or $\mathfrak{B}^+$, or a C-tripod pair to two terms of $\mathfrak{A}^+$ or $\mathfrak{B}^+$.

If the $\mathfrak{C}$-bond involved lies in an A-tripod or A-tripod pair in $\mathfrak{C}$, then one of the A-tripods becomes a bipod, which we then shear, and the other A-tripod then becomes a bipod which we shear. These are the only changes made to $\mathfrak{C}$.

We have now described all the adjustments we make to obtain $\mathfrak{A}'$, $\mathfrak{B}'$ and $\mathfrak{C}'$. $\square$

8.4 Lemma. $c' + 2(n - b') \geq p + h' + d'$.

Proof. We have $c + 2(n - b) \geq p + h + d$, so it suffices to show that

$$(c' - c) - 2(b' - b) \geq (h' - h) + (d' - d),$$

that is, $2(b - b') + (d - d') \geq (c - c') + (h' - h)$.

Let Δ denote the set of (yellow) vertices in $D_0 - D'_0$, so each element of Δ has valence at most one in D_0. For $j = 0, 1$, let Δ_j denote the set of vertices of valence exactly j in D_0.

Now, $h^+ - h = b - b' = |\Delta|$, and $c - c' = |\Delta_1|$. Let $d^+ - d$ denote the number of y such that e_y, described in Case 1.1 above, exists and is C-exceptional, so, in particular, is not an edge of D. Thus, $d^+ - d \leq |\Delta_0|$, and $|\Delta| \geq (c - c') + (d^+ - d)$. But

$$|\Delta| = h^+ - h = 2(b - b') - (h^+ - h),$$

so $2(b - b') + (d - d^+) \geq (c - c') + (h^+ - h)$.

Thus it suffices to show that $d^+ - d' \geq h' - h^+$, so it suffices to show that, starting at (g^+, h^+, i^+), each time we increased the break size by $(0, 1, 1)$, the weight sum dropped by at least one. This can be checked by straightforward case-by-case considerations. □

It is now a simple matter to show that $\mathfrak{ABC}'$ is an admissible factorization. It is clear that $\mathfrak{ABC}'$ is a refinement of $\mathfrak{ABC}$, and it is a proper refinement, since the distinguished term YC is replaced by subsets with fewer red vertices, and restoring yellow vertices in tripods does not alter this.

§9. Type IV

9.1 Hypotheses. Throughout this section, we suppose $(g, h, i) = (1, 1, 0)$, so §4.12 applies. □

For Type IV, D'_0 is obtained from D_0 by removing unipods, and at most one bipod corresponding to a crossover vertex.

9.2 Definitions. If $D'_0 = D_0$, no adjustments will be required, and we take $\mathfrak{ABC}' = \mathfrak{ABC}^+$, and $(g', h', i') = (g, h, i) = (1, 1, 0)$, which is Type IV. It is straightforward to see that we still have the conditions of Type IV, so we have an admissible factorization which is a proper refinement of $\mathfrak{ABC}$. □

9.3 Hypotheses and notation. We may, and shall, assume $D_0' \neq D_0$.

Here, in the generic case, there is an $r \in \{1, \ldots, q\}$ such that $D_0 = D(r)$, and the break of $AB(r)$ has a crossover vertex y, and a vertex x_1 in the south cell, and a crossover edge (an $\mathfrak{AB}$-edge) e_2 incident to a vertex x_2 in the west cell, and, moreover, y lies in D_0.

We shall assume these conditions hold, with this notation; as usual, we leave the degenerate cases to the reader.

Let e_1 denote the bond joining y and x_1. We call $\{y, e_1, e_2\}$ the *y-bipod.*

Since $D_0' \neq D_0$, $h_q \geq 2$.

Recall that the crossover vertex y lies in $AB(r+1)$, although we have formally removed it from $D(r+1)$.

If y lies in $AB(q)$ then it is sheared off in AB_0^+ but added back in ABC_0^+, to formally move y into $D(C_0)$.

