A COMBINATORIAL DECOMPOSITION THEORY

WILLIAM H. CUNNINGHAM AND JACK EDMONDS

1. Introduction. Given a finite undirected graph G and $A \subseteq E(G)$, G(A) denotes the subgraph of G having edge-set A and having no isolated vertices. For a partition $\{E_1, E_2\}$ of E(G), $W(G; E_1)$ denotes the set $V(G(E_1)) \cap V(G(E_2))$. We say that G is *non-separable* if it is connected and for every proper, non-empty subset A of E(G), we have $|W(G; A)| \geq 2$. A *split* of a non-separable graph G is a partition $\{E_1, E_2\}$ of E(G) such that

 $|E_1| \ge 2 \le |E_2|$ and $|W(G; E_1)| = 2$.

Where $\{E_1, E_2\}$ is a split of G, $W(G; E_1) = \{u, v\}$, and e is an element not in E(G), we form graphs G_i , i = 1 and 2, by adding e to $G(E_t)$ as an edge joining u to v. In this situation we write $G \rightarrow \{G_1, G_2\}$, and call $\{G_1, G_2\}$ a simple decomposition of G, associated with the split $\{E_1, E_2\}$ and the marker e. This paper describes a unique decomposition theory which includes among its applications a theory of graph decomposition based on this notion of simple decomposition. In this section we continue with the description of this instance of the theory.

Let G be a non-separable graph. A decomposition D of G is defined inductively to be either $\{G\}$ or a set obtained from a decomposition D' of G by replacing a member G_1 of D' by the members of a simple decomposition of G_1 , where the marker of this simple decomposition is not an edge of any member of D'. If D'' is obtained from D by a (non-empty) sequence of operations of the kind described above, then D'' is said to be a (*strict*) refinement of D. If the sequence consists of exactly one operation, the refinement is *simple*.

We can associate a graph T with any decomposition D of a nonseparable graph G. The vertices of T are the members of D and the edges are the markers of D; each marker joins in T the two members of D of which it is an edge. It is clear that T is a tree. This "decomposition tree" provides a useful way to visualize a decomposition.

Two decompositions D, D' of G are *equivalent* if D' can be obtained from D by replacing some of the markers of D by markers of D'. All unique decomposition theorems of this paper involve uniqueness "up to equivalence", but we will tend not to include this phrase in their statements. The decomposition D of G is *minimal* with some property P if Dhas P and there does not exist a decomposition D' of G also having P,

Received September 19, 1978 and in revised form May 2, 1979.

such that D is a strict refinement of D'. A decomposition D is *trivial* if |D| = 1. A non-separable graph G is *prime* if it has no non-trivial decomposition. We observe that the prime graphs are precisely those which are 3-connected in the sense of Tutte [23].

One might hope that each non-separable graph G has a unique decomposition consisting of prime graphs. Examples of graphs which are badly behaved in this regard are the *polygons* (connected graphs in which each vertex has degree 2) and *bonds* (connected graphs having 2 vertices and no loops). Bonds and polygons having 4 or more edges have at least two different (inequivalent) prime decompositions. (A bond or polygon having 6 or more edges can have two prime decompositions having nonisomorphic decomposition trees.) Other types of graphs can also have more than one prime decomposition, but bonds and polygons play a special role in the uniqueness theory. The following result is the main unique decomposition theorem for graphs; its proof (as well as a discussion of its relation to other work on graph decomposition) appears in Section 4.

THEOREM 1. Let G be a non-separable graph. Then G has a unique minimal decomposition, each of whose members is prime, a polygon, or a bond.

In the remainder of this section, we derive some elementary properties of simple decompositions and splits of non-separable graphs. These properties will motivate the definition of "decomposition frame" (Section 2), which is the context of the main theorems. It is convenient at this point to introduce some notation. If G is a non-separable graph, $\{E_1, E_2\}$ is a split of G, and $e \notin E(G)$, then we use $G(E_1; e)$, $G(E_2; e)$ to denote the members of the (unique) simple decomposition of G associated with $\{E_1, E_2\}$ and e.

LEMMA 1. If G is a non-separable graph and $G \rightarrow \{G_1, G_2\}$, then G_1 and G_2 are non-separable.

Proof. Let $\{E_1, E_2\}$ be the split and e be the marker associated with $\{G_1, G_2\}$. It is easy to see that, for any $A \subseteq E_1$, $W(G_1; A) = W(G; A)$. The result follows.

LEMMA 2. Let $G \to \{G_1, G_2\}$ with marker e, and let $A \subset E_1$. Then $\{A, E(G) \setminus A\}$ is a split of G if and only if $\{A, (E_1 \setminus A) \cup \{e\}\}$ is a split of G_1 .

Proof. Again, the result follows from the fact that $W(G; A) = W(G_1; A)$.

LEMMA 3. Let $\{E_1, E_2\}$ and $\{E_3, E_4\}$ be splits of G such that $E_3 \subset E_1$.

Let e, f be distinct elements which are not edges of G. Then

$$G(E_1; e)(E_3; f) = G(E_3; f), and$$

$$G(E_1; e)((E_1 \setminus E_3) \cup \{e\}; f) = G(E_4; f)((E_4 \setminus E_2) \cup \{f\}; e).$$

Proof. As before, for any $A \subseteq E_1$, we have

$$W(G(E_1; e); A) = W(G; A) \text{ and } W(G(E_1; e); A \cup \{e\})$$

 $= W(G; A \cup E_2).$

The result follows from these formulae, because $W(H; \{h\})$ is the set of ends of h in H, for any edge h of graph H.

LEMMA 4. Let $\{E_1, E_2\}$ and $\{E_3, E_4\}$ be splits of the graph G, such that $|E_1 \cap E_3| \ge 2$ and $E_1 \cup E_3 \ne E(G)$. Then $\{E_1 \cap E_3, E_2 \cup E_4\}$ is a split of G.

Proof. For this proof, we abbreviate W(G; A) to W(A), for any $A \subseteq E(G)$. Clearly, $W(E_1 \cap E_3)$, $W(E_2 \cap E_4) \subseteq W(E_1) \cup W(E_3)$. Any element v of $W(E_1 \cap E_3) \cap W(E_2 \cap E_4)$ must be in $W(E_1) \cap W(E_3)$. (If not, suppose that $v \in W(E_1) \setminus W(E_3)$; then v is incident with no member of E_4 , so $v \notin W(E_2 \cap E_4)$.) Thus

$$|W(E_1 \cap E_3)| + |W(E_2 \cap E_4)| \le |W(E_1) \cap W(E_3)| + |W(E_1) \cup W(E_3)| = |W(E_1)| + |W(E_3)| = 4.$$

But we are given that $E_2 \cap E_4 \neq \emptyset$, and so, since G is non-separable,

 $|W(E_2 \cap E_4)| \ge 2.$

The result follows.

We introduce here some further notational conventions. Where A is a set and e is an element, we abbreviate $A \cup \{e\}$ to A + e and $A \setminus \{e\}$ to A - e. In the absence of parentheses, set operations are to be performed from the left; for example, $A \setminus B + e$ denotes $(A \setminus B) \cup \{e\}$. For sets A and B, A meets B means that $A \cap B \neq \emptyset$. Partitions $\{A_1, A_2\}$ and $\{B_1, B_2\}$ of a set E are said to cross if $A_1 \cap B_1$, $A_1 \cap B_2$, $A_2 \cap B_1$, and $A_2 \cap B_2$ are all non-empty.

The organization of the paper is as follows. In Sections 2 and 3 we introduce "decomposition frames" and prove the main results concerning them. In Section 4 these results are applied to the decomposition of non-separable graphs, which has already been introduced. In Section 5 we describe a much deeper application, a decomposition theory for families of sets. Section 6 is devoted to a special case of the set-family theory, a decomposition theory for matroids; it is proved that, for matroids, "prime" is equivalent to "3-connected". In Section 7 we derive a "substitution" decomposition theory for set families from the theory of Section 5.

For the special case in which the set families are "clutters" this topic has been previously investigated in the contexts of Boolean functions and simple games.

The theory presented in this paper appeared in the thesis [10] of the first author, which was supervised by the second author. It was inspired by the work of Louis Billera and has benefited, at an earlier stage, from conversations with Professors Billera, J. A. Bondy and R. E. Bixby. The authors are grateful to an anonymous referee for a careful reading of the manuscript which resulted in several improvements, including a significant simplification of the proof of Theorem 15. This research has been supported by fellowships and grants from the National Research Council of Canada.

2. Decomposition frames: basic properties. In this section we introduce the notion of decomposition frame, and explain its relevance to the graph decomposition material of Section 1. We derive some elementary properties, and prove a rudimentary unique decomposition theorem which is the basis for the deeper results of Section 3.

In absorbing the following definitions, the reader should keep in mind the graph decomposition example of Section 1. Let \mathscr{N} be a class and E be a function defined on \mathscr{N} such that, for each $N \in \mathscr{N}$, E(N) is a finite set, called the set of *cells* of N. Let \rightarrow be a relation associating elements Nof \mathscr{N} to two-element subsets $\{N_1, N_2\}$ of \mathscr{N} , written $N \rightarrow \{N_1, N_2\}$. The triple $(\mathscr{N}, E, \rightarrow)$ is a *decomposition frame* if F1, F2, F3, F4 below are satisfied.

F1. If $N \to \{N_1, N_2\}$, then for some $e \notin E(N)$ and some partition $\{E_1, E_2\}$ of E(N) with $|E_1| \ge 2 \le |E_2|$, we have $E(N_1) = E_1 + e$, $E(N_2) = E_2 + e$.

In the context of F1, $\{N_1, N_2\}$ is called a *simple decomposition* of N, e is called the *marker* of the simple decomposition, and $\{E_1, E_2\}$ is called the *split* of N corresponding to the simple decomposition.

F2. For a split $\{E_1, E_2\}$ of $N \in \mathcal{N}$ and $e \notin E(N)$, there is exactly one simple decomposition $\{N_1, N_2\}$ of N with marker e corresponding to $\{E_1, E_2\}$.

Given a split $\{E_1, E_2\}$ of N and $e \notin E(N)$, we denote by $N(E_i; e), i = 1$ and 2, the unique element of \mathcal{N} such that $E(N(E_i; e)) = E_i + e$ and $N \rightarrow \{N(E_1; e), N(E_2; e)\}.$

F3. Let $\{E_1, E_2\}$ be a split of $N \in \mathcal{N}$, let $A \subset E_1$, and $e \notin E(N)$. Then $\{A, E(N) \setminus A\}$ is a split of N if and only if $\{A, (E_1 + e) \setminus A\}$ is a split of $N(E_1; e)$. F4. Let $\{E_1, E_2\}$, $\{E_3, E_4\}$ be splits of $N \in \mathcal{N}$ such that $E_3 \subset E_1$, and let $e, f \notin E(N), e \neq f$. Then

$$N(E_1; e)(E_3; f) = N(E_3; f)$$
, and
 $N(E_1; e)(E_1 \setminus E_3 + e; f) = N(E_4; f)(E_4 \setminus E_2 + f; e).$

It follows immediately from the definitions and Lemmas 1, 2, 3 of Section 1 that, where \mathcal{N} is the class of finite non-separable graphs, E(G) denotes the edge-set of G, and \rightarrow is as defined in Section 1, $(\mathcal{N}, E, \rightarrow)$ is a decomposition frame. In general, we will refer to the elements of \mathcal{N} as the *objects* of the decomposition frame. We define the terms *decomposition*, (*strict*) *refinement*, *markers* (of a decomposition), *equivalent*, *minimal*, *trivial*, and *prime* for decomposition frames just as they were defined in Section 1. The only differences are the substitution of "cell" for "edge", and "object" for "non-separable graph". The notion of "decomposition tree" also extends to the present context.

A set of splits of an element N of \mathcal{N} is *compatible* if no two members of the set cross. By (F3) and the definition of decomposition, every decomposition D of N gives rise to a set of compatible splits of N, one for each simple refinement in a derivation of N. Given a decomposition D_{i-1} of N, a split $\{A_i', B_i'\}$ of a member N' of D generates a unique simple refinement D_i of D_{i-1} , and any other split of N' which does not cross $\{A_i', B_i'\}$ induces, by (F3), a unique split of a member of D_i . Thus an ordered set $\{\{A_i, B_i\}: 1 \leq i \leq k\}$ of compatible splits of N generates a unique decomposition of N. In fact, as the next result shows, this decomposition does not depend on the order.

