Database Design for Incomplete Relations

Mark Levene
University College London
and
George Loizou
Birkbeck College

Name: Mark Levene
Address: Gower Street, London WC1E 6BT, U.K., mlevene@cs.ucl.ac.uk
Affiliation: Department of Computer Science, University College London, University of London

Name: George Loizou
Address: Malet Street, London WC1E 7HX, U.K., george@dcs.bbk.ac.uk
Affiliation: Department of Computer Science, Birkbeck College, University of London

Although there has been a vast amount of research in the area of relational database design, to our knowledge, there has been very little work that considers whether this theory is still valid when relations in the database may be incomplete.

When relations are incomplete and thus contain null values the problem of whether satisfaction is additive arises. Additivity is the property of the equivalence of the satisfaction of a set of functional dependencies (FDs), F, with the individual satisfaction of each member of F in an incomplete relation. It is well known that, in general, satisfaction of FDs is not additive. Previously we have shown that satisfaction is additive if and only if the set of FDs is monodependent. A set of FDs is monodependent if it satisfies both the intersection property and the split-freeness property. (The two defining properties of monodependent sets of FDs correspond to the two defining properties of conflict-free sets of multivalued dependencies.)

We conclude that monodependence is a fundamental desirable property of a set of FDs when considering incomplete information in relational database design. This provides the motivation for our investigation. We now outline the desirable properties that monodependent sets of FDs possess.

We show that when the set of FDs F either satisfies the intersection property or the split-freeness property then the problem of finding an optimum cover of F can be solved in polynomial time in the size of F; in general, this problem is known to be NP-complete. We also show that when F satisfies the split-freeness property then deciding whether there is a superkey of cardinality k or less can be solved in polynomial time in the size of F, since all the keys have the same cardinality. If F only satisfies the intersection property then this problem is NP-complete, as in the general case. Moreover, we show that when F either satisfies the intersection property or the split-freeness property then deciding whether an attribute is prime can be solved in polynomial time in the size of F; in general, this problem is known to be NP-complete.

Assume that a relation schema R is in an appropriate normal form with respect to a set of FDs F. We show that when F satisfies the intersection property then the notions of Second Normal Form and Third Normal Form are equivalent. We also show that when R is in Boyce-Codd Normal Form (BCNF), then F is monodependent if and only if either there is a unique key for R, or for all keys X for R, the cardinality of X is one less than the number of attributes associated with R.

Finally, we tackle a long standing problem in relational database theory by showing that when a set of FDs F over R satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R with respect to F is also dependency preserving. As a corollary of this result we are able to show that when F satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R, which is in BCNF, is also dependency preserving. A lossless join decomposition of R, which is in BCNF, is said to be optimum if it has the smallest possible size. Our final result is that when F is monodependent then there exists a unique optimum lossless join decomposition of R, which is in BCNF, and is also dependency preserving. Furthermore, this ultimate decomposition can be attained in polynomial time in the size of F. From the database designer's point of view this is the ideal situation.

Categories and Subject Descriptors: H.2.1 [**Database Management**]: Logical Design—*data models; normal forms*

General Terms: Design, Theory, Complexity

Additional Key Words and Phrases: Incomplete information, null functional dependencies, additivity problem, monodependence, intersection property, split-freeness property, optimum cover, prime attribute problem, superkey of cardinality k problem, normal forms, lossless join decomposition, dependency preserving decomposition

1. INTRODUCTION

Relational database design is a very important branch of relational database theory and is extensively covered in most database textbooks [Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993; Date 1995]. One of the main aims of relational database design is to *normalise* an input relation schema together with a set of data dependencies into an appropriate normal form. The motivation behind the various normal forms is to remove the problems that are caused by the infamous update anomalies and redundancy problems [Codd 1972; Vincent 1994]. When the input set of data dependencies is a set of *Functional Dependencies* (or simply FDs) an appropriate normal form may be *Second*, *Third* or *Boyce-Codd* Normal Form (abbreviated to 2NF, 3NF and BCNF, respectively). As a result of normalisation we obtain a collection of relation schemas, called a *decomposition*. With each relation schema in the decomposition we associate the *projection* of the input set of FDs onto the relation schema. Two desirable properties of decompositions are that they possess the *lossless join property* (or simply are lossless) [Aho et al. 1979] and that they be *dependency preserving* [Beeri and Honeyman 1981] (see also [Rissanen 1977; Vardi 1982]). Although it is always possible to obtain a decomposition which is in 2NF, 3NF or BCNF, and is also lossless, it is not always possible to achieve a decomposition which is in BCNF, and is also both lossless and dependency preserving [Beeri and Bernstein 1979; Osborn 1979].

Despite the vast amount of research in the area of relational database design, to our knowledge, there has been very little work that considers whether this theory is still valid when the relations in the database may be incomplete. This provides the motivation for the current paper. Before outlining our results we briefly describe the underlying model of incomplete information that we use herein. (A detailed account can be found in [Levene and Loizou 1997; Levene and Loizou 1998]; for alternative treatments to dealing with incomplete information we refer the reader to Chapter 5 in [Levene and Loizou 1999]).

We follow Codd's suggestion [Codd 1979] of adding an unmarked null value, denoted by *unk*, to the underlying database domains, whose meaning is "value at present exists but is unknown". We call a relation whose tuples may contain the null value *unk*, an *incomplete relation*. The semantics of an incomplete relation r are defined in terms of the *possible worlds* relative to r; we denote this set of possible worlds by POSS(r). Each possible world in POSS(r) is a *complete relation*, i.e. a relation without any occurrence of *unk*, emanating from a possible substitution of all the occurrences of *unk* in r by nonnull values in the underlying database domains.

FDs are by far the most common integrity constraint in the real world [Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993] and the notion of a key (derived from a given set of FDs) [Codd 1979] is fundamental to the relational model. Given a set of FDs F over a relation schema R and an incomplete relation r over R we say that r satisfies an FD $X \rightarrow Y$, denoted by $r \models X \rightarrow Y$, if there is some complete relation $s \in$ POSS(r) that satisfies $X \rightarrow Y$ in the standard manner [Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993]; we denote this satisfaction by $s \models X \rightarrow Y$. On extending the definition of satisfaction to a set of FDs, it is therefore natural to say that r satisfies F, denoted by $r \models F$, if there is some complete relation s $\in$ POSS(s), such that for all FDs X $\rightarrow$ Y in F, s $\models$ X $\rightarrow$ Y.

This gives rise to the *additivity problem*, which is the problem of whether the statement that r $\models$ F is equivalent to the statement that for all FDs X $\rightarrow$ Y in a *reduced cover* G of F [Beeri and Bernstein 1979] r $\models$ X $\rightarrow$ Y [Levene and Loizou 1997; Levene and Loizou 1998] (cf. [Atzeni and De Antonellis 1993]). (Example 1 shows why we define r $\models$ F in terms of a reduced cover of F.) If these two statements are equivalent for a class of incomplete relations and a class of sets of FDs then we say that *satisfaction is additive* with respect to these classes. If satisfaction is not additive, then a set of FDs F, in this nonadditive class, may be viewed as contradictory. Thus we consider the solution of the additivity problem to be an important prerequisite for any relational database system supporting FDs in the context of incomplete information, since otherwise semantic anomalies may arise. It is well known that, in general, satisfaction of FDs is not additive [Atzeni and Morfuni 1986; Levene and Loizou 1997; Levene and Loizou 1998]. For example, the incomplete relation r_1 [Atzeni and Morfuni 1986] shown in Table 1 is such that $r_1 \models$ A $\rightarrow$ B and $r_1 \models$ B $\rightarrow$ C but $r_1 \not\models$ {A $\rightarrow$ B, B $\rightarrow$ C}. This is due to the fact that $\forall s_1 \in$ POSS(r_1), $s_1 \not\models$ A $\rightarrow$ C, which is inferred from A $\rightarrow$ B and B $\rightarrow$ C by the transitivity rule of Armstrong's axiom system [Armstrong 1974; Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993]. Another example is the incomplete relation r_2 shown in Table 2, where $r_2 \models$ A $\rightarrow$ C and $r_2 \models$ B $\rightarrow$ C, but $r_2 \not\models$ {A $\rightarrow$ C, B $\rightarrow$ C}. This is due to the fact that $\forall s_2 \in$ POSS(r_2), the C-value of the second tuple is either 1 or 0 and therefore either $s_2 \not\models$ A $\rightarrow$ C or $s_2 \not\models$ B $\rightarrow$ C. In this case two incomparable sets of attributes, A and B, functionally determine a common attribute C. It is an interesting fact that it is also the case that $r_2 \not\models$ {A $\rightarrow$ B, B $\rightarrow$ C} and thus the second relation is also a counterexample to the first set of FDs.