If y does not lie in $AB(q)$ then there is a least $r' \in \{r+1, \ldots, q-1\}$ such that y lies in the east or north cell of the break of $AB(r')$. If y lies in the north cell, we may assume that the break of $AB(r'+1)$ consists of removing y as part of a unipod, which then gets deleted. We retain this notation for this case throughout.

If the crossover vertex y lies in $AB(q)$, then we set

$$(g^+, h^+, i^+) = (g_q, h_q, g_q h_q) = (1, h_q, h_q),$$

which is Type V.

If y does not lies in $AB(q)$, then we set

$$(g^+, h^+, i^+) = (g_q, h_q + 1, g_q h_q + 1) = (1, h_q + 1, h_q + 1),$$

which is also Type V. □

We proceed as for Type V.

Consider a south cell in a break in the step from $\mathfrak{AB}^\dagger$ to $\mathfrak{AB}^+$; such a south cell is the intersection of D_0 with the term involved. If y lies in $AB(q)$, then this south cell is 2-partitioned into, first, the intersection with $D_0' \cup \{y\}$, and, second, a subset of $D_0 - (D_0' \cup \{y\})$. These are $(1,2)$- and $(g_q - 1, h_q - 2)$-trivial, respectively, so the south cell is (g_q, h_q)-trivial. If y does not lie in $AB(q)$, then our south cell is 3-partitioned into, first, the intersection with D_0', and second, a subset of $D_0 - (D_0' \cup \{y\})$, and third, the intersection with y. These are $(1,2)$-, $(g_q - 1, h_q - 2)$, and $(0,1)$-trivial, respectively, so the south cell is $(g_q, h_q + 1)$-trivial. Hence $\mathfrak{AB}^+$ is a (g^+, h^+, i^+)-factorization of $\mathfrak{AB}$, and hence is a shearing of a (g^+, h^+, i^+)-factorization of AB which properly refines $\mathfrak{AB}$.

9.4 Adjustments. We proceed as in §8.3. The absence of exceptional bonds for Type IV simplifies the procedure, and many of the cases can be ignored here. Each (yellow) vertex of $D_0 - (D_0' \cup \{y\})$ is dealt with as before, and, as before, the break size is increased by $(0,1,1)$. The only novel feature we have to consider is the bipod $\{y, e_1, e_2\}$ on $AB(r+1)$. Notice that the basic increase in break size associated with y is $(0,1,2)$, but this increase is augmented by $(0,1,1)$ to $(0,2,3)$ if y does not lie in $AB(q)$, so breaks off a second time.

Here $y \in D(C_0^+)$, and, for $j = 1, 2$, e_j may become C-exceptional. By the conditions for Type IV, e_j is not exceptional with respect to $\mathfrak{A}\mathfrak{B}\mathfrak{C}$, so if $x_j = x$, then e_j is not A-exceptional, although $x_j = x \in D_0(A)$.

Notice that e_1 and e_2 must be AB-edges or bonds of $\mathfrak{A}^\dagger \vee \mathfrak{B}^\dagger$, since they occur in the break of $AB(r)$.

Case 1. y lies in $AB(q)$.

Here y behaves like an element of D_0', and no adjustments are required. Here x_1 and x_2 also must lie in $AB(q)$, since otherwise the bipod is replaced with a unipod, which would be broken off in an atomic factorization.

Case 2. y does not lie in $AB(q)$.

Here we have increased the break size by another $(0,1,1)$, so we are free to move y into the south cell in all breaks where necessary. Let x_3 denote the red vertex in the south cell of the break of $AB(r')$, and let e_3 denote the bond joining y and e_3. (We leave to the reader the simpler case where the south cell does not have a red vertex.) In this case, y behaves like a yellow vertex of $D_0(A)$ incident to two A-exceptional bonds in Type V.

Case 2.1. y lies in the north cell of the break of $AB(r')$, and is broken off $AB(r'+1)$ and deleted.