THEOREM 2. For $N \in \mathcal{N}$, any set of compatible splits of N generates a unique decomposition of N.

Proof. Let $\{\{A_i, B_i\}: i \in I\}$ be a set of compatible splits of N. We know that for any fixed ordering of I, the decomposition generated is unique. It is true that any ordering of I can be obtained from any other one by a sequence of interchanges of adjacent elements. Thus it suffices to prove that the decomposition D obtained from the ordering $S:i_1, i_2, \ldots, i_{j-1}, i_j, i_{j+1}, i_{j+2}, \ldots, i_k$ of I is the same as the decomposition D' obtained from the ordering $S':i_1, i_2, \ldots, i_{j-1}, i_j, i_{j+1}, i_{j+2}, \ldots, i_k$ of I is the same as the decomposition D' obtained from the ordering $S':i_1, i_2, \ldots, i_{j-1}, i_{j+1}, i_j, i_{j+2}, \ldots, i_k$. Let $D_0 = D_0' = \{N\}$ and let D_m , (D_m') , $1 \leq m \leq k$, denote the decomposition of N generated by the splits $\{A_i, B_i\}$, where i runs through the first m terms of S(S') in that order. Then $D_k = D$ and $D_k' = D'$. Clearly, $D_{j-1} = D_{j-1}'$. If $\{A_j, B_j\}$ and $\{A_{j+1}, B_{j+1}\}$ induce splits in different members of D_{j-1} , then clearly $D_{j+1} = D_{j+1}'$. If they induce splits of the same member of D_{j-1} , then $D_{j+1} = D_{j+1}'$, by (F4). In either case $D_{j+1}' = D_{j+1}$, and so D = D', as required.

In view of the notion of "decomposition tree", with any set of compatible 2-element partitions of a finite set, Theorem 2 associates a tree structure. Essentially the same concept is an important idea of [13], although the two papers are otherwise quite different.

A split of a member N of \mathscr{N} is said to be *good* if it is crossed by no other split of N. Good splits play an extremely important role in this uniqueness theory. Obviously a prime member of \mathscr{N} can have no good split, but there can exist non-primes having no good split. For example, it is easy to see that, with respect to the graph decomposition described in Section 1, bonds and polygons have no good split.

THEOREM 3. Let $N \in \mathcal{N}$. Then N has a unique minimal decomposition, each of whose members has no good split.

LEMMA 5. If $\{E_1, E_2\}$ is a split of N and $\{E_3, E_4\}$ is a good split of N with $E_3 \subset E_1$, then $\{E_3, E_1 \setminus E_3 + e\}$ is a good split of $N(E_1; e)$.

Proof. If the lemma is not true, then there exists a split $\{A, E_1 \setminus A + e\}$ of $N(E_1; e)$ which crosses $\{E_3, E_1 \setminus E_3 + e\}$. Then by (F3), $\{A, E(N) \setminus A\}$ is a split of N, and it is easy to check that it crosses $\{E_3, E_4\}$, contradicting the goodness of $\{E_3, E_4\}$.

LEMMA 6. If $\{E_1, E_2\}$ is a good split of $N \in \mathcal{N}$ and $\{E_3, E_1 \setminus E_3 + e\}$ is a good split of $N(E_1; e)$, then $\{E_3, E(N) \setminus E_3\}$ is a good split of N.

Proof. By (F3), $\{E_3, E(N) \setminus E_3\}$ is a split of N. If it is not a good split, then there exists a split $\{E_5, E_6\}$ of N which crosses it. Now $\{E_5, E_6\}$ cannot cross $\{E_1, E_2\}$, since $\{E_1, E_2\}$ is good, and cannot be equal to $\{E_1, E_2\}$, since $\{E_1, E_2\}$ does not cross $\{E_3, E(N) \setminus E_3\}$. By interchanging E_5 with E_6 if necessary, we may conclude that $E_5 \subset E_1$ or $E_5 \supseteq E_1$. But in the second case, $E_6 \cap E_3 = \emptyset$, which would imply that $\{E_5, E_6\}$ does not cross $\{E_3, E(N) \setminus E_3\}$. Therefore, $E_5 \subset E_1$, so by (F3), $\{E_5, E_1 \setminus E_6 + e\}$ is a split of $N(E_1; e)$. But it follows from the assumption that $\{E_5, E_6\}$ crosses $\{E_3, E(N) \setminus E_3\}$, that $\{E_5, E_1 \setminus E_5 + e\}$ crosses $\{E_3, E_1 \setminus E_3 + e\}$, contradicting the goodness of $\{E_3, E_1 \setminus E_3 + e\}$.

Proof of Theorem 3. Let D be a decomposition of N, and suppose that D is generated by the set S of splits of N. If there is a good split of N which is not in S then, by Lemma 5, there is a member of D having a good split. Therefore, if D has the property that none of its members has a good split, then every good split of N is in S. It follows that every such decomposition D is a refinement of the (unique, by Theorem 2) decomposition D' of N generated by the set of good splits of N. Thus, to prove the theorem, it suffices to show that no member of D' has a good split. But this follows from Lemma 6.

At this point we introduce another example of a decomposition frame. Define a *split system* N to be a finite set E(N) together with a set of partitions (called *splits*) $\{E_1, E_2\}$ of E(N), such that $|E_1| \ge 2 \le |E_2|$. Given a split $\{E_1, E_2\}$ of N and $e \notin E(N)$, define a split system $N(E_i; e)$, for i = 1 and 2, having $E(N(E_i; e)) = E_i + e$ and having as splits precisely the partitions of the form $\{E_3, E_i \setminus E_3 + e\}$ where $\{E_3, E(N) \setminus E_3\}$ is a split of N, and $E_3 \subset E_i$. Then we write

$$N \rightarrow \{N(E_1; e), N(E_2; e)\}.$$

It is easily proved that, where \mathcal{N} is the class of all split systems, $(\mathcal{N}, E, \rightarrow)$ is a decomposition frame. It is, in a sense, the simplest of all decomposition frames, since its members have no structure other than that necessary to satisfy the frame axioms. (This frame does not satisfy the additional properties which are needed for the results of the next section.)

The split system decomposition frame illustrates an interesting aspect of the present theory. Many decomposition theories, such as the theory of prime factorization of integers, are based on possibilities for reversing some uniquely determined composition. Though several instances described here do have this property, others do not, and the split system frame is an obvious example. If $\{N_1, N_2\}$ is a simple decomposition of the split system N associated with the split $\{E_1, E_2\}$ of N, then $\{N_1, N_2\}$ determines the splits of N not crossing $\{E_1, E_2\}$, but does not generally determine the ones that do cross $\{E_1, E_2\}$. Moreover, while the graph decomposition frame of Section 1 does have a unique composition, a slight variant of it does not. Suppose that we define the frame to have as objects equivalence classes of non-separable graphs, where G and G' are equivalent if there is an isomorphism from G to G' which is the identity on E(G). (This is just a precise way of saying that we want to "forget" vertex names.) Now if we have a simple decomposition $\{G_1, G_2\}$ of G with marker $e_1 \{G_1, G_2\}$ is also a simple decomposition of an object G' obtained by "identifying the ends of *e* in the opposite order". Therefore, this new frame does not have a unique composition, but Lemmas 1 to 4 are still true and, as we shall see, these are the results needed to derive Theorem 1. The point that we are making is that a successful decomposition theory can exist in the absence of a uniquely determined composition.

3. Decomposition frames: main theorems. We say that a decomposition frame $(\mathcal{N}, E, \rightarrow)$ has the *intersection property* if, whenever $\{E_1, E_2\}$ and $\{E_3, E_4\}$ are splits of $N \in \mathcal{N}$ such that $|E_1 \cap E_3| \ge 2$ and $E_1 \cup E_3 \ne E(N)$, then $\{E_1 \cap E_3, E_2 \cup E_4\}$ is a split of N. By Lemma 4, the graph decomposition frame of Section 1 has the intersection property. We say that $N \in \mathcal{N}$ is brittle if every partition $\{E_1, E_2\}$ of N such that $|E_1|, |E_2| \ge 2$, is a split of N. We say that $N \in \mathcal{N}$ is semi-brittle if $E(N) = \{e_0, e_1, \ldots, e_{n-1}\}$, and the splits of N are precisely the partitions of E(N) of the form $\{\{e_{i+1}, e_{i+2}, \ldots, e_{i+j}\}, \{e_{i+j+1}, e_{i+j+2}, \ldots, e_i\}$.

where subscripts are modulo n and j, $n - j \ge 2$. We observe that, if N is brittle or semi-brittle, then N has no good split. We also observe that, with respect to the graph decomposition frame, bonds are brittle and polygons are semi-brittle.

THEOREM 4. Let $(\mathcal{N}, E, \rightarrow)$ be a decomposition frame having the intersection property, and let $N \in \mathcal{N}$. Then N has a unique minimal decomposition, each of whose members is prime, brittle, or semi-brittle.

We say that $(\mathcal{N}, E, \rightarrow)$ has the transitivity property if, whenever

 $\{\{e_1, e_2\}, E(N) \setminus \{e_1, e_2\}\}$ and $\{\{e_2, e_3\}, E(N) \setminus \{e_2, e_3\}\}$

are splits of $N \in \mathcal{N}$, then so also is

 $\{\{e_1, e_3\}, E(N) \setminus \{e_1, e_3\}\}.$

The graph decomposition frame does not have the transitivity property. In fact, it is clear that, if a frame has the transitivity property, then every semi-brittle object N has $E(N) \leq 3$, and therefore is prime. Therefore, the following important result is an immediate consequence of Theorem 4.

THEOREM 5. Let $(\mathcal{N}, E, \rightarrow)$ be a decomposition frame having the intersection and transitivity properties, and let $N \in \mathcal{N}$. Then N has a unique minimal decomposition, each of whose members is prime or brittle.

On the other hand, Theorem 4 follows from Theorem 3, together with the following characterization of objects having no good split.

THEOREM 6. Let $(\mathcal{N}, E, \rightarrow)$ be a decomposition frame having the intersection property. An object $N \in \mathcal{N}$ has no good split if and only if N is prime, brittle, or semi-brittle.

LEMMA 7. Let N be an object of a decomposition frame having the intersection property, and suppose that N is not prime and has no good split. Then there exists an ordering $e_0, e_1, \ldots, e_{n-1}$ of E(N) such that, for $0 \le i \le n-1$,

 $\{\{e_i, e_{i+1}\}, E(N) \setminus \{e_i, e_{i+1}\}\}$ is a split of N.

(Subscripts are modulo n).

Proof. The truth of the lemma is easily checked for |E(N)| = 4 (and for |E(N)| < 4). Suppose that it is true for all N such that $|E(N)| \le m \ge 4$, and suppose that we are given N with |E(N)| = m + 1, not prime and having no good split. Then N has a split $\{E_1, E_2\}$. Let N_i denote $N(E_i; e)$ for i = 1 and 2, where $e \notin E(N)$. We claim that N_1 and N_2 have no good split. Let $\{A, E_1 \setminus A + e\}$ be a split of N_1 . Then $\{A, E(N) \setminus A\}$ is a split of N by (F3). There exists a split $\{B, E(N) \setminus B\}$ of N crossing $\{A, E(N) \setminus A\}$. Rename, if necessary, so that $E(N) \setminus B$ meets $E_1 \setminus A$. Case 1. $B \subseteq E_1$. Then $\{B, E_1 \setminus B + e\}$ is a split of N_1 , by (F3). Now $B \cap A \neq \emptyset$, $(E_1 \setminus A + e) \cap (E_1 \setminus B + e) \neq \emptyset$, and $A \cap (E_1 \setminus B + e) \neq \emptyset$. If $B \cap (E_1 \setminus A + e) = \emptyset$, then $B \subseteq A$, a contradiction. Thus $\{A, E_1 \setminus A + e\}$ is not a good split of N_1 .

Case 2. $B \cap E_2 \neq \emptyset$. Then, by the intersection property, $\{E_1 \setminus B, B \cup E_2\}$ is a split of N. Therefore, $\{E_1 \setminus B, E_1 \cap B + e\}$ is a split of N₁, by (F3). Now A meets $E_1 \setminus B$ and $E_1 \cap B + e$, since $\{A, E(N) \setminus A\}$ crosses $\{B, E(N) \setminus B\}$; $E_1 \setminus B$ meets $E_1 \setminus A + e$, since $E(N) \setminus B$ meets $E_1 \setminus A + e$, because e is an element of each. Thus $\{A, E_1 \setminus A + e\}$ is not a good split of N₁.