A	B	C
0	*unk*	0
0	*unk*	1

Table 1. The counterexample relation r_1

A	B	C
0	*unk*	0
0	0	*unk*
unk	0	1

Table 2. The counterexample relation r_2

In [Levene and Loizou 1997] we introduced the class of *monodependent* sets of FDs. A set of FDs F over a relation schema R is monodependent if the following two properties are satisfied. The first property, called the *intersection property*, informally states that for each attribute A in the attribute set associated with R there is a unique nontrivial and reduced FD in the closure of F that functionally determines A. The second property, called the *split-freeness property*, informally states that there are no two nontrivial FDs in the closure of F such that the right-hand side of each of the two FDs splits the left-hand side of the other FD. (The two defining properties of monodependent sets of FDs correspond to the two defining properties of *conflict-free* sets of *Multivalued Dependencies* (or simply MVDs) [Sciore 1981; Lien 1982; Beeri et al. 1983; Beeri and Kifer 1986]; in Section 4 we exhibit a simple example showing that, in general, monodependence does not imply conflict-freeness.) The main result in [Levene and Loizou 1997] shows that satisfaction is additive with respect to the class of all incomplete relations and a class of sets of FDs **FC**, if and only if all the sets of FDs in **FC** are monodependent sets of FDs. Therefore, monodependence provides a solution to the additivity problem.

We conclude that monodependence is a fundamentally desirable property of a set of FDs when considering incomplete information in relational database design. Therefore the starting point of our investigation is to assume that we are given a monodependent set of FDs F over a relation schema R. We next outline the desirable properties that monodependent sets of FDs possess.

In the process of database design it is common to use a *cover* of a set of FDs F over a relation schema R that represents the FDs in F in a succinct manner [Maier 1980; Mannila and Räihä 1983]. Ideally we would like to use an *optimum* cover of F, i.e. a cover of F with the smallest possible size, but the problem of finding an optimum cover of a set of FDs is, in general, NP-complete [Maier 1980]. On the other hand, an *LR-minimum* cover of F, i.e. a cover of F with the smallest possible cardinality and such that the left-hand sides and right-hand sides of the FDs in the cover are minimal, can be obtained in polynomial time in the size of F [Maier 1980]. (We call an LR-minimum cover of F simply a *minimum* cover of F.)

We show that when F satisfies the intersection property then a minimum cover of F is also an optimum cover of F. It follows that the optimum cover of F can be computed in polynomial time in the size of F. In addition, we show that the optimum cover is unique, that it can have at most n FDs, and that its size can be no more than n^2, where n is the cardinality of the attribute set of R. We also show that when F satisfies the split-freeness property then a minimum cover has optimum left-hand sides and an optimum cover of F can be obtained in polynomial time, although it is not unique.

An important problem in database design is that of deciding whether there is a superkey for R with respect to F of cardinality k or less. This problem, known as the superkey of cardinality k problem, was shown to be NP-complete in [Lucchesi and Osborn 1978]. We show that when F satisfies only the intersection property then this problem remains NP-complete. On the other hand, when F satisfies the split-freeness property then this problem can be solved in polynomial time in the size of F. In fact, in this case all the keys for R with respect to F have the same cardinality and thus the problem reduces to that of finding a single key for R, which can easily be found in polynomial time in the size of F (see Algorithm 3.3. in [Atzeni and De Antonellis 1993] and [Kundu 1985]). A related problem in database design is that of deciding whether an attribute is *prime* with respect to F, i.e. whether an attribute belongs to at least one key for R. This problem, known as the prime attribute problem, is also, in general, NP-complete [Lucchesi and Osborn 1978]. We show that when F either satisfies the intersection property or the split-freeness property then this problem can be solved in polynomial time in the size of F.

Assume that R is in an appropriate normal form with respect to F. We show that when F satisfies the intersection property then the notions of 2NF and 3NF are equivalent. In fact, we show that in this case if the number of keys for R is greater than one then all the attributes associated with R are prime. In the special cases when there is a unique key for R with respect to F, or all the keys for R are *simple*, i.e. their cardinality is one, we show that the notions of 2NF, 3NF and BCNF are all equivalent. We also show that when R is in BCNF, then F is monodependent if and only if either F has a unique key for R, or for all keys X for R, the cardinality of X is one less than the number of attributes associated with R.

Our next result tackles a long standing problem in relational database theory. We show that when F over R satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R with respect to F is also dependency preserving. This result is significant in database design, since assuming that F is monodependent, the users need only concern themselves with attaining a decomposition which is lossless. Dependency preservation comes as a by-product of F being monodependent. As a corollary of this result we are able to show that when F satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R, which is in BCNF, is also dependency preserving.

A lossless join decomposition of R, which is in BCNF, is said to be *optimum* if it has the smallest possible size. Our final result is that there is a unique optimum lossless join decomposition of R, which is in BCNF, and is also dependency preserving. Furthermore, this ultimate decomposition can be attained in polynomial time in the size of F. From the database designer's point of view this is the ideal situation.

Due to the many desirable properties of monodependent sets of FDs in the context of relational database design, our results are relevant whether we allow incomplete relations in the database or not. Our results show an interesting correspondence between monodependence as a desirable property, when the data dependencies are FDs, and conflict-freeness as a desirable property, when the data dependencies are MVDs [Sciore 1981; Lien 1982; Beeri et al. 1983; Beeri and Kifer 1986].

The layout of the rest of the paper is as follows. In Section 2 we formalise the notion of incomplete relations. In Section 3 we define the notion of an FD being satisfied in an incomplete relation. In Section 4 we introduce monodependent sets of FDs. In Section 5 we discuss the additivity problem and state the theorem that shows that monodependent sets of FDs solve this problem. In Section 6 we show that finding an optimum cover of a set of FDs that either satisfies the intersection property or the split-freeness property can be done in polynomial time. In Subsection 6.1 we show that when F satisfies the intersection property and is minimum then it is also optimum and, in addition, this optimum cover is unique. In Subsection 6.2 we show that when F satisfies the split-freeness property and is minimum then it has optimum left-hand sides. In Subsection 6.3 we give some technical lemmas concerning the structure of monodependent sets of FDs, which are used in the remaining sections. In Section 7 we investigate the structure of the set of keys for R, when F is monodependent. In Subsection 7.1 we show that the superkey of cardinality k problem remains NP-complete when F satisfies only the intersection property but can be solved in polynomial time when F satisfies the split-freeness property. In Subsection 7.2 we show that the prime attribute problem can be solved in polynomial time in the size of F when F either satisfies the intersection property or the split-freeness property. In Section 8 we investigate the effect that monodependence has on a relation schema which is in a particular normal form. In Section 9 we show that when a set of FDs F over R satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R is also dependency preserving, and show that a unique, optimum, lossless join decomposition of R, which is in BCNF, and is also dependency preserving can be obtained in polynomial time in the size of F. Finally, in Section 10 we give our concluding remarks.

Table 3 summarises the main complexity results obtained in the paper: FD denotes the general class of FDs, IP denotes the class of FDs satisfying the intersection property, SF denotes the class of FDs satisfying the split-freeness property and MONO denotes the class of monodependent FDs. NPC stands for NP-complete, NPH stands for NP-hard, P stands for polynomial time and "–" stands for not investigated. The leftmost column denotes the problem considered.

Problem / Class	FD	IP	SF	MONO
Optimum cover	NPC	P	P	P
Superkey of cardinality k	NPC	NPC	P	P
Prime attribute	NPC	P	P	P
3NF	NPC	P	P	P
Optimum BCNF	NPH	–	–	P

Table 3. Summary of main complexity results

2. RELATIONS THAT MODEL INCOMPLETE INFORMATION

In this section we extend relation schemas and relations so as to model incomplete information.

We use the notation |S| to denote the cardinality of a set S. If S is a subset of T we write S $\subseteq$ T and if S is a proper subset of T we write S $\subset$ T. Furthermore, S and T are *incomparable* if S $\not\subseteq$ T and T $\not\subseteq$ S; otherwise S and T are *comparable*. We often denote the singleton {A} simply by A, and the union of two sets S, T, i.e. S $\cup$ T, simply by ST. We will make use of the index set $I = \{1, \ldots, n\}$.

Definition 1. A *relation schema* R is a finite set of attributes which we denote by schema(R); we denote the cardinality of schema(R) by type(R). From now on we abbreviate schema(R) to sch(R).

We assume a countably infinite domain of constants, *Dom*, containing two distinguished constants *unk* and *inc*, denoting the null values "unknown" and "inconsistent", respectively.

A type(R)-tuple (or simply a tuple whenever type(R) is understood from context) is a total mapping t from sch(R) into *Dom* such that $\forall A_i \in$ sch(R), t(A_i) $\in$ *Dom*. A relation over R is a finite set of type(R)-tuples.

From now on we let R *be a relation schema and* r *be a relation over* R. As usual uppercase letters (which may be subscripted) from the end of the alphabet such as X, Y, Z will be used to denote sets of attributes, while those from the beginning of the alphabet such as A, B, C will be used to denote single attributes or singleton sets of attributes.

We note that we have actually included two types of null value in our formalism: *unk* and *inc*. The inclusion of *unk* was motivated in the introduction and the inclusion of *inc* is motivated by the fact that it allows us to easily detect unwanted inconsistency.

DEFINITION 2. The *projection* of a type(R)-tuple t onto a set of attributes Y $\subseteq$ sch(R), denoted by t[Y] (and also called the Y-value of t), is the restriction of t to Y. The projection of a relation r onto Y, denoted as π_Y(r), is defined by π_Y(r) = {t[Y] | t $\in$ r}.

DEFINITION 3. A type(R)-tuple is said to be *complete* if $\forall A_i \in$ sch(R), t[A_i] $\neq$ *unk* and t[A_i] $\neq$ *inc*, i.e. t[A_i] is a nonnull value, otherwise t is said to be *incomplete*; t is said to be *inconsistent* if $\exists A_i \in$ sch(R) such that t[A_i] = *inc*, otherwise t is said to be *consistent*.