Here we argue as in §8.3, Case 1.2.2.2.1, since y behaves like an element of $D(C_0)$ incident to two C-exceptional bonds, or like a yellow vertex which escapes from the nucleus in the north cell. In the generic case, x_1 lies in the west cell and x_2 in the east cell. On moving y into the south cell, we add a bipod $\{y, e_2, e_3\}$ to the east child, and will shear it. Then e_2 and e_3 become unexceptional, although e_1 may still become C-exceptional. Recall that y is restored to the rest of the main line.

Case 2.1. y lies in the east cell of the break of $AB(r')$.

Here x_1 and x_2 lie in the union of the east and south cells. We assume they both lie in the east cell, since the case where $x_1 = x_3$ is similar. By §6, x_1 and x_2 do not lie in D_0. On moving y to the south cell, we add a unipod $\{y, e_3\}$ to the west child, so e_1 and e_2 become unexceptional, while e_3 may become C-exceptional. □

We now have $\mathfrak{A}\mathfrak{B}\mathfrak{C}'$, (g', h', i'), and d'.

9.5 Lemma. $c' + 2(n - b') \geq p + h' + d'$.

Proof. For Type IV, $(a, b, c) = (m - 1, n - 1, p)$, $(g, h, i) = (1, 1, 0)$.

Thus $c + 2(n - b) = p + 2(1) = p + h + d + 1$, so it suffices to show that $(c' - c) - 2(b' - b) \geq (h' - h) + (d' - d) - 1$, or equivalently,

$$2(b - b') + (d - d') + 1 \geq (c - c') + (h' - h).$$

Let D_0^* denote the graph obtained from D_0 by deleting (the intersection of D_0 and) the y-bipod. Let $\text{size}(D_0^*) = (a^*, b^*, c^*)$, so $a^* = m-1$, $b^* = n-2$, $p \geq c^* \geq p - 2$. The number of bonds in the bipod which are not edges of D is $2 - (p - c^*)$, so the contribution d^* of the two bonds in the y-bipod to the weight sum is at most $2 - (p - c^*)$. Thus $c^* + 2 \geq p + d^*$. Since $2(n - b^*) = 4$ and $h^* = 2$, we see $c^* + 2(n - b^*) \geq p + h^* + d^*$.

Now we proceed as in the proof of Lemma 8.4. □

It is now a simple matter to check that $\mathfrak{A}\mathfrak{B}\mathfrak{C}'$ is an admissible factorization. It is clear that $\mathfrak{A}\mathfrak{B}\mathfrak{C}'$ is a refinement of $\mathfrak{A}\mathfrak{B}\mathfrak{C}$, and it is a proper refinement, since the distinguished term YC is replaced by subsets with fewer red vertices, and adding back yellow vertices in tripods does not alter this.

§10. Type III

10.1 Hypotheses. Throughout this section, we suppose $(g, h, i) = (1, 1, 1)$, so §4.10 applies. □

For Type III, D_0' is obtained from D_0 by removing unipods.

10.2 Definition. If $D_0' \neq D_0$, we set $(g^+, h^+, i^+) = (g_q, h_q, g_q h_q)$, which is Type V. We make adjustments as in §8.3, although many of the cases obviously do not apply. Set $d = 1$ if there is an exceptional bond and it is not an edge of D; otherwise, set $d = 0$. Thus if $d = 1$ then $c = p = p - 1 + d$, and if $d = 0$ then $c \geq p - 1 = p - 1 + d$ Hence

$$c + 2(n - b) = c + 2(1) \geq (p - 1 + d) + 2(1) = p + 1 + d = p + h + d.$$

As in the proof of Lemma 8.4, we find $c' + 2(n - b') \geq p + h' + d'$. We then obtain an admissible factorization which is a proper refinement of $\mathfrak{A}\mathfrak{B}\mathfrak{C}$. □

Thus we may assume that $D_0' = D_0$ for the remainder of this section.