Therefore, N_1 has no good split; similarly, N_2 has no good split. There exists a split of N crossing $\{E_1, E_2\}$. It follows from the intersection property and (F3) that, for i = 1 and 2, N_i is not prime unless $|E_i| = 2$. From this fact, the induction hypothesis, and (F3), we conclude that there exist orderings c_1, c_2, \ldots, c_p of E_1 and d_1, d_2, \ldots, d_q of E_2 such that

 $\{\{c_i, c_{i+1}\}, E(N) \setminus \{c_i, c_{i+1}\}\}\$ is a split of N for $i = 1, 2 \dots, p - 1$,

and

$$\{\{d_i, d_{i+1}\}, E(N) \setminus \{d_i, d_{i+1}\}\}$$
 is a split of N for $i = 1, 2, ..., q - 1$.

It follows from the intersection property that

$$\{\{c_i, c_{i+1}, \ldots, c_j\}, E(N) \setminus \{c_i, c_{i+1}, \ldots, c_j\}\}$$
 is a split of N

whenever $1 \leq i < j \leq p$,

and similarly for the d_i .

We can choose a split $\{E_3, E_4\}$ of N and the orderings c_1, \ldots, c_p and d_1, \ldots, d_q so that $\{E_3, E_4\}$ crosses $\{E_1, E_2\}$ and $\{c_p, d_1\} \subseteq E_3$. (Proof. If not, there exists such a split $\{E_3', E_4'\}$ with $c_1, c_p \in E_3'$ and $d_1, d_q \in E_4'$. Then $|E_1| > 2$, and so by the intersection property $\{E_1 - c_p, E_2 + c_p\}$ is a split of N. Then

 $\{E_3, E_4\} = \{E_3' \cap (E_2 + c_p), E_4' \cup (E_1 - c_p)\}$

is a split having the required property.) Now, by the intersection property, the partitions $\{E_5, E_6\}$ and $\{E_7, E_8\}$ of E(N), given by

$$E_5 = E_3 \cap (E_1 + d_1)$$
 and $E_7 = E_5 \cap (E_2 + c_p)$,

are splits of N. That is, $\{\{c_p, d_1\}, E(N) \setminus \{c_p, d_1\}\}\$ is a split of N. It is easy to use this fact and the intersection property to show that $\{\{c_1, d_q\}, E(N) \setminus \{c_1, d_q\}\}\$ is also a split of N. Therefore, $c_1, c_2, \ldots, c_p, d_1, \ldots, d_q$ is the required ordering of E(N), and the lemma is proved by induction.

Proof of Theorem 6. Suppose that N has no good split and is neither prime nor semi-brittle. Then there is an ordering $e_0, e_1, \ldots, e_{n-1}$ of E(N) as in Lemma 7, and there exist a split $\{E_1, E_2\}$ of N and integers i, j, k with 0 < i < j < k < n such that $e_0, e_j \in E_1$ and $e_i, e_k \in E_2$.

Applying the intersection property, we obtain a split $\{E_3, E_4\}$ where

$$E_3 = E_1 \cap \{e_0, e_1, \ldots, e_j\};$$

applying it again, we obtain a split $\{E_5, E_6\}$ where

$$E_5 = E_3 \cap \{e_j, e_{j+1}, \ldots, e_{n-1}, e_0\} = \{e_0, e_j\}$$

Now let *m* be an integer such that $2 \leq m \leq j - 1$. By the intersection property, $\{E_7, E_8\}$ is a split of *N*, where

$$E_8 = E_2 \cup \{e_1, e_2, \ldots, e_{j-1}\}.$$

Again, $\{E_9, E_{10}\}$ is a split, where

$$E_9 = E_7 \cup \{e_m, e_{m+1}, \ldots, e_j\}.$$

If we repeat the argument of the first part of the proof, with $\{E_9, E_{10}\}$ replacing $\{E_1, E_2\}$, e_1 replacing e_i , and e_m replacing e_j , we can conclude that $\{\{e_0, e_m\}, E(N) \setminus \{e_0, e_m\}\}$ is a split of N. A similar argument can be applied to the case where $j + 1 \leq m \leq n - 2$. It follows that, for $1 \leq m \leq n - 1$, $\{\{e_0, e_m\}, E(N) \setminus \{e_0, e_m\}\}$ is a split of N. Now let $\{E_1, E_2\}$ be any partition of N such that $|E_1| \geq 2 \leq |E_2|$; we may assume that $e_0 \in E_1$. Using the splits $\{\{e_0, e_j\}, E(N) \setminus \{e_0, e_j\}\}$ for $e_j \in E_1 - e_0$ and applying the intersection property repeatedly, we conclude that $\{E_1, E_2\}$ is a split of N. Therefore, N is brittle, and the proof is complete.

4. Graph decomposition. In this section we apply the theory of the last two sections to prove Theorem 1. We also discuss some other aspects of the graph decomposition theory.

Proof of Theorem 1. It is a consequence of Lemmas 1 to 4 that, where \mathscr{G} is the class of non-separable graphs, E means "edge-set", and \rightarrow is as defined in Section 1, $(\mathcal{G}, E, \rightarrow)$ is a decomposition frame having the intersection property. Now let $G \in \mathscr{G}$ and let $e, f \in E(G)$, where $|E(G)| \ge 4$. Then $\{\{e, f\}, E(G) \setminus \{e, f\}\}$ is a split of G if and only if e and f are in parallel (have the same ends in G) or e and f are in series (are the two edges incident with a vertex of degree 2 in G). Moreover, if e and f are in parallel in G, and f and g are in parallel in G, then e and g are in parallel in G. On the other hand, if e and f are in series in G, and fand g are in series in G, then e and g are not in series in G. Finally, it is not possible for both e and f to be in parallel and f and g to be in series. From these observations we deduce that if G is brittle, every two edges of G are in parallel, so that G is a bond. Also, if G is semi-brittle, then there is an ordering $e_0, e_1, \ldots, e_{n-1}$ of E(G) such e_i is in series with e_{i+1} for $0 \leq i \leq i$ n-1, so that G is a polygon. Therefore, Theorem 1 follows immediately from Theorem 4.

Given a non-separable graph G, let us call the decomposition of G whose uniqueness is asserted in Theorem 1 the *standard* decomposition of G. A consequence of Theorem 1 is that every decomposition of G consisting of 3-connected graphs, polygons, and bonds is a refinement of the standard decomposition D of G. Any such decomposition D' must result from replacing each bond and polygon of D by the members of a decomposition of the bond or polygon. If $D' \neq D$, this means that D' has two polygons or two bonds which share a marker. On the other hand, it is easy to show, using the fact that D is generated by the good splits of G (in any order), that D cannot have two such bonds or polygons. Therefore, we can avoid the "minimal" in Theorem 1 by stating it thus: Every non-separable graph has a unique decomposition consisting of 3-connected graphs, polygons, and bonds with the property that no two bonds and no two polygons share a marker.

According to a notion of 3-connectivity which is perhaps more common than Tutte's, a non-separable graph G is 3-connected if and only if it has no "Whitney split"; that is, there does not exist a partition $\{E_1, E_2\}$ of E(G) such that

 $|W(G; E_1)| = 2$ and $|V(G(E_1))|, |V(G(E_2))| \ge 3$.

Every Whitney split is a Tutte split, and a Tutte split $\{E_1, E_2\}$ is a Whitney split if and only if neither $G(E_1)$ nor $G(E_2)$ is a bond. One might want to develop a decomposition theory for non-separable graphs using this more restrictive notion of split, together with the same notion of simple decomposition. However, this cannot be done in the context of decomposition frames, because (F3) is violated; moreover, Whitney splits do not satisfy the intersection property.

In [23, Chapter 11], W. T. Tutte describes a decomposition of a nonseparable graph G, whose members are called the "cleavage units" of G. Tutte defines this decomposition (uniquely) by restricting simple decompositions to splits satisfying certain additional requirements. He then proves that the resulting decomposition has a number of attractive properties; in particular, its members are 3-connected graphs, polygons, and bonds, and (though Tutte does not state this explicitly) no two bonds and no two polygons share a marker. It follows from these results and Theorem 1, that the cleavage units of G are precisely the members of the standard decomposition of G. Therefore, Tutte's work and our own produce the same canonical decomposition of G, but the theorems are different. While Tutte defines the decomposition and establishes some of its properties, we prove that it is characterized by certain of these properties.

Hopcroft and Tarjan [14], [15] have discovered Theorem 1 independently, and have applied it to extend an algorithm for isomorphism of planar 3-connected graphs to an algorithm for isomorphism of arbitrary planar graphs. (The uniqueness theorem stated in [14] is not correct, but it has been corrected in [15], and a proof appears in an unpublished version of [15]). They also present an algorithm [15] for computing the standard decomposition of a non-separable graph G, for which the amount of computation is bounded by a linear function of |E(G)|.

In yet another approach to the decomposition of non-separable graphs, MacLane [17] uses the following notion of simple decomposition. Let G be a non-separable graph which is not a polygon and let $\{E_1, E_2\}$ be a split of G such that each of $G(E_1)$, $G(E_2)$ contains a circuit. Where $W(G; E_1) = \{u, v\}$, let P_i be a simple path joining u to v in $G(E_{3-i})$ for i = 1 and 2. Let $G_i = G(E_i \cup E(P_i))$ for i = 1 and 2, and call $\{G_1, G_2\}$ a *simple* decomposition of G. As usual, we define a (general) decomposition by iterating the simple decomposition. Clearly, a graph will be prime with respect to this notion of decomposition if and only if it is "nodally 3-connected" [23]. An *atom* of G is a member of a prime decomposition of G which is not homeomorphic to a bond. MacLane's theorem is that the atoms of G are unique up to a homeomorphism which is the identity on vertices of degree at least three. It is clear that the atoms of G are homeomorphic in this way to the set of non-bond, non-polygon members of our standard decomposition of G, and thus that MacLane's theorem is a consequence of Theorem 1.

Finally, we have recently learned, from T. R. S. Walsh, of a theorem of Trakhtenbrot [22] which is closely related to Theorem 1. (The approach taken in [22] is that of "substitution decomposition" as in Section 7.) The paper [25] of Walsh contains a translation of Trakhtenbrot's proof, while in [26], Walsh has used this theorem to derive (independently) Theorem 1.

5. Set-family decomposition. A set family H is a pair (E, \mathscr{F}) , where E is a finite set of cells of H and \mathscr{F} is a set of non-empty subsets of E. A cell e of H such that $\{e\} \in \mathscr{F}$ is called a *loop* of H. A subset A of E is a separator of H if no member of \mathscr{F} meets both A and $E \setminus A$. We say that H is non-separable if its only separators are E and \emptyset . If $A \subseteq E$, then the restriction of H to $E \setminus A$ is $H \setminus A = (E \setminus A, \mathscr{F} \setminus A)$ where $\mathscr{F} \setminus A$ denotes $\{F \in \mathscr{F} : F \subseteq E \setminus A\}$. If $e \in E$, we may abbreviate $H \setminus \{e\}$ to $H \setminus e$ and $\mathscr{F} \setminus \{e\}$ to $\mathscr{F} \setminus e$.

Let \mathscr{H} denote the class of non-separable, loopless set families. (We explain these restrictions later. In the terminology of [2], the members of \mathscr{H} are "finite, simple, loopless, connected hypergraphs".) If $H = (E, \mathscr{F}) \in \mathscr{H}$ and $H_i = (E_i + e, \mathscr{F}_i) \in \mathscr{H}$ for i = 1 and 2, where $\{E_1, E_2\}$ is a partition of E such that $|E_1| \ge 2 \le |E_2|$ and $e \notin E$, we define \rightarrow by $H \rightarrow \{H_1, H_2\}$ if and only if

$$\mathscr{F} = (\mathscr{F}_1 \setminus e) \cup (\mathscr{F}_2 \setminus e) \cup \{F_1 \cup F_2 - e : e \in F_1 \in \mathscr{F}_1, e \in F_2 \in \mathscr{F}_2\};$$

in this situation $\{E_1, E_2\}$ is said to be a *split* of H. For $H \in \mathcal{H}$, let E(H) denote the set of cells of H.