A relation r is said to be *complete* if $\forall t \in r$, t is complete, otherwise r is said to be *incomplete* (when no confusion arises we use the term relation to mean incomplete relation). The relation r is said to be *inconsistent* if $\exists t \in r$ such that t is inconsistent, otherwise r is said to be *consistent*.

DEFINITION 4. We define a partial order in *Dom*, denoted by $\sqsubseteq$, as follows:

$$u \sqsubseteq v \text{ if and only if } u = v \text{ or } u = unk \text{ or } v = inc, \text{ where } u, v \in Dom.$$

We extend $\sqsubseteq$ to be a partial order in the set of type(R)-tuples as follows: where t_1 and t_2 are type(R)-tuples, t_1 is *less informative than* t_2, written $t_1 \sqsubseteq t_2$, if $\forall A_i \in$ sch(R), $t_1[A_i] \sqsubseteq t_2[A_i]$.

DEFINITION 5. The set of all *possible worlds* relative to a relation r, denoted by POSS(r), is defined by

$$\begin{aligned} POSS(r) = \{s \mid\; & s \text{ is a relation over R and there exists a total and onto mapping} \\ & f : r \rightarrow s \text{ such that } \forall t \in r, t \sqsubseteq f(t) \text{ and } f(t) \text{ is complete}\}. \end{aligned}$$

Our definition of possible worlds embodies the *Closed World Assumption* [Reiter 1978; Imielinski and Lipski Jr. 1984], since POSS(r) allows only completed tuples from the relation r to be present in POSS(r). On the other hand, if we had taken the *Open World Assumption* [Reiter 1978; Imielinski and Lipski Jr. 1984], then the definition of POSS(r) would have to be relaxed so as to allow additional tuples to be present in POSS(r), which do not necessarily result from completed tuples in the relation r. It can be shown that the results obtained herein also hold when the OWA approach is taken, since a closed world in POSS(r) can be viewed as the greatest lower bound of all open worlds emanating from it.

PROPOSITION 1. A relation r is inconsistent if and only if POSS(r) = $\emptyset$. □

Hereafter we assume that *relations are consistent*.

3. NULL FUNCTIONAL DEPENDENCIES

We now define Functional Dependencies (FDs) and their satisfaction in the context of incomplete relations.

DEFINITION 6. A *functional dependency* over R (or simply an FD) is a statement of the form X $\rightarrow$ Y, where X, Y $\subseteq$ sch(R).

We call an FD of the form X $\rightarrow$ Y, where Y $\subseteq$ X, a *trivial* FD. Two nontrivial FDs of the forms X $\rightarrow$ A and Y $\rightarrow$ A are said to be *incomparable* if X and Y are incomparable. Two nontrivial FDs of the forms XB $\rightarrow$ A and YA $\rightarrow$ B are said to be *cyclic*.

We stress the fact that we allow FDs whose left-hand side is the empty set. From now on we let F *be a set of FDs over* R. We define the *size* of F, denoted by $||F||$, to be the number of attributes appearing in F including repetitions.

DEFINITION 7. An FD X $\rightarrow$ Y is satisfied in a relation r, denoted by r $\models$ X $\rightarrow$ Y, if and only if $\exists$ s $\in$ POSS(r) such that $\forall t_1, t_2 \in$ s, if t_1[X] = t_2[X] then t_1[Y] = t_2[Y].

We note that the definition of satisfaction of an FD in a relation reduces to the standard definition of the satisfaction of an FD when the relation is complete [Ullman 1988] (in that case there exists exactly one s $\in$ POSS(r)).

The following lemma, which gives a syntactic characterisation of satisfaction of an FD, follows from Definition 7 (cf. Lemma 6.2 in [Atzeni and De Antonellis 1993]).

LEMMA 2. Let r be a relation and X, Y $\subseteq$ sch(R). Then r $\models$ X $\rightarrow$ Y if and only if $\forall t_1, t_2 \in$ r, if t_1[X] and t_2[X] are complete and t_1[X] = t_2[X], then $t_1[Y] \sqcup t_2[Y]$ is consistent, where $t_1[Y] \sqcup t_2[Y]$ denotes the least upper bound of t_1[Y] and t_2[Y] with respect to $\sqsubseteq$. □

DEFINITION 8. A set of FDs F *logically implies* an FD X $\rightarrow$ Y, written F $\models$ X $\rightarrow$ Y, if whenever r is a relation then the following condition is true:

$$\text{if } \forall W \rightarrow Z \in F, r \models W \rightarrow Z \text{ holds then } r \models X \rightarrow Y \text{ also holds.}$$

4. MONODEPENDENT SETS OF FUNCTIONAL DEPENDENCIES

We assume that the reader is familiar with Armstrong's axiom system for FDs [Armstrong 1974; Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993], consisting of the reflexivity, augmentation and transitivity inference rules. A fundamental result in relational database theory is that Armstrong's axiom system is sound and complete for FDs holding in complete relations. We denote the closure of a set of FDs F over R with respect to Armstrong's axiom system by F^+. Lien [Lien 1982] and Atzeni and Morfuni [Atzeni and Morfuni 1986] showed that the inference rules: reflexivity, augmentation, decomposition and union, are sound and complete for FDs holding in incomplete relations; we call this axiom system, Lien and Atzeni's axiom system. That is, by dropping the transitivity rule from Armstrong's axiom system and adding the decomposition and union rules, we obtain Lien and Atzeni's axiom system. We denote the closure of a set of FDs F over R with respect to Lien and Atzeni's axiom system by F^*. The soundness and completeness of Lien and Atzeni's axiom system for FDs holding in incomplete relations can be written symbolically as the statement: F $\models$ X $\rightarrow$ Y if and only if X $\rightarrow$ Y $\in F^*$.

DEFINITION 9. A set of FDs G over R is a *cover* of F if $F^+ = G^+$.

DEFINITION 10. An FD X $\rightarrow$ Y $\in F^+$ is *reduced* [Beeri and Bernstein 1979] if there does not exist a set of attributes W $\subset$ X such that W $\rightarrow$ Y $\in F^+$. A set of FDs F is *reduced* if all the FDs in F are reduced.

We note that a reduced cover G of F can be obtained in polynomial time in the size of F [Beeri and Bernstein 1979].

DEFINITION 11. A set of FDs F is a *monodependent* set of FDs over R (or simply monodependent whenever R is understood from context) if $\forall$A $\in$ sch(R), the following two conditions are true:

(1) Whenever there exist incomparable FDs, X $\rightarrow$ A, Y $\rightarrow$ A $\in F^+$, then X $\cap$ Y $\rightarrow$ A $\in F^+$; we call this property the *intersection property*.
(2) Whenever there exist cyclic FDs, XB $\rightarrow$ A, YA $\rightarrow$ B $\in F^+$, then either Y $\rightarrow$ B $\in F^+$ or (X $\cap$ Y)A $\rightarrow$ B $\in F^+$; we call this property the *split-freeness property*.

An immediate consequence of the above definition is that if G over R is a cover of F, then F is monodependent if and only if G is monodependent. In addition, we have shown in [Levene and Loizou 1997] that monodependence of F can be checked in polynomial time in the size of F.

We observe that the two defining properties of monodependent sets of FDs correspond to the two defining properties of *conflict-free* sets of *multivalued dependencies* (MVDs) [Sciore 1981; Lien 1982; Beeri et al. 1983; Beeri and Kifer 1986]. We further observe that the set of MVDs that is logically implied by a monodependent set of FDs may not be conflict-free and thus monodependence is a weaker notion than conflict-freeness. For example, let F = {A $\rightarrow$ B, B $\rightarrow$ A}, with sch(R) = {A, B, C}. It can easily be verified that F is monodependent but that the set of MVDs logically implied by F is not conflict-free.

The next theorem, whose proof can be found in [Levene and Loizou 1997], shows that assuming that F is reduced and satisfies the intersection property, then the closure of F with respect to Armstrong's axiom system (i.e. F^+) is equal to the closure of F with respect to Lien and Atzeni's axiom system (i.e. F^*). This result is fundamental to the theory of FDs in incomplete relations, since it justifies the use of Armstrong's axiom system in the context of incomplete relations when F is monodependent.

THEOREM 3. Assume that F is reduced. If F satisfies the intersection property then $F^+ = F^*$. $\square$

The converse of Theorem 3 is, in general, false. For example, let F = {A $\rightarrow$ C, B $\rightarrow$ C}, with sch(R) = {A, B, C}. It can be easily verified that $F^+ = F^*$; however, F does not satisfy the intersection property, since $\emptyset \rightarrow$ C $\notin F^+$. Moreover, if F satisfies the split-freeness property but not the intersection property, then, in general, $F^+ \neq F^*$. For example, let F = {A $\rightarrow$ B, B $\rightarrow$ C}, with sch(R) = {A, B, C}. It can be easily verified that F satisfies the split-freeness property but not the intersection property, and also that $F^+ \neq F^*$.

The following proposition pertaining to the property of derivations of FDs, using Armstrong's axiom system, which appears as Lemma 2 in [Beeri and Bernstein 1979], will be frequently used in subsequent proofs. We state it in a slightly altered form from the original; herein the use of an FD W $\rightarrow$ Z $\in$ F in a derivation of an FD X $\rightarrow$ Y $\in F^+$ is said to be *nonredundant* if we cannot omit W $\rightarrow$ Z from the derivation and still infer X $\rightarrow$ Y $\in F^+$.