10.3 Definition. If there is an exceptional bond and it interferes with the factorization tree of $\mathfrak{AB}$ then we change the break size to $(1,1,0)$, Type IV, and move the yellow vertex into the south cell as a crossover vertex, and shear. All the conditions of Type IV are then satisfied, and we obtain an admissible factorization which is a proper refinement of $\mathfrak{ABC}$. □

Thus, for the remainder of this section, we may further assume that, even if there is an exceptional bond, it does not interfere with the factorization tree of $\mathfrak{AB}$.

10.4 Definition. Here we take $(g', h', i') = (g, h, i) = (1,1,1)$, Type III, and take $\mathfrak{ABC}' = \mathfrak{ABC}^+$. There is no shearing involved. The only verification that is not completely straightforward, is the fact that $\mathfrak{BC}'$ is a (g, h, i)-factorization of BC, with permutations of A, B, C applying. The only difficulty arises if $x \in D(A_0) - D(A'_0)$, which means that we pass from the inner case to the outer case. In the generic case, there is a unique term Z of $\mathfrak{A}'$ whose intersection with D has more than one red vertex, and this intersection consists of two red vertices, x and x_1, and two yellow vertices, y and y_1, and the bond (x_1, y_1) is a D'_0-bond, and there is an exceptional bond, and it joins x to y or y_1. In $\mathfrak{A}'|_{BC}$, all terms other than $Z|_{BC}$ have at most one red vertex so contain no (g, h, i)-atomic factor, and can be discarded. Then $Z|_{BC}$ is left with four vertices and two bonds, with respect to $(\mathfrak{ABC}' - \mathfrak{A}')$-bonds and BC-edges. This can then be discarded. Now $D(A'_0)$ can be removed by discarding unipods, and we see that $\mathfrak{BC}'$ is a (g, h, i)-factorization of BC, as desired. Hence all the conditions of Type III are satisfied, so we have a proper refinement. □

§11. Type II

11.1 Hypotheses. Throughout this section, we suppose $(g, h, i) = (1, 0, 0)$, so §4.8 applies. □

For Type II, D'_0 is obtained from D_0 by removing unipods, and each unipod which escapes from the nucleus is the restriction of a specified unipod of AB, namely, that determined by the red vertex, if any, which lies in the south cell of the break of AB where the unipod escapes from the nucleus.

11.2 Definitions. We take $\mathfrak{ABC}' = \mathfrak{ABC}^+$. In the factorization tree, we make the adjustments as in §8.3, moving (yellow) vertices of $D_0 - D'_0$ into the south cell of any break where the bond in the unipod of AB interferes.

Case 1. No unipods escape from the nucleus.

We take $(g', h', i') = (g, h, i) = (1, 0, 0)$, Type II.

Case 2. Exactly one unipod escapes from the nucleus.

Let y denote the yellow vertex. In the generic case, there is a red vertex x_y in the south cell of the break in question. Let e_y denote the bond joining y and x_y. It is potentially exceptional.

Case 2.1. e_y does not interfere in the factorization tree.

We take the break size (g', h', i') to be $(1, 1, 1)$, Type III, and we have one exceptional bond.

Case 2.2. e_y does interfere in the factorization tree.

Thus there is a break in which x_y is in the east cell and y is in the west cell, or vice-versa, or there is an $r \in \{1, \ldots, q-1\}$ such that $x_y \in AB(r) - AB(r+1)$.

In the generic case there is a red vertex x_2 (and nothing else) in the south cell, and we let e_2 denote the bond joining x_2 and y.

We take the break size (g', h', i') to be $(1, 1, 0)$, Type IV, and move y into the south cell creating a bipod $\{y, e_y, e_2\}$, which we then shear, so there are no exceptional bonds. It is interesting to note that we are free to decide whether y is a crossover vertex or not.

Case 3. Two or more unipods escape from the nucleus.

We take $(g', h', i') = (g_q, h_q, g_q h_q)$, Type V, and move the escaping yellow vertices into the south cell wherever necessary. Here we set $d = 0$, so

$$c + 2(n - b) = p + h + d$$

since $n - b$ and h are both 0. As in Lemma 8.4, we get

$$c' + 2(n - b') = p' + h' + d'. \quad \square$$

In all cases it is straightforward to check that we have a proper refinement.