To provide an example of the above simple decomposition, we return to graphs. Given a finite graph G, we can associate with G a set family PM(G), called the *polygon matroid* of G; $PM(G) = (E(G), \mathscr{F})$, where the members of \mathscr{F} are the edge-sets of simple circuits of G. If G has no isolated vertices, then it is a consequence of a result of [28] that PM(G)is a non-separable set family if and only if G is a non-separable graph. Now let G be a non-separable graph having a (graph) split $\{E_1, E_2\}$. It is easy to see that $\{E_1, E_2\}$ is also a split of PM(G), and that

 $PM(G) \rightarrow \{PM(G(E_1; e)), PM(G(E_2; e))\}.$

At this point, it may appear that the set family decomposition directly generalizes the graph decomposition; that this is not the case is demonstrated by two closely-related facts. First, two non-isomorphic non-separable graphs can have the same polygon matroid; second, the polygon matroid of a non-separable graph G can have (set family) splits which are not (graph) splits of G. Nevertheless, there is an extremely close relationship between the set family decomposition theory for PM(G), and the graph decomposition theory for G, and we will return to this in the next section.

We now present the main results of the set family decomposition theory.

THEOREM 7. $(\mathcal{H}, E, \rightarrow)$ is a decomposition frame, which has the intersection and transitivity properties.

THEOREM 8. Each set family $H \in \mathcal{H}$ has a unique minimal decomposition, each of whose members is prime or brittle.

Theorem 8, of course, follows immediately from Theorems 5 and 7. We can also strengthen Theorem 8 by characterizing the brittle set families; they are of a few simple types, which we now describe. Let $H = (E, \mathscr{F})$ be a set family. Then H is a *bond* if

$$\mathscr{F} = \{F \subseteq E \colon |F| = 2\};$$

H is a *star* if

$$\mathscr{F} = \{\{e, e'\} : e \in E - e'\} \text{ for some } e' \in E;$$

H is a k-superstar if

 $\mathscr{F} = \{F \subseteq E : A \subseteq F, |F| \ge 2\}$ for some $A \subseteq E, |A| = k$.

An |E|-superstar is also called a *polygon*. A *superstar* is a set family which is a k-superstar for some k. Theorem 9 below provides a classification of

the brittle set families, and Theorem 10, the main uniqueness theorem for set-family decomposition, is a consequence of Theorems 8 and 9.

THEOREM 9. Set family $H \in \mathscr{H}$ is brittle if and only if H is a bond, a star, or a superstar.

THEOREM 10. Each $H \in \mathscr{H}$ has a unique minimal decomposition, each of whose members is a prime, a bond, a star, or a superstar.

An important special case of the present theory, which in fact was developed earlier, occurs when attention is restricted to *clutters*: set families (E, \mathscr{F}) in which no member of \mathscr{F} contains another. The truth of the following proposition is easy to check, and the resulting theorem (Theorem 11) is a consequence of Theorem 10, when we observe which brittle members of \mathscr{H} are clutters.

PROPOSITION 1. If $H \in \mathscr{H}$ and $H \to \{H_1, H_2\}$, then H is a clutter if and only if H_1 and H_2 are clutters.

THEOREM 11. Each clutter $H \in \mathscr{H}$ has a unique minimal decomposition, each of whose members is a prime, a bond, a star, or a polygon.

If $H_1 = (E_1, \mathscr{F}_1)$ and $H_2 = (E_2, \mathscr{F}_2)$ are set families such that $E_1 \cap E_2 = \emptyset$, then the *direct sum* $H_1 \oplus H_2$ of H_1 and H_2 is the set family $(E_1 \cup E_2, \mathscr{F}_1 \cup \mathscr{F}_2)$. This composition is clearly associative and commutative. The next result is an easy consequence of the definitions.

PROPOSITION 2. The set $A \subseteq E$ is a separator of $H = (E, \mathscr{F})$ if and only if $H = (H \setminus A) \oplus (H \setminus (E \setminus A))$.

It is easy to see that the complement of a separator is a separator, and that the intersection or union of separators is a separator. Thus the minimal non-empty separators (elementary separators) of H partition E. The restrictions of H to its elementary separators are called the components of H; clearly, they are non-separable. A set family $H = (E, \mathscr{F})$ is said to be null if $E = \emptyset$. The following theorem is easy to derive from the above remarks.

THEOREM 12. Each non-null set family H has a unique expression as the direct sum of non-null, non-separable set families.

The existence of the elementary theory of direct sum decomposition justifies the restriction of our theory to non-separable set families. The exclusion of loops is partly explained by observing that, if the marker e were allowed to be a loop in H_1 or H_2 , the result below would not be true.

LEMMA 8. If $\{H_1, H_2\}$ is a simple decomposition of $H \in \mathcal{H}$, where $H_i = (E_i + e, \mathcal{F}_i)$ for i = 1 and 2, then H_i is uniquely determined from H, E_i , and e by the formula:

$$\mathscr{F}_i = (\mathscr{F} \setminus E_{3-i}) \cup \{F \cap E_i + e: F \in \mathscr{F} \text{ meets } E_1 \text{ and } E_2\}.$$

Proof. Suppose that $F_1 \in \mathscr{F}_1$. If $e \notin F_1$, then $F_1 \in \mathscr{F}$, so $F_1 \in \mathscr{F} \setminus E_2$. If $e \in F_1$, then (since H_1 is loopless and non-separable) there exists $F_2 \in \mathscr{F}_2$ with $e \in F_2 \neq \{e\}$. Then $F = F_1 \cup F_2 - e \in \mathscr{F}$, F meets E_1 and E_2 , and $F_1 = F \cap E_1 + e$. Thus

 $\mathscr{F}_1 \subseteq (\mathscr{F} \setminus E_2) \cup \{F \cap E_1 + e: F \in \mathscr{F} \text{ meets } E_1 \text{ and } E_2\}.$

Now let F_1 be a member of $(\mathscr{F} \setminus E_2) \cup \{F \cap E_1 + e: F \in \mathscr{F} \text{ meets } E_1$ and $E_2\}$. If $e \notin F_1$, then $F_1 \in \mathscr{F} \setminus E_2$, so $F_1 \in \mathscr{F}_1$ (since e is not a loop of H_2). If $e \in \mathscr{F}_1$, we may choose $F \in \mathscr{F}$ meeting E_1 and E_2 such that $F_1 = F \cap E_1 + e$. By definition of simple decomposition, F = $F_1' \cup F_2' - e$, where $e \in F_i' \in \mathscr{F}_i$, for i = 1 and 2. Then $F_1 = F_1'$, so $F_1 \in \mathscr{F}_1$.

The following useful characterization of splits of set families is a consequence of Lemma 8 and the definitions.

LEMMA 9. Partition $\{E_1, E_2\}$ of E is a split of $H = (E, \mathscr{F}) \in \mathscr{H}$ if and only if $|E_1| \ge 2 \le |E_2|$, and whenever $F_1, F_2 \in \mathscr{F}$ meet both E_1 and E_2 , then

 $(F_1 \cap E_1) \cup (F_2 \cap E_2) \in \mathscr{F}.$

We have explained why loops are not allowed as marker elements. By Lemma 9, the presence of a loop in H will not affect what simple decompositions H can have. Therefore, for consistency, we do not allow loops at all. (We note that, if H is a non-separable clutter having more than one cell, H is necessarily loopless.) We thus arrive at the triple $(\mathcal{H}, E, \rightarrow)$ described above. We begin the proofs of the theorems by verifying that the decomposition frame axioms are satisfied.

LEMMA 10. If $H \to \{H_1, H_2\}$, where $H = (E, \mathscr{F})$, $H_1 = (E_1 + e, \mathscr{F}_1)$, and $H_2 = (E_2 + e, \mathscr{F}_2)$, and $\{E_3, E_4\}$ is a partition of E such that $E_3 \subset E_1$, then $\{E_3, E_4\}$ is a split of H if and only if $\{E_3, E_1 \setminus E_3 + e\}$ is a split of H_1 .

Proof. We begin by observing that $|E_3| \ge 2 \le |E_4|$ if and only if $|E_3| \ge 2 \le |E_1 \setminus E_3 + e|$. Suppose that $\{E_3, E_1 \setminus E_3 + e\}$ is a split of H_1 , and let $F_1, F_2 \in \mathscr{F}$ meet E_3 and E_4 . Let $F_3 = (F_1 \cap E_3) \cup (F_2 \cap E_4)$. For i = 1 and 2, let $F_i' = F_i$ if $F_i \subseteq E_1$ and let $F_i' = F_i \cap E_1 + e$ otherwise. Then $F_1', F_2' \in \mathscr{F}_1$. By assumption,

$$F_{3}' = (F_{1}' \cap E_{3}) \cup (F_{2}' \cap (E_{1} \setminus E_{3} + e)) \in \mathscr{F}_{1}.$$

If $e \notin F_3'$, then $F_3 = F_3' \in \mathscr{F}$. Otherwise,

$$F_3 = ((F_2 \cap E_2 + e) \cup F_3') - e \in \mathscr{F}.$$

Thus $\{E_3, E_4\}$ is a split of H.

Now suppose that $\{E_3, E_4\}$ is a split of H. For every $F \in \mathscr{F}$ meeting E_3 and E_4 , there exists $F' \in \mathscr{F}_1$ meeting both E_3 and $E_1 \setminus E_3 + e$, such that $F \cap E_3 = F' \cap E_3$. (Choose F' = F if $F \subseteq E_1$, and $F' = F \cap E_1 + e$, otherwise.) Moreover, every $F' \in \mathscr{F}_1$ meeting E_3 and $E_1 \setminus E_3 + e$ arises in this way from some F (there may be several). Let F_1', F_2' be arbitrary members of \mathscr{F}_1 meeting E_3 and $E_1 \setminus E_3 + e$. Then

$$(F_1' \cap E_3) \cup (F_2' \cap (E_1 \setminus E_3 + e))$$

= $(F_1 \cap E_3) \cup (F_2' \cap (E_1 \setminus E_3 + e))$
= $((F_1 \cap E_3) \cup (F_2 \cap E_4))' \in \mathscr{F}_1.$

Thus $\{E_3, E_1 \setminus E_3 + e\}$ is a split of H_1 .

LEMMA 11. Let $\{E_1, E_2\}$, $\{E_3, E_4\}$ be splits of $H \in \mathscr{H}$ with $E_3 \subset E_1$. Then

$$H(E_1; e)(E_3; f) = H(E_3; f) and H(E_1; e)(E_1 \setminus E_3 + e; f) = H(E_4; f)(E_1 \setminus E_3 + f; e).$$

Proof. $\mathscr{F}(E_1; e)$, et cetera, are defined as expected. Now

$$\mathcal{F}(E_1; e)(E_3; f) = \{F: F \in \mathcal{F}(E_1; e), F \subseteq E_3\} \cup \{F \cap E_3 + f: F \in \mathcal{F}(E_1; e), F \in E_3 \} \cup \{F \cap E_3 + f: F \in \mathcal{F}, F \text{ meets } E_1 \text{ and } E_1 \setminus E_3 + e\} = \{F: E_3 \supseteq F \in \mathcal{F}\} \cup \{F \cap E_3 + f: F \in \mathcal{F}, F \text{ meets } E_3 \text{ and } E_1 \setminus E_3 \text{ but not } E_2\} \cup \{((F \cap E_1) + e) \cap E_3) + f: F \in \mathcal{F}, F \text{ meets } E_3 \text{ and } E_2\} = \{F: E_3 \supseteq F \in \mathcal{F}\} \cup \{F \cap E_3 + f: F \in \mathcal{F}, F \text{ meets } E_3 \text{ and } E_4\} = \mathcal{F}(E_3; f).$$

This proves the first part.

Now

$$\mathcal{F}(E_1; e) (E_1 \setminus E_3 + e; f) = \{F: E_1 \setminus E_3 \supseteq F \in \mathcal{F}(E_1; e)\} \cup \{(F \cap (E_1 \setminus E_3 + e)) + f: F \in \mathcal{F}(E_1; e), F \text{ meets } E_3 \text{ and } E_1 \setminus E_3 + e\} \\ = \{F: E_1 \setminus E_3 \supseteq F \in \mathcal{F}\} \cup \{(F \cap (E_1 \setminus E_3)) + e: F \in \mathcal{F}, F \text{ meets } E_1 \setminus E_3 \text{ and } E_2, \text{ but not } E_3\} \cup \{((F \cap (E_1 \setminus E_3)) + f: F \in \mathcal{F}, F \text{ meets } E_1 \setminus E_3 \text{ and } E_3 \text{ but not } E_2\} \cup \{(F \cap (E_1 \setminus E_3)) + f + e: F \in \mathcal{F}, F \text{ meets } E_3 \text{ and } E_2\}.$$

By symmetry, $\mathscr{F}(E_4; f)(E_1 \setminus E_3 + f; e)$ is obtained from the last expression by interchanging E_3 with E_2 , E_4 with E_1 , and e with f. Since this does not change the expression, the result follows.