PROPOSITION 4. Let F be a set of FDs and assume that W $\rightarrow$ Z $\in$ F is used nonredundantly in a derivation of an FD X $\rightarrow$ Y $\in F^+$ from F by using Armstrong's axiom system. Then X $\rightarrow$ W $\in$ (F − {W $\rightarrow$ Z})$^+$. □

In some subsequent proofs we utilise the *pseudo-transitivity inference rule* [Ullman 1988], which states that if X $\rightarrow$ Y $\in F^+$ and WY $\rightarrow$ Z $\in F^+$, then it is also the case that XW $\rightarrow$ Z $\in F^+$.

5. THE ADDITIVITY PROBLEM FOR NULL FUNCTIONAL DEPENDENCIES

We next generalise the definition of satisfaction of a single FD to satisfaction of a set of FDs, and discuss the semantics of this generalised definition.

DEFINITION 12. A set of FDs F is satisfied in a relation r, denoted by r $\models$ F, if and only if $\exists$ s $\in$ POSS(r) such that $\forall$ X $\rightarrow$ Y $\in$ F, s $\models$ X $\rightarrow$ Y.

Following [Atzeni and De Antonellis 1993] we next define additive satisfaction.

DEFINITION 13. Satisfaction is *additive* with respect to a class of relations **RC** over R and a class of sets of FDs **FC** over R, whenever: $\forall$ r $\in$ **RC**, $\forall$ F $\in$ **FC**, r $\models$ F if and only if there exists a reduced cover G of F such that $\forall$ X $\rightarrow$ Y $\in$ G, r $\models$ X $\rightarrow$ Y.

The next example shows that in Definition 13 we cannot relax the requirement that G be a reduced cover.

EXAMPLE 1. Consider the relation r_1 shown in Table 1 and let F = {A $\rightarrow$ B, AB $\rightarrow$ C}. It can easily be verified that $r_1 \models$ A $\rightarrow$ B and $r_1 \models$ AB $\rightarrow$ C but $r_1 \not\models$ F. On the other hand, if we let G = {A $\rightarrow$ B, A $\rightarrow$ C}, i.e. G is a reduced cover of F, then $r_1 \not\models$ A $\rightarrow$ C.

The following fundamental result was shown in [Levene and Loizou 1997].

THEOREM 5. Satisfaction is additive with respect to the class of all relations over R and a class of sets of FDs **FC** if and only if all the sets of FDs in **FC** are monodependent. □

6. OPTIMUM COVERS OF MONODEPENDENT SETS OF FDS

A set of FDs F over R is *minimum* if there does not exist a cover of F with fewer FDs than F and all the left-hand sides and right-hand sides of the FDs in F are minimal; in [Maier 1980] a minimum set of FDs is called an *LR-minimum* set of FDs.

DEFINITION 14. A set of FDs F is a *minimum* set of FDs if there is no cover G of F such that G has fewer FDs than F, all the FDs in F are reduced and for every FD X $\rightarrow$ Y $\in$ F and for every Z $\subset$ Y, $((\mathrm{F} - \{\mathrm{X} \rightarrow \mathrm{Y}\}) \cup \{\mathrm{X} \rightarrow \mathrm{Z}\})^+ \neq F^+$.

A set of FDs F over R is *optimum* if there does not exist a cover of F whose size is less than the size of F [Maier 1980].

DEFINITION 15. F is an *optimum* set of FDs if there does not exist a cover G of F such that $||\mathrm{G}|| < ||\mathrm{F}||$.

The following result was shown in [Maier 1980].

LEMMA 6. An optimum set of FDs is minimum. □

DEFINITION 16. Given a set of FDs F the sets of attributes X, Y $\subseteq$ sch(R) are *equivalent* under F, if X $\rightarrow$ Y, Y $\rightarrow$ X $\in F^+$. We denote the subset of FDs in F whose left-hand sides are equivalent to a set of attributes X $\subseteq$ sch(R) by $E_F(\mathrm{X})$; we call the sets $E_F(\mathrm{X})$ the *equivalence classes* of F, and denote the set of left-hand sides of $E_F(\mathrm{X})$ by LHS(X) assuming that F is understood from context.

The set $E_F(\mathrm{X})$ is said to be an *optimum equivalence class* of F if for all sets of FDs G over R satisfying $((\mathrm{F} - E_F(\mathrm{X})) \cup \mathrm{G})^+ = F^+$, we have that $||\mathrm{G}|| \geq ||\mathrm{F}||$.

The following result, which was shown in [Mannila and Räihä 1983], implies that in order to ascertain whether F is optimum we need only consider its equivalence classes separately.

LEMMA 7. A minimum set of FDs F is optimum if and only if all the equivalence classes $E_F(\mathrm{X})$ of F, where X $\subseteq$ sch(R), are optimum. □

The following result, which was shown in [Maier 1980], implies that the cardinalities of the equivalence classes of a set of FDs F are invariant with respect to the cover of F that is used.

LEMMA 8. Let F and G be two minimum sets of FDs over R such that G is a cover of F. Then $\forall$ X $\subseteq$ sch(R), $|E_F(\mathrm{X})| = |E_G(\mathrm{X})|$. □

LEMMA 9. If F is minimum and X $\rightarrow$ Y $\in$ F, then $\forall$ A $\in$ Y, X $\rightarrow$ A is reduced.

Proof. We prove the result by contradiction. Suppose that for some A $\in$ Y, X $\rightarrow$ A is not reduced and thus for some W $\subset$ X, W $\rightarrow$ A $\in F^+$. There are two cases to consider. Firstly, W $\rightarrow$ X $\in F^+$, and thus by Proposition 4 W $\rightarrow$ X $\in (\mathrm{F} - \{\mathrm{X} \rightarrow \mathrm{Y}\})^+$. The result follows, since X $\rightarrow$ Y can be replaced by W $\rightarrow$ Y without changing F^+ contradicting the fact that F is minimum. Secondly, W $\rightarrow$ X $\notin F^+$, and therefore W $\rightarrow$ A $\in$ (F $-$ E_F(X))$^+$, since otherwise W $\rightarrow$ Z $\in$ F^+, with Z $\rightarrow$ X $\in$ F^+, implying that W $\rightarrow$ X $\in$ F^+ leading to a contradiction. The result follows, since we can remove A from Y without changing F^+ contradicting the fact that F is minimum. □

We close this section with the definition of the concept of optimum left-hand sides [Mannila and Räihä 1983].

Definition 17. A minimum set of FDs F has *optimum left-hand sides* if for all minimum covers G of F we have that

$$\sum_{X \rightarrow Y \in F} |X| \leq \sum_{W \rightarrow Z \in G} |W| .$$

6.1 LR-Minimum covers of sets of FDs that satisfy the intersection property are optimum

We show that sets of FDs which satisfy the intersection property have the desirable property that a minimum set of FDs is also optimum. This implies that an optimum cover of F that satisfies the intersection property can be computed in polynomial time in the size of F due to the fact that a minimum cover of F can be computed in polynomial time in the size of F [Maier 1980]. In general, when a set of FDs does not satisfy the intersection property, then finding an optimum cover is NP-complete [Maier 1980]. We show that a set of FDs which satisfies the intersection property has a unique optimum cover. (We observe that a set of FDs can be optimum but not satisfy the intersection property such as the set of FDs {B $\rightarrow$ A, C $\rightarrow$ A}.)

The ensuing lemmas provide the technical background needed to prove the main result of this section, namely Theorem 15.

Lemma 10. Let F be a set of FDs that is minimum and satisfies the intersection property. Then $\forall$ A $\in$ sch(R), there is at most one FD, X $\rightarrow$ Y $\in$ F, such that A $\in$ Y.

Proof. Suppose to the contrary that X $\rightarrow$ Y, W $\rightarrow$ Z $\in$ F are distinct FDs such that A $\in$ Y $\cap$ Z. If X and W are comparable, then the result follows, since we can remove A from one of Y or Z without changing F^+, and thus F cannot be minimum. On the other hand, if X and W are incomparable, then X $\cap$ W $\rightarrow$ A $\in$ F^+, since F satisfies the intersection property. Now, both X $\cap$ W $\rightarrow$ X $\notin$ F^+ and X $\cap$ W $\rightarrow$ W $\notin$ F^+ hold, since otherwise we can replace X $\rightarrow$ Y by X $\cap$ W $\rightarrow$ Y, or respectively, we can replace W $\rightarrow$ Z by X $\cap$ W $\rightarrow$ Z, without changing F^+, and thus F cannot be minimum. Therefore, there exists an FD V $\rightarrow$ U $\in$ F, which is distinct from X $\rightarrow$ Y and W $\rightarrow$ Z, such that X $\cap$ W $\rightarrow$ V $\in$ (F $-$ {X $\rightarrow$ Y, W $\rightarrow$ Z})$^+$ and A $\in$ U. It follows that A can be removed from both Y and Z without changing F^+, and thus F is not minimum. □

The next result, which is a corollary of Lemma 10, gives an upper bound for the cardinality of F when F is minimum and satisfies the intersection property.