§12. Type I

12.1 Hypotheses. In this section, we suppose $(g, h, i) = (0, h, 0)$, $h \geq 0$, so §4.8 applies. $\square$

For Type I, D_0' is obtained from D_0 by removing nullipods, and at most one red vertex x incident to at most h yellow vertices of valence two or more, and unipods attached to x.

12.2 Definition. The procedure is as in the previous sections. We take $\mathfrak{ABC}' = \mathfrak{ABC}^+$, and assume the appropriate adjustments have been made in the factorization tree.

Case 1. No red vertex escapes.

We take $(g', h', i') = (g_q, h_q, g_q h_q) = (0, h_q, 0)$, Type I.

Case 2. The red vertex x escapes.

Here so $g_q = 1$ and there exists $r \in \{0, \ldots, q-1\}$ such that

$$x \in AB(r) - AB(r+1).$$

Let $y_1, \ldots, y_N$ be an enumeration of the (at most h) (yellow) vertices in the south cell of the break of $AB(r)$ together with the (at most $h_q - h$) yellow vertices of D_0 in the east cell. (The north cell is empty.) Thus $N \leq h_q$. For each $i \in \{1, \ldots, N\}$, let e_i be the bond joining x to y_i. These are the only bonds that require attention. If e_i does not interfere with the factorization tree of AB, then e_i becomes C-exceptional. If e_i does interfere with the factorization tree, then we move y into the south cell and shear. We have to describe the break size implicit in this procedure.

Case 2.1. $h_q = 0$.

We take $(g', h', i') = (g_q, h_q, g_q h_q) = (1, 0, 0)$, Type II.

Case 2.2. $h_q = 1$.

Here $0 \leq N \leq 1$.

Case 2.2.1. $N = 0$.

We take $(g', h', i') = (g_q, h_q, g_q h_q) = (1, 1, 1)$, Type III.

Case 2.2.2. $N = 1$.

Case 2.2.2.1. e_1 does not interfere with the factorization tree.

We take $(g', h', i') = (g_q, h_q, g_q h_q) = (1, 1, 1)$, Type III.

Case 2.2.2.2. e_1 interferes with the factorization tree.

We take $(g', h', i') = (g_q, h_q, g_q h_q - 1) = (1, 1, 0)$, Type IV, and, wherever e_1 interferes, we move y_1 into the south cell *as a crossover vertex* and shear.

Case 2.3. $h_q \geq 2$.

We take $(g', h', i') = (g_q, h_q, g_q h_q) = (1, h_q, h_q)$, Type V. □

Again we have a proper refinement.

§13. Conclusion

We have now covered all the cases and arrived at a proper refinement in each one. This gives a contradiction, as noted in §5.3. This we have proved Theorem 1.3, and hence Theorem 0.1.

References

1. R. G. Burns, *On the intersection of finitely generated subgroups of a free group*, Math. Z. **119** (1971), 121–130.
2. Warren Dicks, *Equivalence of the strengthened Hanna Neumann conjecture and the amalgamated graph conjecture*, Invent. Math. **117** (1994), 373–389.
3. H. Neumann, *On intersections of finitely generated subgroups of free groups. Addendum*, Publ. Math. Debrecen **5** (1958), 128.
4. W. D. Neumann, *On intersections of finitely generated subgroups of free groups*, Lecture Notes in Mathematics **1456** (1990), 161–170.
5. G. Tardos, *On the intersection of subgroups of a free group*, Invent. Math. **108** (1992), 29–36.
6. G. Tardos, *Toward the Hanna Neumann conjecture using Dicks' method*, Invent. Math. **123** (1996), 95–104.

Warren Dicks, Departament de Matemàtiques, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
E-mail address: dicks@mat.uab.es

Edward Formanek, Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, U.S.A.
E-mail address: formanek@math.psu.edu