Combining Lemmas 8, 10, and 11, we obtain the following result.

THEOREM 13. $(\mathcal{H}, E, \rightarrow)$ is a decomposition frame.

To complete the proof of Theorem 7, and thus of Theorem 8, we must prove that $(\mathscr{H}, E, \rightarrow)$ has the intersection and transitivity properties. Neither of these results seems to be easy to prove, although the transitivity proof is fairly straightforward.

THEOREM 14. $(\mathcal{H}, E, \rightarrow)$ has the intersection property.

Proof. Let $\{E_1, E_2\}$ and $\{E_3, E_4\}$ be splits of $H = (E, \mathscr{F}) \in \mathscr{H}$, such that $E_1 \cup E_3 \neq E$, and $|E_1 \cap E_3| \geq 2$. Let F_1 , F_2 be members of \mathscr{F} meeting both $E_1 \cap E_3$ and $E_2 \cup E_4$. We must show that

$$(F_1 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4)) \in \mathscr{F}.$$

Case 1. $F_1 \cup F_2 \not\subseteq E_1 \cup E_3$. Case 1a. $F_2 \cap E_2$, $F_2 \cap E_4 \neq \emptyset$. Then

$$(((F_1 \cap E_1) \cup (F_2 \cap E_2)) \cap E_2) \cup (((F_1 \cap E_3) \cup (F_2 \cap E_4)) \cap E_1) \in \mathscr{F}.$$

But this set is

$$(F_2 \cap E_2) \cup (F_1 \cap E_1 \cap E_3) \cup (F_2 \cap E_1 \cap E_4) = (F_1 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4)),$$

as required.

Case 1b. $F_2 \cap E_4 = \emptyset$. Then $F_1 \cap E_2$, $F_2 \cap E_2 \neq \emptyset$, since F_1 and F_2 meet $E_2 \cup E_4$ and $F_1 \cup F_2$ meets $E_2 \cap E_4$.

Case 1b(i). $F_1 \cap E_1 \cap E_4 = \emptyset$. Then

$$(F_1 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4))$$

= $(F_1 \cap E_1 \cap E_3) \cup (F_2 \cap E_2)$
= $(F_1 \cap E_1) \cup (F_2 \cap E_2) \in \mathscr{F}.$

Case 1b(ii). $F_1 \cap E_1 \cap E_4 \neq \emptyset$. Then $F_3 = (F_1 \cap E_2) \cup (F_2 \cap E_1) \in \mathscr{F}$, and F_3 meets E_3, E_4 . Thus $F_4 = (F_1 \cap E_3) \cup (F_3 \cap E_4) \in \mathscr{F}$, and we show that F_4 meets E_2 as follows. One of F_1, F_2 meets $E_2 \cap E_4$, but $F_2 \cap E_4 = \emptyset$, so $F_1 \cap E_2 \cap E_4 \neq \emptyset$. Therefore, the set $F_5 = (F_4 \cap E_1) \cup (F_2 \cap E_2) \in \mathscr{F}$. But

$$F_{5} = (F_{1} \cap E_{1} \cap E_{3}) \cup (F_{3} \cap E_{1} \cap E_{4}) \cup (F_{2} \cap E_{2})$$

= $(F_{1} \cap E_{1} \cap E_{3}) \cup (F_{2} \cap E_{1} \cap E_{4}) \cup (F_{2} \cap E_{2})$
= $(F_{1} \cap E_{1} \cap E_{3}) \cup (F_{2} \cap (E_{2} \cup E_{4})),$

as required.

Case 1c. $F_2 \cap E_2 = \emptyset$. This case is similar to Case 1b. Case 2. $F_1, F_2 \subseteq E_1 \cup E_3$.

Case 2a. There exists $F_3 \in \mathscr{F}$ meeting both $E_1 \cap E_3$ and $E_2 \cap E_4$. Then we define

$$F_4 = \begin{cases} (F_1 \cap E_1) \cup (F_3 \cap E_2), & \text{if } F_1 \cap E_2 \neq \emptyset; \\ (F_1 \cap E_3) \cup (F_3 \cap E_4), & \text{otherwise.} \end{cases}$$

Then $F_4 \in \mathscr{F}$ and $F_4 \cap E_1 \cap E_3 = F_1 \cap E_1 \cap E_3$. Also,

$$(F_4 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4)) \in \mathscr{F}$$

since $F_4 \not\subseteq E_1 \cup E_3$ and Case 1 has been proved. Thus

 $(F_1 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4)) \in \mathscr{F}.$

Case 2b. No member of \mathscr{F} meets both $E_1 \cap E_3$ and $E_2 \cap E_4$. Then, since H is non-separable, there exists $F_3 \in \mathscr{F}$ meeting both $E_1 \cup E_3$ and $E_2 \cap E_4$.

Case 2b(i). F_3 meets $E_1 \cap E_4$. If $F_4 \in \mathscr{F}$ meets both $E_1 \cap E_3$ and $E_2 \cap E_3$, then $(F_3 \cap E_2) \cup (F_4 \cap E_1) \in \mathscr{F}$ meets both $E_1 \cap E_3$ and $E_2 \cap E_4$, a contradiction. Thus every member of \mathscr{F} meeting $E_1 \cap E_3$ is contained in E_1 . But then

$$(F_1 \cap E_1 \cap E_3) \cup (F_2 \cap (E_2 \cup E_4)) \\ = (F_1 \cap E_3) \cup (F_2 \cap E_4) \in \mathscr{F}.$$

Case 2b(ii). F_3 meets $E_2 \cap E_3$. This case is similar to Case 2b(i). The proof is complete.

THEOREM 15. $(\mathcal{H}, E, \rightarrow)$ has the transitivity property.

Proof. Suppose that $\{\{x, y\}, E \setminus \{x, y\}\}$ and $\{\{y, z\}, E \setminus \{y, z\}\}$ are splits of $H = (E, \mathscr{F}) \in \mathscr{H}$ for distinct cells x, y, z of H. We must show that $\{\{x, z\}, E \setminus \{x, z\}\}$ is a split of H. If |E| > 4, then $\{\{x, y, z\}, E \setminus \{x, y, z\}\}$ is a split of H by Theorem 14. Thus $H \rightarrow \{H_1, H_2\}$, where $H_1 =$ $(\{x, y, z, w\}, \mathscr{F}_1)$ and by Proposition 5, $\{\{x, y\}, \{z, w\}\}$ and $\{\{y, z\},$ $\{x, w\}\}$ are splits of H_1 , and $\{\{x, z\}, E \setminus \{x, z\}\}$ is a split of H if and only if $\{\{x, z\}, \{y, w\}\}$ is a split of H_1 . Therefore, it is sufficient to prove the theorem in the case where |E| = 4, say $E = \{x, y, z, w\}$.

Now let F_1 , F_2 be arbitrary members of \mathscr{F} meeting $\{x, z\}$ and $\{y, w\}$. We must show that

$$F = (F_1 \cap \{x, z\}) \cup (F_2 \cap \{y, w\}) \in \mathscr{F}.$$

If $F_1 \cap \{y, w\} = F_2 \cap \{y, w\}$, then $F = F_1 \in \mathscr{F}$. Similarly, if $F_1 \cap \{x, z\} = F_2 \cap \{x, z\}$, there is no problem. Also, for example, the case $y \in F_1$, $w \notin F_1$ is symmetrical to the case $y \notin F_1$, $w \in F_1$. By such considerations, we can assume that $x, y \in F_1$ and $z, w \in F_2$. For i = 1, 2 and c = x, y, z, w, let $F_i(c) = \{c\}$ if $c \in F_i$, and \emptyset otherwise. Thus

$$F_1 = \{x, y\} \cup F_1(z) \cup F_1(w); F_2 = F_2(x) \cup F_2(y) \cup \{z, w\}.$$

 $\{\{y, z\}, \{x, w\}\}$ is a split, and so

$$(F_1 \cap \{y, z\}) \cup (F_2 \cap \{x, w\}) \in \mathscr{F} ext{ and}$$

 $(F_2 \cap \{y, z\} \cup (F_1 \cap \{x, w\}) \in \mathscr{F};$

that is

$$F_3 = \{y, w\} \cup F_1(z) \cup F_2(x) \in \mathscr{F} ext{ and } F_4 = \{x, z\} \cup F_1(w) \cup F_2(y) \in \mathscr{F}.$$

But $\{\{x, y\}, \{z, w\}\}$ is a split, and so $(F_4 \cap \{x, y\}) \cup (F_3 \cap \{z, w\}) \in \mathscr{F}$; that is,

$$\{x, w\} \cup F_1(z) \cup F_2(y) \in \mathscr{F}.$$

But this is $(F_1 \cap \{x, z\}) \cup (F_2 \cap \{y, w\})$, as required.

We have completed the proof of Theorem 7. We now prove Theorem 9, and thus Theorems 10 and 11. A set family $H = (E, \mathscr{F})$ is a *throng* if, whenever $E \supseteq F_2 \supseteq F_1 \in \mathscr{F}$, then $F_2 \in \mathscr{F}$.

Proof of Theorem 9. It is straightforward to verify that bonds, stars, and superstars are brittle and non-separable. Now suppose that $H = (E, \mathscr{F}) \in \mathscr{H}$ is brittle. We first prove:

Claim 1. If $F_1 \subseteq F_2 \subseteq F_3$ and $F_1, F_3 \in \mathscr{F}$, then $F_2 \in \mathscr{F}$.

Choose $x \in F_1$, and put $E_1 = F_2 - x$, $E_2 = E \setminus E_1$. Then $|E_1| \ge 2 \le |E_2|$, and so $\{E_1, E_2\}$ is a split. But F_1 , F_3 meet E_1 , E_2 , and so

 $(F_3 \cap E_1) \cup (F_1 \cap E_2) = F_2 \in \mathscr{F},$

proving Claim 1. We now prove:

Claim 2. H is a throng or a bond or a star.

Let F_1 be a maximal member of \mathscr{F} . If H is not a throng, then by Claim 1 $F_1 \neq E$, and there exists a maximal member F_2 of \mathscr{F} distinct from F_1 . By the non-separability of H, we may choose F_2 so that F_1 meets F_2 . Suppose that $|F_1 \cap F_2| \geq 2$, and choose $x \in F_1 \cap F_2$. Let $E_1 =$ $(F_1 \setminus F_2) + x$, and let $E_2 = E \setminus E_1$. Then F_1 , F_2 meet E_1 , E_2 so

 $(F_1 \cap E_1) \cup (F_2 \cap E_2) = F_1 \cup F_2 \in \mathscr{F},$

which contradicts the maximality of F_1 and F_2 . Thus we may assume that $F_1 \cap F_2 = \{x\}$. Now suppose that $|F_1| \ge 3$. Choose $y \in F_1 - x$ and let $E_1 = \{x, y\}$ and $E_2 = E \setminus E_1$. Then F_1 , F_2 meet E_1 , E_2 so

 $(F_1 \cap E_1) \cup (F_2 \cap E_2) = F_2 + y \in \mathscr{F},$

contradicting the maximality of F_2 . Thus $|F_1| = 2$. It follows that all maximal members of \mathscr{F} , and thus all members of \mathscr{F} , have cardinality 2. Now we may assume that E is the vertex-set of a simple, connected graph G and that \mathscr{F} consists of adjacent pairs of elements of E. Suppose that, for distinct elements x, y, z, w of E, we have $\{w, x\}, \{x, y\}, \{y, z\} \in \mathscr{F}$. Then $\{w, x\}, \{y, z\}$ meet $\{w, z\}, E \setminus \{w, z\}$, so

$$(\{w, x\} \cap \{w, z\}) \cup (\{y, z\} \cap E \setminus \{w, z\})) = \{w, y\} \in \mathscr{F}.$$

It follows that vertices of G are at distance 2 only if they have degree 1. Since G is connected, no distances greater than 2 can occur. Thus every vertex of G has degree 1 or is adjacent to every other vertex. If all the vertices are of the second type, G is complete and so H is a bond. If there exist vertices of degree 1, then there can be only one vertex which is adjacent to all of them. Thus H is a star. This completes the proof of Claim 2.