Corollary 11. Let F be a set of FDs that is minimum and satisfies the intersection property. Then |F| $\leq$ type(R). □

LEMMA 12. Let F and G be two sets of FDs over R, which are minimum, satisfy the intersection property and are covers of each other. Then Y = Z, whenever X $\rightarrow$ Y $\in$ F and X $\rightarrow$ Z $\in$ G.

Proof. Assume to the contrary that Y $\neq$ Z and that $\exists$ A $\in$ Y such that A $\notin$ Z; a symmetric argument can be used if we assume that $\exists$ A $\in$ Z such that A $\notin$ Y. It follows by Lemma 10 that there is a unique FD V $\rightarrow$ U $\in$ G such that A $\in$ U. Moreover, by Proposition 4 X $\rightarrow$ V $\in$ (G $-$ {V $\rightarrow$ U})$^+$ and by the intersection property X $\cap$ V $\rightarrow$ A $\in G^+$.

We claim that V $\subset$ X holds. Firstly, if X $\subseteq$ V, then G is not minimum, since X $\rightarrow$ Z and V $\rightarrow$ U can be replaced by X $\rightarrow$ UZ. Secondly, if V $\not\subseteq$ X, then $|$X $\cap$ V$|$ < $|$V$|$. Furthermore, by Proposition 4 X $\cap$ V $\rightarrow$ V $\in$ (G $-$ {V $\rightarrow$ U})$^+$, since by Lemma 10 there can be only one FD in G with A in its right-hand side. This contradicts the fact that G is minimum, since we can replace V $\rightarrow$ U $\in$ G by X $\cap$ V $\rightarrow$ U without changing G^+. Therefore, the claim that V $\subset$ X follows.

It now follows that V $\rightarrow$ X $\in$ (F $-$ {X $\rightarrow$ Y})$^+$, since by Lemma 10 X $\rightarrow$ Y $\in$ F is the only FD in F with A in its right-hand side. This contradicts the fact that F is minimum, since we can replace X $\rightarrow$ Y $\in$ F by V $\rightarrow$ Y without changing F^+. The result that Y = Z follows. □

LEMMA 13. Let F and G be two sets of FDs over R, which are minimum, satisfy the intersection property and are covers of each other. Then for any distinct FDs, X $\rightarrow$ Y $\in$ F and W $\rightarrow$ Z $\in$ G, either Y $\subseteq$ W and Z $\subseteq$ X, or Y $\cap$ W = $\emptyset$ and X $\cap$ Z = $\emptyset$. Moreover, X $\neq$ W and Y $\cap$ Z = $\emptyset$.

Proof. Suppose that $\exists$ A $\in$ Y $\cap$ W; a symmetric argument can be used if we assume that $\exists$ A $\in$ X $\cap$ Z. (We prove the result by contradiction assuming that the statement of the theorem is false.)

Firstly, assume that Y $\subseteq$ W and thus $\exists$ B $\in$ Z$-$X, since it follows that Z $\not\subseteq$ X. It follows by the pseudo-transitivity inference rule that X(W$-$A) $\rightarrow$ B $\in G^+$ and thus by the intersection property it follows that (W$-$A) $\rightarrow$ B $\in G^+$, since A $\notin$ X. Thus on using Proposition 4 and Lemma 10 it follows that (W$-$A) $\rightarrow$ W $\in$ (G $-$ {W $\rightarrow$ Z})$^+$, since otherwise we can replace W $\rightarrow$ Z $\in$ G by W $\rightarrow$ Z$-$B without changing G^+. However, this also contradicts the fact that G is minimum, since we can replace W $\rightarrow$ Z $\in$ G by (W$-$A) $\rightarrow$ Z without changing G^+.

Secondly, assume that Y $\not\subseteq$ W and thus $\exists$ C $\in$ Y$-$W. If it is also the case that Z $\not\subseteq$ X, then the result follows by an argument identical to that made in the first case. So assume that Z $\subseteq$ X and thus $\exists$ D $\in$ X $\cap$ Z. By an argument analogous to that of the first case it follows that (X$-$D) $\rightarrow$ X $\in$ (F $-$ {X $\rightarrow$ Y})$^+$, since otherwise we can replace X $\rightarrow$ Y $\in$ F by X $\rightarrow$ Y$-$C without changing F^+. However, this also contradicts the fact that F is minimum, since we can replace X $\rightarrow$ Y $\in$ F by (X$-$D) $\rightarrow$ Y without changing F^+.

By Lemma 12 it now follows that X $\neq$ W, since X $\rightarrow$ Y and W $\rightarrow$ Z are distinct FDs. It remains to show that Y $\cap$ Z = $\emptyset$.

Suppose that Y $\subseteq$ W and Z $\subseteq$ X but Y $\cap$ Z $\neq$ $\emptyset$. It follows that Y $\cap$ Z $\subseteq$ X contradicting the fact that F is minimum, since F^+ = ((F $-$ {X $\rightarrow$ Y}) $\cup$ {X $\rightarrow$ Y$-$Z })$^+$.

Next, suppose that Y $\cap$ W = $\emptyset$ and X $\cap$ Z = $\emptyset$ but that Y $\cap$ Z $\neq \emptyset$. Let A $\in$ Y $\cap$ Z and thus A $\notin$ XW. It follows by the intersection property that (X $\cap$ W) $\rightarrow$ A $\in G^+$, with |X $\cap$ W| < |X| or |X $\cap$ W| < |W|, since X $\neq$ W. Let V = X $\cap$ W and assume without loss of generality that |V| < |W|.

We claim that V $\rightarrow$ A $\in$ (G $-$ {W $\rightarrow$ Z})$^+$. Suppose that this is not the case. Then, by Proposition 4 and Lemma 10, it follows that V $\rightarrow$ W $\in$ (G $-$ {W $\rightarrow$ Z})$^+$. Consequently, G is not minimum, since we can replace W $\rightarrow$ Z $\in$ G by V $\rightarrow$ Z without changing G^+. Thus the claim is true, but then we obtain a contradiction to the fact that G is minimum, since we can replace W $\rightarrow$ Z by W $\rightarrow$ Z$-$A without changing G^+. It follows that Y $\cap$ Z = $\emptyset$ as required. □

Lemma 14. Let F and G be two sets of FDs over R, which are minimum, satisfy the intersection property and are covers of each other, and also let X $\rightarrow$ Y $\in E_F$(X) and W $\rightarrow$ Z $\in E_G$(X) be distinct FDs. Then X and W are incomparable, Y $\subseteq$ W, Z $\subseteq$ X and Y $\cap$ Z = $\emptyset$.

Proof. We first prove that X and W are incomparable. By Lemma 12 X $\neq$ W, since X $\rightarrow$ Y and W $\rightarrow$ Z are distinct FDs. So assume that X $\subset$ W; a symmetric argument can be used if we assume that W $\subset$ X. Now, since X and W are equivalent under G^+, it follows from Proposition 4 that X $\rightarrow$ W $\in$ (G $-$ {W $\rightarrow$ Z})$^+$. This contradicts the fact that G is minimum, since we can replace W $\rightarrow$ Z $\in$ G by X $\rightarrow$ Z without changing G^+.

It remains to show that Y $\subseteq$ W and that X $\subseteq$ Z, since by Lemma 13 this implies that Y $\cap$ Z = $\emptyset$. By Lemma 13 it is sufficient to show that either Y $\cap$ W $\neq \emptyset$ or X $\cap$ Z $\neq \emptyset$. Assume that Y $\cap$ W = $\emptyset$ and that X $\cap$ Z = $\emptyset$. Let A $\in$ Z. It follows that X $\rightarrow$ A $\in G^+$ is a nontrivial FD, since X $\rightarrow$ W $\in G^+$, and thus (X $\cap$ W) $\rightarrow$ A $\in G^+$, due to the fact that G satisfies the intersection property. Moreover, (X $\cap$ W) $\subset$ X and (X $\cap$ W) $\subset$ W, since X and W were shown to be incomparable. It follows by Proposition 4 that there exists an FD V $\rightarrow$ U $\in$ G, such that (X $\cap$ W) $\rightarrow$ V $\in G^+$, with V $\rightarrow$ A $\in G^+$.

We claim that V $\rightarrow$ A $\in$ (G $-$ {W $\rightarrow$ Z})$^+$. Suppose that this is not the case. Then by Proposition 4 V $\rightarrow$ W $\in$ (G $-$ {W $\rightarrow$ Z})$^+$ and therefore V $\rightarrow$ Z $\in G^+$ and also (X $\cap$ W) $\rightarrow$ Z $\in G^+$. It follows that G is not minimum, since we can replace W $\rightarrow$ Z $\in$ G by (X $\cap$ W) $\rightarrow$ Z without changing G^+. The result that A $\notin$ Z now follows, since otherwise G is not minimum, because we can replace W $\rightarrow$ Z $\in$ G by W $\rightarrow$ (Z$-$A) without changing G^+. □

Theorem 15. If a set of FDs F is minimum and satisfies the intersection property then it is also optimum.

Proof. By Lemma 7 it is sufficient to show that all the equivalence classes of F are optimum. Assume that E_F(X) is not optimum for some X $\subseteq$ sch(R), where E_F(X) is nonempty. Thus, by Lemma 8, there exists an optimum cover G of F in which an FD X $\rightarrow$ Y $\in E_F$(X) has been replaced by some other distinct FD W $\rightarrow$ Z $\in F^+-$ F, with W $\rightarrow$ Z $\in E_G$(X). Moreover, |W| + |Z| < |X| + |Y|.

By Lemma 6 G is minimum, therefore by Lemma 14 it follows that X and W are incomparable, Y $\subseteq$ W, Z $\subseteq$ X and Y $\cap$ Z = $\emptyset$. (Since F satisfies the intersection property, so does G.)

By Proposition 4, X $\rightarrow$ W $\in$ (G $-$ {W $\rightarrow$ Z})$^+$. Therefore, X $\rightarrow$ Y $\in$ (G $-$ {W $\rightarrow$ Z})$^+$, since Y $\subseteq$ W. Let A $\in$ Y, and thus A $\notin$ Z. Hence for some FD V $\rightarrow$ U $\in$ G, which is distinct from W $\rightarrow$ Z, we have that A $\in$ U. Furthermore, V $\neq$ X, since otherwise by Lemma 12 U = Y contradicting the fact that X $\rightarrow$ Y $\notin$ G. There are three cases to consider.

Firstly, assume that X $\subset$ V. It follows by Proposition 4 that X $\rightarrow$ V $\in$ (G $-$ {V $\rightarrow$ U})$^+$, since by Lemma 10 V $\rightarrow$ U is the only FD in G with A in its right-hand side. This contradicts the fact that G is optimum, since we can replace V $\rightarrow$ U $\in$ G by X $\rightarrow$ U without changing G^+.

Secondly, assume that V $\subset$ X. It follows by Proposition 4 that V $\rightarrow$ X $\in$ (F $-$ {X $\rightarrow$ Y})$^+$, since by Lemma 10 X $\rightarrow$ Y is the only FD in F with A in its right-hand side. This contradicts the fact that F is minimum, since we can replace X $\rightarrow$ Y $\in$ F by V $\rightarrow$ Y without changing F^+.

Thirdly, assume that X and V are incomparable. Thus by the intersection property X $\cap$ V $\rightarrow$ A $\in$ F^+, with |X $\cap$ V| $<$ |X|. It follows by Proposition 4 that X $\cap$ V $\rightarrow$ X $\in$ (F $-$ {X $\rightarrow$ Y})$^+$, since by Lemma 10 X $\rightarrow$ Y is the only FD in F with A in its right-hand side. This contradicts the fact that F is minimum, since we can replace X $\rightarrow$ Y $\in$ F by X $\cap$ V $\rightarrow$ Y without changing F^+. It follows that E_F(X) is optimum as required. □

The following corollary follows from Theorem 15 on using the result in [Maier 1980], namely that a minimum cover of a set of FDs F over R can be found in polynomial time in the size of F. The result is important, since it was shown in [Maier 1980] that finding an optimum cover of a set of FDs is, in general, an NP-complete problem.

Corollary 16. An optimum cover of a set of FDs F over R that satisfies the intersection property can be found in polynomial time in the size of F. □

We next show that there exists only one optimum cover of a set of FDs, which satisfies the intersection property. Therefore, by Theorem 15 a set of FDs which satisfies the intersection property has only one minimum cover.

Theorem 17. A set of FDs F, which satisfies the intersection property, has a unique optimum cover.

Proof. Suppose to the contrary that F has two distinct optimum covers G and H and thus E_G(X) $\neq$ E_H(X), for some X $\subseteq$ sch(R). It follows that there is an FD X $\rightarrow$ Y $\in$ E_G(X) which is not in H.

Let A $\in$ Y; A $\notin$ X due to the fact that X $\cap$ Y = $\emptyset$. Therefore, X $\rightarrow$ A $\in$ H^+ and by Lemma 10 there exists an FD W $\rightarrow$ Z $\in$ H which is distinct from X $\rightarrow$ Y and such that A $\in$ Z; A $\notin$ W due to the fact that W $\cap$ Z = $\emptyset$. This leads to a contradiction, since by Lemma 13 Y $\cap$ Z = $\emptyset$ and thus A $\notin$ Z. □

The next corollary follows from Theorem 15, Lemma 10 and the fact that if X $\rightarrow$ Y is an FD over R, with X $\cap$ Y = $\emptyset$, then |X| + |Y| $\leq$ type(R).

Corollary 18. If a set of FDs F over R is optimum and satisfies the intersection property, then |F| $\leq$ type(R) and ||F|| $\leq$ (type(R))^2. □

6.2 LR-Minimum covers of sets of FDs that satisfy the split-freeness property have optimum left-hand sides

The next example shows that the analogue of Theorem 15 does *not* hold for sets of FDs that satisfy the split-freeness property.

EXAMPLE 2. Let R be a relation schema, with sch(R) = {A, B, C, D} and let F = {AD → BC, BD → A, CD → A}. It can easily be verified that F is minimum and satisfies the split-freeness property. On the other hand, F is not optimum, since G = {AD → B, BD → C, CD → A} is an optimum cover of F with ||G||<||F||.

We nevertheless establish the weaker result that sets of FDs, which satisfy the split-freeness property, have the desirable property that such minimum sets of FDs have optimum left-hand sides. In addition, we show that an optimum cover of a set of FDs that satisfies the split-freeness property can be found in polynomial time in the size of F using the notion of an *annual* cover [Maier 1983]. Such an optimum cover is not, in general, unique.

LEMMA 19. Let F be a set of FDs that satisfies the split-freeness property and let XB → A, YA → B ∈ F^+ be cyclic and reduced FDs. Then X = Y.

Proof. We prove the result by contradiction. Suppose that X ≠ Y. There are two cases to consider. Firstly, assume that X ⊂ Y; a symmetric argument can be used if we assume that Y ⊂ X. Therefore, Y ≠ ∅ and X ∩ Y ⊂ Y. It follows that (X ∩ Y)A → B ∉ F^+ and Y → B ∉ F^+, since YA → B is reduced. Secondly, assume that X and Y are incomparable. Therefore X, Y ≠ ∅ and X ∩ Y ⊂ Y. As in the first case it follows that (X ∩ Y)A → B ∉ F^+ and Y → B ∉ F^+, since YA → B is reduced. □

LEMMA 20. Let F and G be two sets of FDs over R, which are minimum, satisfy the split-freeness property and are covers of each other, and also let X → Y ∈ E_F(X) and W → Z ∈ E_G(X) be distinct FDs. Then either X = W or X and W are incomparable, and in addition |X| = |W|.

Proof. Suppose that X ≠ W and assume without loss of generality that X ⊂ W; a symmetric argument can be used if we assume that W ⊂ X. It follows that X → Z ∈ G^+, since W → Z ∈ E_G(X) and G is a cover of F, thus contradicting the fact that G is minimum. So X and W must be incomparable.

It remains to show that |X| = |W| assuming that X and W are incomparable. Suppose that |X| ≠ |W| and assume without loss of generality that |X| < |W|; a symmetric argument can be used if we assume that |W| < |X|. Let V = X ∩ W. Thus for some A ∈ X−W, for some V_1 ⊆ V and for some W_1 ⊆ W−X, we have that V_1W_1 → A ∈ G^+ is reduced with $|W_1| > 1$; otherwise G would not be minimum, since W → Z can be replaced by U → Z, for some U ⊂ W. Correspondingly it follows that for some B ∈ W−X, for some V_2 ⊆ V and for some X_1 ⊆ X−W, we have that V_2X_1 → B ∈ F^+ is reduced with A ∈ X_1; otherwise F would not be minimum, since X → Y can be replaced by U → Y, for some U ⊆ X−A. There are two cases to consider.

Firstly, B ∈ W_1 and thus by Lemma 19 V_1W_1A = V_2X_1B, since F and G are covers of each other. It follows that $V_1 = V_2$ and that W_1−B = X_1−A, since

$V_1 \cap X_1 = \emptyset$ and $V_2 \cap W_1 = \emptyset$. Furthermore, we have that $X_1 -$ A $\neq \emptyset$, since $|W_1| > 1$ thus leading to a contradiction of the fact that $W_1 \cap X_1 = \emptyset$. The claim that |X| = |W| follows.

Secondly, B $\notin W_1$ and thus B $\in$ (W−X) $-W_1$. It follows by the pseudo-transitivity inference rule that $V_1V_2W_1(X_1-$ A) $\rightarrow$ B $\in G^+$; assume without loss of generality that this FD is reduced. We conclude the result by induction on k = $|X_1-$ A|. In the basis step when k = 0 the result follows, since W $\rightarrow$ Z $\in$ G can be replaced by W−B $\rightarrow$ Z without changing G^+, thus contradicting the fact that G is minimum.

In the induction step when k > 0, we have that X_1−A $\neq \emptyset$; let $A' \in X_1$−A. It follows that for some $W_2 \subseteq$ W−X and for some $V_3 \subseteq$ V, $V_3W_2 \rightarrow A' \in G^+$ is reduced.

If B $\in W_2$ then by Lemma 19 $V_1V_2 = V_3$, and in addition we have that W_2−B = $W_1(X_1$−A$A')$ implying that X_1 = AA' and that W_2−B = W_1, in which case $V_1V_2W_2 \rightarrow A' \in F^+$ and $V_2X_1 \rightarrow$ B $\in F^+$ are reduced and cyclic FDs, and thus by Lemma 19 we conclude that $X_1 = A'$, due to the fact that $X_1 \cap W_2 = \emptyset$. It follows that $|X_1| = 1$, which contradicts the fact that k > 0.