Let \mathscr{C} be the set of minimal members of \mathscr{F} . We now prove:

Claim 3. If H is a throng, then where $A = \bigcup \{C: C \in \mathcal{C}\}, |A| \ge 2$ and $H' = (A, \mathcal{C})$ is a non-separable brittle clutter.

Choose $E_1 \subseteq E$ and suppose that $C_1, C_2 \in \mathscr{C}$ meet E_1 and $E_2 = E \setminus E_1$. Then

$$F = (C_1 \cap E_1) \cup (C_2 \cap E_2) \in \mathscr{F}.$$

Suppose $F \notin \mathscr{C}$. Then since H is a throng, there exists $x \in F$ such that $F - x \in \mathscr{F}$. Assume $x \in E_1$. (The case $x \in E_2$ is similar.) If $C_1 \cap E_1 = \{x\}$, then F - x is a proper subset of C_2 , which is not possible. We conclude that F - x meets E_1 and E_2 . Thus

 $F_1 = ((F - x) \cap E_1) \cup (C_1 \cap E_2) \in \mathscr{F}.$

But this is a contradiction, since F_1 is a proper subset of C_1 . Thus $F \in \mathscr{C}$. Now suppose there exist distinct elementary separators S_1 , S_2 of (E, \mathscr{C}) with $|S_1| \ge 2 \le |S_2|$. Choose $D_1 \subseteq E$ such that S_1 and S_2 meet both D_1 and $D_2 = E \setminus D_1$, and choose $C_i \in \mathscr{C}$ such that $C_i \subseteq S_i$ and meets D_1, D_2 for i = 1 and 2. Then

 $C = (C_1 \cap D_1) \cup (C_2 \cap D_2) \in \mathscr{C}.$

But *C* meets S_1 , S_2 which contradicts the fact that S_1 , S_2 are elementary separators of (E, \mathscr{C}) . Thus there is at most one elementary separator of (E, \mathscr{C}) having cardinality greater than 1. Also, unless $|E| \leq 1$, when the theorem is obvious, there is one such elementary separator, say *A*. Since *H* has no loops, $\mathscr{C} = \mathscr{C} \setminus (E \setminus A)$. Thus $H' = (A, \mathscr{C})$ is a non-separable brittle clutter, and Claim 3 is proved.

Now H' must be a bond or a star or a polygon by Claim 2. If H' is a bond with A = E, then H is a 0-superstar. If H' is a star with A = E, then H is a 1-superstar. If H' is a polygon and $|A| = k \ge 2$, then H is a *k*-superstar. Thus we must eliminate the possibility that H' is a star or a bond (not a polygon) with $A \ne E$. If this happens, we have $|A| \ge 3$, since a bond or star H' with |A| = 2 is also a polygon. There exist distinct elements x, y, z of A with $\{x, y\}, \{y, z\} \in \mathscr{C}$. Choose $w \in E \setminus A$ and $E_1 \subseteq E$ such that $x, y \in E_1$ and $w, z \in E_2 = E \setminus E_1$. By the definition of

 $\mathscr{C}, \{x, y, w\} \in \mathscr{F}.$ Also $\{x, y, w\}, \{y, z\}$ meet E_1, E_2 so $(\{x, y, w\} \cap E_2) \cup (\{y, z\} \cap E_1) = \{y, w\} \in \mathscr{F}.$

But this is a contradiction, since $\{y, w\}$ does not contain a member of \mathscr{C} . The proof of the theorem is complete.

The proof of the above theorem, revealing as it does the close connection between a brittle throng and its clutter of minimal members, might lead one to suspect that the clutter of minimal members of an arbitrary set family would yield a great deal of information about the possible decompositions of the set family. This does not seem to be the case, and indeed the authors believe that the decomposition theory for arbitrary set families cannot easily be derived from the corresponding clutter theory.

6. Matroid decomposition. A set family $M = (E, \mathscr{F})$ is said to be a *matroid* if \mathscr{F} is a clutter, and the set

 $\mathscr{I} = \{J \subseteq E: J \text{ contains no member of } \mathscr{F}\}$

(called the set of *independent* sets of M) satisfies:

I1. For all $A \subseteq E$, any two maximal independent subsets of A have the same cardinality.

A maximal independent subset of A is called a *basis* of A, and its cardinality (which depends only on A) is called the *rank*, r(A), of A. Since we often deal with more than one matroid, we occasionally prefix terms by the name of the appropriate matroid; for example, "*M*-independent." Similarly, the rank function of matroid M (or M_i , or M') will be denoted by r (or r_i , or r'). The set family which we have associated with the matroid M has as its members the minimal dependent (non-independent) sets of M, called the *circuits* of M. For this special case, the set-family composition was formulated by Minty [18]. We will define any matroid terminology used here, but we occasionally use well-known, elementary facts without proving them; a good matroid-theory reference is [27].

The notion of separability which we have used for set families is well known for matroids. A standard result concerning it says that $A \subseteq E$ is a separator of M if and only if $r(A) + r(E \setminus A) = r(E)$. Tutte [24] has defined a partition $\{E_1, E_2\}$ of M to be a *k*-separation of M, for k a positive integer, if $|E_1| \ge k \le |E_2|$ and $r(E_1) + r(E_2) \le r(E) + k - 1$. For n a positive integer, the matroid M is said to be *n*-connected if M has no *k*-separation for any positive integer k < n. Thus, in particular, every matroid is 1-connected, and a matroid $M \in \mathcal{H}$ is prime if and only if it is non-separable. We show that a matroid $M \in \mathcal{H}$ is prime if and only if it

754

is 3-connected. This result has also been obtained in [6] and [19]; related work is that of [8] and [21].

THEOREM 16. Let $M \in \mathcal{H}$ be a matroid, and let $\{E_1, E_2\}$ be a partition of E = E(M). Then $\{E_1, E_2\}$ is a split of M if and only if $\{E_1, E_2\}$ is a 2-separation of M.

LEMMA 12. Let $\{E_1, E_2\}$ be a 2-separation of the non-separable matroid M and let B_1 be a basis of E_1 . Then B_1 contains a set D such that every circuit C having $\emptyset \neq C \cap B_1 = C \cap E_1$ satisfies $C \cap B_1 = D$.

Proof. Where B_2 is a basis of E_2 , $B_1 \cup B_2$ contains a unique circuit C, because $B_1 \cup B_2 = B + x$ for some basis B of E and $x \notin B$. It will be enough to show that, for every choice of B_2 , $C \cap B_1$ is the same set D, since every circuit of the kind described in the lemma arises in this way. Moreover, since $b \in B_2 \setminus B_2'$ for bases B_2, B_2' of E_2 implies that there exists $b' \in B_2' \setminus B_2$ such that $B_2 - b + b'$ is also a basis of E_2 , it will be enough to consider bases B_2 and $B_2' = B_2 - b + b'$ of E_2 . Let C, C' be the circuit contained in $B_1 \cup B_2$, $B_1 \cup B_2'$, respectively. There is a circuit $F \subseteq B_2 + b'$, and $b' \in F$. For any $x \in (C' \cap B_1) \setminus C$, there exists a circuit G such that $x \in G \subseteq (C' \cup F) - b'$. But $G \subseteq B_1 \cup B_2$, so G = C; that is, $C \supseteq C' \cap B_1$. Similarly, $C' \supseteq C \cap B_1$, and the proof is finished.

Proof of Theorem 16. Suppose that $\{E_1, E_2\}$ is a split of M. Let B_1 be a basis of E_1 . Extend B_1 to a basis B of E, and extend $B \cap E_2$ to a basis B_2 of E_2 . We must show that $|B_2 \setminus B| = 1$. If not, then there exist distinct elements x, y of $B_2 \setminus B$. There exist circuits C_x, C_y such that $x \in C_x \subseteq B + x, y \in C_y \subseteq B + y$, and both C_x and C_y must meet E_1 . Then

$$C_x' = (C_x \cap E_1) \cup (C_y \cap E_2)$$

is a circuit and there exists $z \in C_x' \cap C_y$. Thus there exists a circuit C such that

$$x \in C \subseteq (C_x' \cup C_y) - z.$$

Now C must meet E_1 , since otherwise $C \subseteq B_2$, but then

$$C' = (C_x' \cap E_1) \cup (C \cap E_2)$$

is a circuit properly contained in C_x' , a contradiction. Therefore, $\{E_1, E_2\}$ is a 2-separation of M.

Now suppose that $\{E_1, E_2\}$ is a 2-separation of M, and let C, C' be circuits of M meeting E_1 and E_2 . Extend $C \cap E_1$ to a basis B_1 of E_1 , and extend $C' \cap E_2$ to a basis B_2 of E_2 . It follows from Lemma 12 that every circuit F meeting E_1 and E_2 such that $F \cap E_1 \subseteq B_1$ satisfies $F \cap E_1 = C \cap E_1$; similarly, if $F \cap E_2 \subseteq B_2$, then $F \cap E_2 = C' \cap E_2$. But $B_1 \cup B_2$ contains such a circuit F, so $F = (C \cap E_1) \cup (C' \cap E_2)$.

The decomposition (and composition) for matroids can be related to certain standard matroid constructions, as follows. Given a matroid $M = (E, \mathscr{F})$, the set family $M \setminus A$ is also a matroid, obtained from M by deleting A. The set family $M \setminus A$ can also be proved to be a matroid, obtained from M by contracting A; $M \setminus A$ is defined to have as circuits the minimal sets $C \subseteq E \setminus A$ such that $C \cup A$ contains an M-circuit meeting C. Where $e \in E$, we will abbreviate $M \setminus \{e\}$ to $M \setminus e$ and $M \setminus \{e\}$ to M/e. A set $A \subseteq E$ is a series set of M if every M-circuit C satisfies $C \supseteq A$ or $C \cap A = \emptyset$. A series contraction is a contraction of a proper subset of a series set. A series minor of M is a matroid obtained from Mby a deletion followed by a sequence of series contractions. Finally, the sum [11], $M_1 + M_2$, of matroids M_1 and M_2 is a matroid satisfying $E(M_1 + M_2) = E(M_1) \cup E(M_2)$, a set being $(M_1 + M_2)$ -independent if and only if it can be expressed as a union of an M_1 -independent set with an M_2 -independent set. (Note that $E(M_1), E(M_2)$ need not be disjoint.)

THEOREM 17. Let $M = (E, \mathscr{F}) \in \mathscr{H}$ be a matroid, and let $\{M_1, M_2\}$ be the simple decomposition of M associated with the split $\{E_1, E_2\}$ of M and the marker e. Then M_1 and M_2 are isomorphic to series minors of M (and therefore are matroids). Moreover, $M = (M_1 + M_2)/e$.

Proof. Since $M \in \mathcal{H}$, we can choose an *M*-circuit *C* meeting E_1 and E_2 . Let $D = E_2 \cap C$, and choose $x \in D$. Let

 $M_1' = (M \setminus (E_2 \setminus D)) / (D - x).$

It is easy to see that D is a series set of $M \setminus (E_2 \setminus D)$, and thus M_1' is a series minor of M. We claim that M_1 can be obtained from M_1' by replacing x by e. To prove this, it will be enough to show that the set of circuits of M_1' is

 $\{C \subseteq E_1: C \text{ an } M\text{-circuit}\} \cup \{C \cap E_1 + x:$

C an M-circuit meeting E_1 and E_2 .

But this follows from the fact that D is a series set of $M \setminus (E_2 \setminus D)$. Therefore, M_1 is isomorphic to a series minor of M, and similarly for M_2 .

To prove the second part of the theorem, observe that the set of $(M_1 + M_2)$ -circuits is

 $\{C \subseteq E_1: C \text{ an } M_1\text{-circuit}\} \cup \{C \subseteq E_2: C \text{ an } M_2\text{-circuit}\}$

 $\cup \{C_1 \cup C_2: e \in C_1 \cap C_2, C_i \text{ an } M_i \text{-circuit for } i = 1 \text{ and } 2\}.$

It follows that the circuits of $(M_1 + M_2)/e$ are precisely the circuits of M.

The only brittle set families which are matroids are the bonds and polygons. Therefore, we derive from Theorem 10 the following unique decomposition theorem for matroids. THEOREM 18. Every non-separable matroid has a unique minimal decomposition, each of whose members are bonds, polygons, or 3-connected matroids.