On the other hand, if B $\notin W_2$ then by the pseudo-transitivity inference rule V$W_1W_2(X_1-$ A$A')$ $\rightarrow$ B $\in G^+$. The result follows by inductive hypothesis, since $|X_1-$ A$A'| <$ k. The claim that |X| = |W| now follows. □

Theorem 21. Let F be a set of FDs that is minimum and satisfies the split-freeness property. Then F has optimum left-hand sides.

Proof. Let G be a minimum cover of F. Then G also satisfies the split-freeness property, since by Definition 11 the split-freeness property is cover invariant. Let X $\rightarrow$ Y $\in$ F. Then by Lemma 20, $\forall$ W $\rightarrow$ Z $\in E_F$(X) $\cup$ E_G(X), |W| = |X|. The result follows, since by Lemma 8 $|E_F$(X)| = $|E_G$(X)|. □

Lemma 22. Let F be a set of FDs that is minimum and satisfies the split-freeness property, and let X $\rightarrow$ Y, W $\rightarrow$ Z be distinct FDs in E_F(X). Then X and W are incomparable and |X $\cap$ W|=|X| −1 =|W| −1.

Proof. Obviously X $\neq$ W, since F is minimum, so by Lemma 20 X and W are incomparable. We claim that $\exists$ A $\in$ W−X, such that X $\rightarrow$ A $\in F^+$ is a reduced FD; a symmetric argument can be used to show that $\exists$ B $\in$ X−W such that W $\rightarrow$ B $\in F^+$ is a reduced FD. Suppose to the contrary that $\forall$ A $\in$ W−X, X $\rightarrow$ A $\in F^+$ is *not* reduced. It follows by Lemma 9 that Y $\cap$ W = $\emptyset$. Let A $\in$ W−X and V $\rightarrow$ A $\in F^+$ be a reduced FD, with V $\subset$ X. There are two cases to consider. Firstly, V $\rightarrow$ X $\in F^+$, and thus by Proposition 4 V $\rightarrow$ X $\in$ (F − {X $\rightarrow$ Y}$)^+$. The result follows, since X $\rightarrow$ Y can be replaced by V $\rightarrow$ Y without changing F^+ contradicting the fact that F is minimum. Secondly, V $\rightarrow$ X $\notin F^+$, and therefore V $\rightarrow$ A $\in$ (F − {X $\rightarrow$ Y}$)^+$. We conclude that X $\rightarrow$ W $\in$ (F − {X $\rightarrow$ Y}$)^+$, and thus X $\rightarrow$ Y can be removed from F and W $\rightarrow$ Z can be replaced by W $\rightarrow$ YZ, contradicting the fact that F is minimum.

The result that |X $\cap$ W|=|X| −1 follows by Lemma 19, since $\exists$ A $\in$ W−X, $\exists$ B $\in$ X−W such that X $\rightarrow$ A, W $\rightarrow$ B $\in F^+$ are cyclic and reduced FDs. □

We next define the notion of an *annual* cover of a set of FDs that satisfies the split-freeness property; the name originates from [Maier 1983] (cf. *circular* set of FDs [Atzeni and De Antonellis 1993]).

DEFINITION 18. Let F be a set of FDs that is minimum and satisfies the split-freeness property and let $E_F(\mathrm{X}) = \{X_1 \rightarrow Y_1, \ldots, X_n \rightarrow Y_n\}$ be a nonempty equivalence class of F for some X $\subseteq$ sch(R). We denote the set of attributes $\bigcup_{i=1}^{n} Y_i - \bigcup_{i=1}^{n} X_i$ by ACY(X) and the set of attributes $\bigcup_{i=1}^{n} Y_i \cap \bigcup_{i=1}^{n} X_i$ by CYC(X) assuming that F is understood from context.

The set of FDs F is *annual* if each nonempty equivalence class $E_F(\mathrm{X})$ of F has the form, $\{X_1 \rightarrow \mathrm{Y}A_1, \ldots, X_n \rightarrow A_n\}$, where Y = ACY(X). In addition, if $n > 1$ then $\forall i \in I, A_i \in \mathrm{CYC(X)}{-}X_i$, otherwise, if $n = 1$ then $A_1 = \emptyset$, i.e. $E_F(X)$ has the form $\{X_1 \rightarrow \mathrm{Y}\}$.

LEMMA 23. Let F be a set of FDs that is minimum and satisfies the split-freeness property, and let B $\in$ ACY(X) for some X $\subseteq$ sch(R). Then there does not exist a reduced FD WB $\rightarrow$ A $\in F^+$ such that ZA $\in$ LHS(X) for some Z $\subseteq$ sch(R).

Proof. Suppose to the contrary that there exists a reduced FD WB $\rightarrow$ A $\in F^+$ with ZA $\in$ LHS(X) for some Z $\subseteq$ sch(R). Then by a similar argument to that made in the proof of Lemma 9 we have that ZA $\rightarrow$ B $\in F^+$ is reduced. Therefore WB $\rightarrow$ A $\in F^+$ and ZA $\rightarrow$ B $\in F^+$ are reduced and cyclic FDs and thus by Lemma 19 we have that W = Z. It follows that WB $\in$ LHS(X) and thus B $\in$ CYC(X) leading to a contradiction of the assumption that B $\in$ ACY(X). □

LEMMA 24. Let F be a set of FDs which satisfies the split-freeness property. Then an annual cover of F can be obtained in polynomial time in the size of F.

Proof. By Theorem 3 in [Maier 1980] a minimum cover of F can be computed in polynomial time in the size of F. So assume that F is minimum and let $E_F(\mathrm{X}) = \{X_1 \rightarrow Y_1, \ldots, X_n \rightarrow Y_n\}$ $(n > 1)$ be a nonempty equivalence class of F for some X $\subseteq$ sch(R); if $n = 1$ then the result holds trivially. We next proceed to construct an annual cover G of F by considering the equivalence classes of the latter; let Y = ACY(X). Let $E_G(\mathrm{X}) = \{X_1 \rightarrow \mathrm{Y}(X_2 - X_1), X_2 \rightarrow (X_3 - X_2), \ldots, X_n \rightarrow (X_1 - X_n)\}$. By Lemma 22 we have that $\forall i \in I$, $|X_{i+1} - X_i| = 1$, taking $n+1$ to be 1. Therefore, $E_G(\mathrm{X})$ can be rewritten as $\{X_1 \rightarrow \mathrm{Y}A_1, \ldots, X_n \rightarrow A_n\}$ such that $\forall i \in I$, $A_i = X_{i+1} - X_i$ and $A_i \notin X_i$. It only remains to show that G is a minimum cover of F.

It can be verified that $E_G(\mathrm{X})^+ = E_F(\mathrm{X})^+$ and thus G is a cover of F. Furthermore, all the FDs in $E_G(\mathrm{X})$ are reduced; this can be proved by using a similar argument to that given in the proof of Lemma 9. Let us denote X_1 by X and A_1 by A. We claim that for every Z $\subset$ YA, $((\mathrm{G} - \{\mathrm{X} \rightarrow \mathrm{YA}\}) \cup \{\mathrm{X} \rightarrow \mathrm{Z}\})^+ \neq G^+$. Assume that $((\mathrm{G} - \{\mathrm{X} \rightarrow \mathrm{YA}\}) \cup \{\mathrm{X} \rightarrow \mathrm{Z}\})^+ = G^+$. There are two cases to consider.

Firstly, A $\notin$ Z. There are two subcases to consider. In the first subcase X $\rightarrow$ A $\in$ $(\mathrm{F} - E_F(\mathrm{X}))^+$ leading to a contradiction of the fact that F is minimum, since we can replace $E_F(\mathrm{X})$ by $\{X_2 \rightarrow \mathrm{Y}A_2, \ldots, X_n \rightarrow A_n\}$ without changing F^+. In the second subcase X $\rightarrow$ A $\notin$ $(\mathrm{F} - E_F(\mathrm{X}))^+$ and thus due to our assumption $\exists$ B $\in$

Y such that for some W $\subset$ sch(R), WB $\rightarrow$ A $\in F^+$ is reduced and WB $\notin$ LHS(X) leading to a contradiction of Lemma 23, since A $\in$ CYC(X).

Secondly, A $\in$ Z. Let B $\in$ Y$-$Z and therefore $\exists i \in I$ such that B $\in Y_i$. In view of our assumption it follows that ((F $-$ $\{X_i \rightarrow Y_i\}$) $\cup$ $\{X_i \rightarrow Y_i$$-B\}$)$^+$ = F^+ leading to a contradiction of the fact that F is minimum. The result now follows by Definition 18. $\square$

The following theorem is the companion result to Theorem 15.

THEOREM 25. If a set of FDs F is annual then it is also optimum.

Proof. Let #LHS(F) denote the sum, $\sum_{X \rightarrow Y \in F} \mid X \mid$, and #RHS(F) denote the sum, $\sum_{X \rightarrow Y \in F} \mid Y \mid$. By Theorem 21, #LHS(F) = #LHS(G), where G is an optimum cover of F. It remains to show that #RHS(F) = #RHS(G).