We point out that the proof we have given for Theorem 18 is far from being the shortest possible. In particular, Theorem 18 can be derived quite simply from Theorem 5, since the intersection and transitivity properties are extremely easy to prove for matroids, as is a characterization of the brittle matroids. In fact, even the work of Section 3 can be avoided; in [10], Theorem 18 is derived from Theorem 3. A main tool in that derivation is a matroid-theoretic characterization of the good splits of a matroid, which we give here without proof.

THEOREM 19. Let M be a non-separable matroid. The 2-separation $\{E_1, E_2\}$ of M is a good split of M if and only if at least one of $M \setminus E_1$, $M \setminus E_2$ is non-separable, and at least one of M/E_1 , M/E_2 is non-separable.

As before, let us call the decomposition whose uniqueness is asserted in Theorem 18, the *standard* decomposition of M. The next result (whose proof we omit; see [10]) extends some basic connectivity results of Tutte. He proves [24] that, where e is a cell of a non-separable matroid M, at least one of $M \setminus e$, M/e is non-separable; moreover, if M is 3-connected, and has at least 4 cells, both $M \setminus e$ and M/e are non-separable.

THEOREM 20. Let e be a cell of a non-separable matroid M having at least 3 cells, and let M' be the member of the standard decomposition of M which has e as a cell. Then

(a) $M \setminus e$ is separable if and only if M' is a polygon;

(b) M/e is separable if and only if M' is a bond.

The fact that the members of any decomposition of a matroid M are series minors of M implies that many important classes of matroids are "closed under decomposition". These include polygon matroids, matroids linear over a given field, and transversal matroids. (Similarly, many well-known classes are "closed under composition".) In the case of polygon matroids, Tutte [24] has shown that a connected graph G is 3-connected if and only if PM(G) is a 3-connected matroid. Therefore, the polygon matroids of the members of the standard (graph) decomposition D of G constitute a (matroid) decomposition D' of PM(G), each of whose members is a bond, a polygon, or a 3-connected matroid, and D' is minimal with this property. Therefore, D' is the standard decomposition of PM(G). In spite of this close relationship between the two standard decompositions, Theorem 1 applied to G and Theorem 18 applied to PM(G) are quite different uniqueness results.

Perhaps the most important aspect of matroid decomposition as a special case of the set-family theory is that matroids constitute the only identifiable class of set families for which efficient methods of constructing the standard decomposition are known. These methods are based on algorithms for finding matroid k-separations. In [10] a particularly simple and efficient algorithm for separability was described; this method, for the case of binary matroids, is implicit in the papers of Tutte. A good algorithm for finding a k-separation, if one exists, for any fixed k, based on the matroid partition [11] (or matroid intersection [12]) algorithm, was described in [10]. More recently, an efficient recursive algorithm for testing for 3-connectivity was discovered [7].

The expression for M in terms of M_1 and M_2 in Theorem 17 suggests a more general composition for matroids: forming the matroid $(M_1 + M_2)/(E_1 \cap E_2)$, where M_i is on E_i for i = 1 and 2. This composition is investigated in [10] (and also in [19]); many, but not all, of the attractive properties of the cases $|E_1 \cap E_2| = 0$ or 1 extend to the more general composition. Another topic of investigation [10] has been a decomposition theory for systems of homogeneous linear equations, for which "k-decomposability" is equivalent to (k + 1)-separability of the associated matroid. In particular, there is a unique decomposition theory for the case in which simple decompositions are pairs of systems having exactly one common variable, and this theory is closely related to the matroid decomposition theory for the matroid associated with a linear system. We do not treat the linear system decomposition here, partly because it does not satisfy the decomposition frame axioms. The linear system theory, the generalized matroid decomposition, and the matroid connectivity algorithms, will be described in another paper.

7. The substitution decomposition. In this section we investigate a notion of decomposition, "substitution decomposition", which has been studied by several other authors. The theory associated with this decomposition is shown to be a special case of the clutter decomposition theory of Section 5. By applying the general set family theory, we generalize the substitution decomposition theory for clutters to arbitrary set families.

Let $H_1 = (E_1 + e, \mathscr{F}_1)$ and $H_2 = (E_2, \mathscr{F}_2)$ be set families such that $e \notin E, \{E_1, E_2\}$ is a partition of E, and $|E_1| \ge 1, |E_2| \ge 2$. We define $H_1[H_2; e]$ to be the set family $H = (E, \mathscr{F})$, where

$$\mathscr{F} = (\mathscr{F} \setminus e) \cup \{F_1 \cup F_2 - e : e \in F_1 \in \mathscr{F}_1, F_2 \in \mathscr{F}_2\}.$$

We call $H_1[H_2; e]$ a substitution composition. Despite certain similarities between this composition and the set family composition described in Section 5, there are the following obvious differences. Here the two sets of cells are disjoint, the composition is not commutative, and the composition depends on the choice of a "special" cell e of H_1 .

The substitution composition for the special case in which the set families are clutters has been studied previously in several contexts: Boolean functions ([1], [5]), simple games ([20]), and clutters ([4]). As an illustration of the substitution composition, we briefly outline the connection with Boolean functions. Let $E = \{1, 2, ..., n\}$. A function ffrom $\{0, 1\}^{E}$ to $\{0, 1\}$ is monotone if

$$f(x_1, x_2, \dots, x_n) \leq f(y_1, y_2, \dots, y_n) \text{ whenever } x_i \leq y_i$$

for $i = 1, 2, \dots, n$.

We assume that $f(0, 0, \ldots, 0) = 0$. If $A \subseteq E$, $x^A \in \{0, 1\}^E$ denotes $(x_i: i \in E)$ where $x_i = 1$ if and only if $i \in A$. Associated with the monotone function f is the clutter $H = (E, \mathscr{F})$ defined by: $F \in \mathscr{F}$ if and only if F is minimal such that $f(x^F) = 1$. It is clear that f is also recoverable from H. In these circumstances we denote H by c(f). The following result, whose proof is straightforward, can be found in [3].

PROPOSITION 3. Let $E_1 = \{1, 2, ..., k\}$ and $E_2 = \{k + 1, ..., n\}$, where $k \ge 1$ and $n - k \ge 2$. Let e be an element not in E. For monotone functions g from $\{0, 1\}^{E_1+e}$ to $\{0, 1\}$ and h from $\{0, 1\}^{E_2}$ to $\{0, 1\}$, we have

$$f(x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_k, h(x_{k+1}, \ldots, x_n))$$

for every $(x_1, \ldots, x_n) \in \{0, 1\}^E$ if and only if c(f) = c(g)[c(h); e].

If $\{E_1, E_2\}$ is a partition of E, we say that E_2 is a *committee* of $H = (E, \mathscr{F})$ if $|E_1| \ge 1$, $|E_2| \ge 2$ and, whenever $F_1, F_2 \in \mathscr{F}$ both meet E_2 ,

 $(F_1 \cap E_1) \cup (F_2 \cap E_2) \in \mathscr{F}.$

A cell e of H is an *isthmus* of H if there exists no member F of \mathscr{F} containing e.

PROPOSITION 4. Let $H = (E, \mathscr{F})$ be a set family and let $\{E_1, E_2\}$ be a partition of E with $|E_1| \ge 1$, $|E_2| \ge 2$. There exist set families $H_1 = (E_1 + e, \mathscr{F}_1)$ and $H_2 = (E_2, \mathscr{F}_2)$, where $e \notin E$, such that $H = H_1[H_2; e]$ if and only if E_2 is a committee of H. Moreover, if H has no isthmus, then H_1 and H_2 are uniquely determined, by

$$\mathscr{F}_1 = (\mathscr{F} \setminus E_2) \cup \{F \cap E_1 + e: F \in \mathscr{F} \text{ meets } E_2\} \text{ and}$$

 $\mathscr{F}_2 = \{F \cap E_2: F \in \mathscr{F} \text{ meets } E_2\},$

and neither H_1 nor H_2 has an isthmus.

Proof. It is clear from the definitions that, if $H = H_1[H_2; e]$, then E_2 is a committee of H.

If E_2 is a committee of H, define \mathscr{F}_1 to be

$$(\mathscr{F} \setminus E_2) \cup \{F \cap E_1 + e: F \in \mathscr{F}, F \text{ meets } E_2\}$$

and \mathcal{F}_2 to be

 $\{F \cap E_2: F \in \mathscr{F}, F \text{ meets } E_2\}.$

Then it is straightforward to check that, if $H_1 = (E_1 + e, \mathscr{F}_1)$ and $H_2 = (E_2, \mathscr{F}_2)$, we have $H = H_1[H_2; e]$.

Now suppose that E_2 is a committee of H and $H = H_1'[H_2'; e]$, where $H_1' = (E_1 + e, \mathscr{F}_1')$ and $H_2' = (E_2, \mathscr{F}_2')$, and suppose that H has no isthmus. Then

 $\{F \in \mathscr{F}: F \text{ meets } E_2\} \neq \emptyset.$

Thus $\mathcal{F}_{2}' \neq \emptyset$.

Let $F_1' \in \mathscr{F}_1'$. If $e \notin F_1'$, then $F_1' \in \mathscr{F} \setminus E_2$, so $F_1' \in \mathscr{F}_1$. If $e \in F_1'$, choose $F_2' \in \mathscr{F}_2'$. Then $F = F_1' \cup F_2' - e \in \mathscr{F}$, and $F_1' = F \cap E_1 + e$, so $F_1' \in \mathscr{F}_1$. Thus $\mathscr{F}_1' \subseteq \mathscr{F}_1$.

Now suppose $F_1 \in \mathscr{F}_1'$. If $e \notin F_1'$, then $F_1 \in \mathscr{F} \setminus E_2$. Thus $F_1 \in \mathscr{F}_1' \setminus e$, so $F_1 \in \mathscr{F}_1'$. If $e \in F_1$, then $F_1 = F \cap E_1 + e$ for some $F \in \mathscr{F}$ such that F meets E_2 . Then $F = F_1' \cup F_2' - e$, where $e \in F_1' \in \mathscr{F}_1'$ and $F_2' \in \mathscr{F}_2'$. But then $F_1 = F_1'$, so $F_1 \in \mathscr{F}_1'$. It follows that $\mathscr{F}_1' = \mathscr{F}_1$.

Now suppose that $F_{2'} \in \mathscr{F}_{2'}$. Since some $F \in \mathscr{F}$ meets E_2 , e is not an isthmus of H_1' . Therefore, there exists $F_1' \in \mathscr{F}_1'$ with $e \in F_1'$. Then $F = F_1' \cup F_2' - e \in \mathscr{F}$, so $F_2' = F \cap E_2 \in \mathscr{F}_2$. Thus $\mathscr{F}_2' \subseteq \mathscr{F}_2$.

Finally, suppose that $F_2 \in \mathscr{F}_2$. Then $F_2 = F \cap E_2$ for some $F \in \mathscr{F}$ such that F meets E_2 . Now $F = F_1' \cup F_2' - e$, where $e \in F_1' \in \mathscr{F}_1'$ and $F_2' \in \mathscr{F}_2'$. Then $F_2 = F_2'$, so $F_2 \in \mathscr{F}_2'$. It follows that $\mathscr{F}_2 = \mathscr{F}_2'$.

Therefore, $H_1' = H_1$ and $H_2' = H_2$, provided H has no isthmus. In this case, it is clear that H_1 , H_2 also have no isthmus, and the proof is complete.

It is possible to give examples to show that the uniqueness of H_1 , H_2 can fail if H is allowed to have isthmuses. Therefore, set families considered in this section will not have isthmuses. We do not exclude loops, or require non-separability.

The substitution composition seems to be harder to work with than the composition studied (implicitly) in Section 5. In developing a decomposition theory based on this composition, it will be convenient to consider the objects being decomposed to be pairs (H, e), where H is a set family and e is not a cell of H. We will see that this device makes the composition easier to handle by symmetrizing it.

Suppose that $H = H_1[H_2; e_2]$. Let e_1 be an element which is not a cell of H and is different from e_2 . Then we say that $\{(H_1, e_1), (H_2, e_2)\}$ is a *simple factorization of* (H, e_1) . We say that e_2 is the *marker* of the simple factorization. A *factorization* of (H, e) is defined inductively to be either $\{(H, e)\}$ or a set D' obtained from a factorization D of (H, e) by replacing a member (H_1, e_1) of D by the members of a simple factorization of (H_1, e_1) , such that the marker of this simple factorization is neither a cell of H nor an element e' such that $(H', e') \in D$ for some H'. If H''is obtained from H by a (non-empty) sequence of operations of the kind described above, then D'' is said to be a (*strict*) refinement of D. A marker of a factorization D of (H, e) is an element $e' \neq e$ such that $(H', e') \in D$ for some H'. A component of D is a set family H' such that $(H', e') \in D$ for some e'. Clearly each marker of D is a cell of exactly one component of D. A factorization D is *trivial* if |D| = 1.

Given a factorization D of (H, e), we may form a directed graph G as follows: The vertices of G are the components of D and the edges are its markers; the marker e' is directed from H_1 to H_2 , where H_1 is the component of D such that $(H_1, e') \in D$ and H_2 is the component of D of which e' is a cell. It is easy to see that G is a tree with the property that every vertex but one is the tail of exactly one edge; the exceptional vertex is the component H' of D such that $(H', e) \in D$. As for the tree associated with a decomposition in earlier sections, this directed tree provides a convenient way to visualize a factorization.

The terms "equivalent" and "minimal" are defined for factorizations in the same way as for decompositions. If e is not a cell of the set family $H = (E, \mathscr{F})$, we define eH to be the set family $(E + e, e\mathscr{F})$, where $e\mathscr{F} = \{F + e: F \in \mathscr{F}\}$. It is easy to see that eH is loopless; also, eH is non-separable if and only if H has no isthmus. The following result links the present notion of factorization to that of decomposition, discussed in Section 5.

THEOREM 21. D is a factorization of (H, e) if and only if $D' = \{e'H': (H', e') \in D\}$ is a decomposition of eH.

Proof. The result is clearly true for |D| = 1. From the definitions of factorization and decomposition, it is enough to prove the result for |D| = 2. Suppose that $\{(H_1, e_1), (H_2, e_2)\}$ is a simple factorization of (H, e_1) , where

$$H_1 = (E_1 + e_2, \mathscr{F}_1), H_2 = (E_2, \mathscr{F}_2), \text{ and } H = (E, \mathscr{F}).$$

Then

$$\mathscr{F} = (\mathscr{F}_1 \setminus e_2) \cup \{F_1 \cup F_2 - e_2 : e_2 \in F_1 \in \mathscr{F}_1, F_2 \in \mathscr{F}_2\}.$$

Thus

$$e_{1}\mathscr{F} = (e_{1}(\mathscr{F}_{1}\backslash e_{2})) \cup \{F_{1} \cup F_{2} - e_{2} + e_{1}: e_{2} \in F_{1} \in \mathscr{F}_{1}, F_{2} \in \mathscr{F}_{2}\}$$
$$= (e_{1}(\mathscr{F}_{1}\backslash e_{2})) \cup \{F_{1} \cup F_{2} - e_{2}; e_{2} \in F_{1} \in e_{1}\mathscr{F}_{1}, e_{2} \in F_{2} \in e_{2}\mathscr{F}_{2}\}.$$

Since $(e_2 \mathscr{F}_2) \setminus e_2 = \emptyset$, it follows that $\{e_1 H_1, e_2 H_2\}$ is a simple decomposition of $e_1 H'$ with marker e_2 .

Now suppose that $\{e_1H_1, e_2H_2\}$ is a simple decomposition of e_1H . Then,

since $(e_2 \mathscr{F}_2) \setminus e_2 = \emptyset$, we have

$$e_{1}\mathscr{F} = ((e_{1}\mathscr{F}_{1})\backslash e_{2}) \cup \{F_{1} \cup F_{2} - e_{2} : e_{2} \in F_{1} \in e_{1}\mathscr{F}_{1},$$

$$e_{2} \in F_{2} \in e_{2}\mathscr{F}_{2}\}$$

$$= ((e_{1}\mathscr{F}_{1})\backslash e_{2}) \cup \{F_{1} \cup F_{2} - e_{2} + e_{1} : e_{2} \in F_{1} \in \mathscr{F}_{1}, F_{2} \in \mathscr{F}_{2}\}.$$

Thus

$$\mathscr{F} = (\mathscr{F}_1 \setminus e_2) \cup \{F_1 \cup F_2 - e_2 \colon e_2 \in F_1 \in \mathscr{F}_1, F_2 \in \mathscr{F}_2\},\$$

so $\{(H_1, e_1), (H_2, e_2)\}$ is a simple factorization of (H, e_1) .

COROLLARY. If $\{E_1, E_2\}$ is a partition of E, then E_1 is a committee of H if and only if $\{E_1 + e, E_2\}$ is a split of eH.

A set family *H* is *irreducible* if, whenever *e* is not a cell of *H*, (*H*, *e*) has no non-trivial factorization; $H = (E, \mathscr{F})$ is *fragile* if every subset $A \subset E$ such that $|A| \ge 2$, is a committee of *H*. Applying Theorem 8, we obtain the following result.

THEOREM 22. Let H be a set family having no isthmus and let e be an element which is not a cell of H. Then (H, e) has a unique minimal factorization, each of whose components is irreducible or fragile.

To characterize the fragile set families, we merely check the list (Theorem 9) of brittle set families; a set family H is fragile if and only if eH is brittle. Every brittle set family $H = (E, \mathscr{F})$, which has a cell $e \in E$ such that $e \in F$ for every $F \in \mathscr{F}$, gives rise to a fragile set family in this way. The ones that do not have this property are bonds and 0-superstars. The fact that not all brittle families correspond in this way to fragile families is one indication that the theory of Section 5 is a proper generalization of the present theory. A set family $H = (E, \mathscr{F})$ is called a *revised superstar* if

 $\mathscr{F} = \{F: A \subseteq F \subseteq E, F \neq \emptyset\}$ for some $A \subseteq E$;

H is called a *flower* if

 $\mathscr{F} = \{F: F \subseteq E, |F| = 1\}.$

The following results are immediate consequences of Theorems 9 and 10.

THEOREM 23. A set family having no isthmus is fragile if and only if it is a flower or a revised superstar.

THEOREM 24. Let H be a set family having no isthmus, and e be an element which is not a cell of H. Then (H, e) has a unique minimal factorization D such that each component of D is irreducible, a flower, or a revised superstar.

Just as in Section 5, we can restrict the theory to clutters. It is easy to see that the components of a factorization of a clutter are themselves clutters. Notice that the fragile clutters are the flowers and the polygons. Thus we have the following result.

THEOREM 25. Let H be a clutter having no isthmus and let e be an element which is not a cell of H. Then (H, e) has a unique minimal factorization D such that each component of D is an irreducible clutter, a flower or a polygon.

Theorem 25 could be obtained from Theorem 11 in exactly the same way that Theorem 24 was obtained from Theorem 10. Thus Theorem 10 may be said to generalize Theorem 25 in two different directions.

In view of Proposition 3, Theorem 25 yields a unique decomposition theorem for monotone Boolean functions. The fragile clutters, namely the polygons and flowers, have associated Boolean functions which are particularly simple. If f is a monotone Boolean function on $\{0, 1\}^{E}$, where $E = \{1, 2, ..., n\}$, then c(f) is a polygon if and only if f is given by

$$f(x_1, x_2, \ldots, x_n) = x_1 x_2 \ldots x_n,$$

and c(f) is a flower if and only if f is given by

$$f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$$

It should be remarked that the generalization of the substitution decomposition to arbitrary set families does not, as one might hope, provide a decomposition theory for arbitrary Boolean functions.

The substitution decomposition has been studied previously ([1], [5], [20], [4], [6]), and a uniqueness theorem has been proved ([20], [4]). It is shown in [20] that a certain uniquely-determined procedure can be carried out; that procedure constructs a factorization of a clutter, and it is easy to see that that factorization is our "standard" factorization. As is shown in [10], this theorem can be derived from the present theory; however, it is not equivalent to our Theorem 25.

Bixby [6] has applied the substitution decomposition to matroids in the following way. Let e be a cell of a non-separable matroid $M = (E, \mathscr{F})$. The clutter $M(e) = (E - e, \mathscr{F}(e))$, where $\mathscr{F}(e) = \{F - e: e \in F \in \mathscr{F}\}$, is called a (matroidal) path clutter. (A fundamental result of [16] states that M(e) determines M.) Perhaps surprisingly, the substitution decomposition theory for M(e) can be shown to be equivalent to the theory of Section 6 for M. In particular, Bixby shows that M(e) is irreducible if and only if M is 3-connected.

The *clique clutter* of a finite simple graph G is the set family $C(G) = (V(G), \mathscr{F})$, where the members of \mathscr{F} are the vertex-sets of maximal complete subgraphs of G. Clearly, C(G) determines G (except for the names of the edges). Moreover, C(G) has no isthmuses. The substitution composition for clique clutters is easily seen to yield the following graph

composition, called by Chvátal [9], "graph substitution". Given graphs G_1, G_2 and $v \in V(G_1), G_1[G_2; v]$ is the graph obtained from G_1 by replacing v by G_2 and joining every vertex of G_2 to every neighbour in G_1 of v. Using Theorem 25, and supplying appropriate definitions, it is easy to derive a unique factorization theorem for graphs, based on this composition; the components of the factorization are irreducible graphs, complete graphs, and edgeless graphs.

References

- 1. R. L. Ashenhurst, The decomposition of switching functions, in: Proceedings of the international symposium on the theory of switching (Part 1), (Harvard University Press, Cambridge, 1959).
- 2. C. Berge, Graphs and hypergraphs (North Holland, Amsterdam, 1973).
- L. J. Billera, *Clutter decomposition and monotonic Boolean functions*, Annals of the New York Academy of Sciences 175, Article 1 (1970), 41–48.
- On the composition and decomposition of clutters, J. Combinatorial Theory 11 (1971), 234-245.
- 5. Z. W. Birnbaum and J. D. Esary, Modules of coherent binary systems, SIAM J. Applied Math. 13 (1965), 444-462.
- 6. R. E. Bixby, Composition and decomposition of matroids and related topics, Thesis, Cornell University (1972).
- 7. R. E. Bixby and W. H. Cunningham, Matroids, graphs, and 3-connectivity, in: Graph theory and related topics (Academic Press, New York, 1979), 91-103.
- T. H. Brylawski, A decomposition for combinatorial geometries, Trans. Amer. Math. Soc. 171 (1972), 235–282.
- 9. V. Chvátal, On certain polytopes associated with graphs, J. Combinatorial Theory B 18 (1975), 138-154.
- 10. W. H. Cunningham, A combinatorial decomposition theory, Thesis, University of Waterloo (1973).
- J. Edmonds, Matroid partition, Math. of the Decision Sciences, A.M.S. Lectures in Applied Math. 11 (1968), 335-346.
- 12. Matroid intersection, Annals of Discrete Math. 4 (1979), 39-47.
- J. Edmonds and R. Giles, A min-max relation for submodular functions on graphs, Annals of Discrete Math. 1 (1977), 185-204.
- 14. J. E. Hopcroft and R. E. Tarjan, Isomorphism of planar graphs, in: Complexity of computer computations (Plenum Press, New York, 1972).
- Dividing a graph into triconnected components, SIAM J. Computing 2 (1973), 135–158.
- 16. A. Lehman, A solution of the Shannon switching game, SIAM J. Applied Math. 12 (1964), 687–725.
- S. Maclane, A structural characterization of planar combinatorial graphs, Duke Math. J. 3 (1937), 460-472.
- 18. G. J. Minty, On the axiomatic foundations of the theories of directed linear graphs, electrical networks, and network-programming, J. Math. Mech. 15 (1966), 485-520.
- 19. W. Richardson, Connectivity, decomposition and complexity in matroids, Thesis, University of Waterloo (1973).
- 20. L. S. Shapley, On committees, in: New methods of thought and procedure (Springer, New York, 1968).
- 21. C. A. B. Smith, Patroids, J. Combinatorial Theory B 16 (1974), 64-76.
- 22. B. A. Trakhtenbrot, Towards a theory of non-repeating contact schemes (Russian), Trudi MIAN SSSR 51 (1958), 226-269.

- 23. W. T. Tutte, Connectivity in graphs (University of Toronto Press, Toronto, 1966).
- 24. Connectivity in matroids, Can. J. Math. 18 (1966), 1301-1324.
- 25. T. R. S. Walsh, Counting labelled three-connected and homeomorphically irreducible 2-connected graphs, submitted for publication.
- 26. —— Counting unlabelled three-connected graphs, submitted for publication.
- 27. D. J. A. Welsh, Matroid theory (Academic Press, London, 1976).
- H. Whitney, Non-separable and planar graphs, Trans. Amer. Math. Soc. 34 (1932), 339-362.

Carleton University, Ottawa, Ontario; University of Waterloo, Waterloo, Ontario