By Lemma 7 F is optimum if and only if all the equivalence classes of F are optimum. Let E_F(X) be an equivalence class of F for some X $\subseteq$ sch(R). By Lemma 24 we can assume that E_F(X) has the form $\{X_1 \rightarrow$ Y$A_1, \ldots, X_n \rightarrow A_n\}$ with Y = ACY(X). Thus we need to show that #RHS(E_F(X)) = #RHS(E_G(X)).

Assume that E_F(X) is not optimum and that E_G(X) = $\{X_1 \rightarrow$ W$A_1, \ldots, X_n \rightarrow A_n\}$ is an optimum equivalence class of G, with |W| < |Y|. There are two cases to consider. Firstly, assume that W $\subset$ Y. It follows that F is not minimum, since we can replace $X_1 \rightarrow$ Y$A_1 \in E_F$(X) by $X_1 \rightarrow$ WA_1 without changing F^+.

Secondly, assume that W and Y are incomparable and that W_1 = W$-$Y and Y_1 = Y$-$W. It follows that $|W_1| < |Y_1|$. We claim that for some attribute B $\in W_1$ and distinct attributes $C_1, C_2 \in Y_1$, there exists a nontrivial and reduced FD Z $\rightarrow$ B $\in F^+$ such that $C_1, C_2 \in$ Z.

Let H = F $-$ $\{X_1 \rightarrow$ Y$A_1\}$. Obviously, there must be a nontrivial and reduced FD Z $\rightarrow$ B $\in F^+$, since $X_1 \rightarrow$ B $\in F^+$. Now, if for all nontrivial and reduced FDs Z $\rightarrow$ B $\in F^+$, Z $\cap$ $Y_1 = \emptyset$, then $X_1 \rightarrow$ B $\in$ (H $\cup$ $\{X_1 \rightarrow A_1\}$)$^+$. It follows that we can replace $X_1 \rightarrow$ W$A_1 \in E_G$(X) by $X_1 \rightarrow$ (W$-$B)A_1 without changing G^+, thus contradicting the fact that E_G(X) is optimum.

Furthermore, since $|W_1| < |Y_1|$, it must be the case that for some attribute B $\in W_1$, there exists a nontrivial and reduced FD Z $\rightarrow$ B $\in F^+$ such that |Z $\cap Y_1$| > 1. If this is not the case then for some C $\in Y_1$, $X_1 \rightarrow W_1 \in$ (H $\cup$ $\{X_1 \rightarrow$ (Y$-$C)$A_1\}$)$^+$. It then follows that we can replace $X_1 \rightarrow$ Y$A_1 \in E_F$(X) by $X_1 \rightarrow$ (Y$-$C)A_1 without changing F^+, thus contradicting the fact that F is minimum. The claim follows and thus we can assume without loss of generality that $C_1, C_2 \in$ Z. In fact, we can strengthen the above claim by asserting that for some attribute B $\in W_1$ and distinct attributes $C_1, C_2 \in Y_1$ it must be the case that for all nontrivial and reduced FDs in F^+ of the form U $\rightarrow$ B we have that $C_1, C_2 \in$ U. Otherwise, by the above argument F would not be minimum.

Since B $\in W_1$, it must be the case, in view of the above argument, that $C_1, C_2 \in X_1$. Consequently, we can replace $X_1 \rightarrow$ Y$A_1 \in E_F$(X) by $X_1 \rightarrow$ (Y $-C_1C_2$)A_1, thus contradicting the fact that F is minimum. It therefore follows that |W|=|Y| and by Lemma 8 $|E_F$(X)$|$ = $|E_G$(X)$|$, so F is optimum as required. $\square$

The following corollary is the companion result to Corollary 16.

COROLLARY 26. An optimum cover of a set of FDs F over R that satisfies the split-freeness property can be found in polynomial time in the size of F.

Proof. The result is a direct consequence of Lemma 24 and Theorem 25. □

The next example shows that a result corresponding to Theorem 17 does not hold when F satisfies the split-freeness property; that is, F may have more than one optimal cover.

EXAMPLE 3. Let R be a relation schema, with sch(R) = $\{A, B, C\}$ and let F = $\{A \to C, B \to C, C \to B\}$. It can easily be verified that F is optimum and that F satisfies the split-freeness property. Now, let G = $\{A \to B, B \to C, C \to B\}$ and thus G $\neq$ F. It can also easily be verified that G is a cover of F and that G is also optimum.

6.3 The structure of monodependent sets of FDs which are optimum

Herein we give some more technical results which help us to understand the structure of monodependent sets of FDs which are also optimum. From now on we will assume, unless it is stated otherwise, that F *is a set of monodependent FDs over* R.

LEMMA 27. Let F be monodependent and optimum, and $X \to Y$, $W \to Z$ be distinct FDs in E_F(X), for some X $\subseteq$ sch(R). Then $|X|=|W|$, $|X \cap W|=|X| -1$ and $|Y|=|Z|= 1$.

Proof. By Lemma 20 $|X|=|W|$ and by Lemma 22 $|X \cap W|=|X| -1$. Assume to the contrary that $|Y|> 1$; a symmetric argument can be used if we assume that $|Z|> 1$. By Lemma 14, we have that Y $\subseteq$ W. The result follows, since we arrive at the contradictory conclusion that $|X \cap W|\leq|X| -2$, due to the fact that $X \cap Y = \emptyset$. □

The following corollary is an immediate consequence of Lemma 14 and Lemma 27.

COROLLARY 28. Let F be monodependent and optimum, and $X \to Y$, $W \to Z$ be FDs in E_F(X), for some X $\subseteq$ sch(R). Then XY = WZ. □

LEMMA 29. Let F be monodependent and optimum and let $X \to Y$, $W \to Z \in$ F be FDs such that $E_F(X) \neq E_F(W)$. Then

(1) $X \neq W$ and $Y \cap Z = \emptyset$.
(2) $Y \cap W = \emptyset$ and $X \cap Z = \emptyset$.

Proof. The first part follows from Lemma 13. For the second part we show that $Y \cap W = \emptyset$; a symmetric argument can be used to show that $X \cap Z = \emptyset$. Assume to the contrary that $A \in Y \cap W$. By Lemma 13 it follows that Y $\subseteq$ W and Z $\subseteq$ X.

We now claim that it is also the case that X and W are incomparable. We already have that $X \neq W$ by the first part. So assume that $X \subset W$; a symmetric argument can be used if we assume that $W \subset X$. It follows that $Z \subset W$, since $Z \subseteq X$. Thus $W \to Z$ is a trivial FD, contradicting the fact that F is optimum.

So far, on assuming that A $\in$ Y $\cap$ W, we have that Y $\subseteq$ W, Z $\subseteq$ X and that X and W are incomparable. Thus $\exists$ A $\in$ Y $\cap$ W, $\exists$ B $\in$ X $\cap$ Z, such that X $\rightarrow$ A and W $\rightarrow$ B $\in F^+$ are cyclic and reduced FDs. Therefore by Lemma 19 we can rewrite X $\rightarrow$ A as VB $\rightarrow$ A and W $\rightarrow$ B as VA $\rightarrow$ B. However, this contradicts the fact that E_F(X) $\neq$ E_F(W). Therefore, by Lemma 13 it must be the case that Y $\cap$ W = $\emptyset$ and that X $\cap$ Z = $\emptyset$ as required. □

We close this subsection by showing that sets of FDs which are monodependent and optimum are closed under the proper subset operation.

LEMMA 30. Let F be a monodependent and optimum set of FDs. Then $\forall$ G $\subset$ F, G is a monodependent and optimum set of FDs over R.

Proof. Let G $\subset$ F. By Lemmas 29 and 27, and Corollary 28 we can deduce that X $\rightarrow$ A $\in F^+$ and A $\in$ Z for some FD W $\rightarrow$ Z $\in$ G if and only if X $\rightarrow$ A $\in G^+$. We call this statement Observation 1. Therefore, G must satisfy the intersection property, since otherwise there must exist incomparable FDs X $\rightarrow$ A, Y $\rightarrow$ A $\in G^+$, but X $\cap$ Y $\rightarrow$ A $\in F^+ - G^+$, which contradicts Observation 1. Similarly, G must satisfy the split-freeness property, since otherwise there must exist cyclic FDs, XB $\rightarrow$ A, YA $\rightarrow$ B $\in G^+$, but either Y $\rightarrow$ B $\in F^+ - G^+$ or (X $\cap$ Y)A $\rightarrow$ B $\in F^+ - G^+$, which again contradicts Observation 1.

Next, suppose that G is not optimum and that H is an optimum cover of G, with ||H||<||G||. Therefore, ((F $-$ G) $\cup$ H)$^+$ = F^+. This leads to a contradiction that F is optimum, since ||(F $-$ G) $\cup$ H||<||F||. The result that G is optimum follows. □

7. KEYS OF MONODEPENDENT SETS OF FDS

Herein we characterise the set of keys of a monodependent set of FDs. We also show that two problems that are in general NP-complete, i.e. the superkey of cardinality k problem and the prime attribute problem [Lucchesi and Osborn 1978], are solvable in polynomial time when the set of FDs is monodependent.

DEFINITION 19. A set of attributes X $\subseteq$ sch(R) is a *superkey* for R with respect to F (or simply a superkey for R whenever F is understood from context), if X $\rightarrow$ sch(R) $\in F^+$. A set of attributes X $\subseteq$ sch(R) is a *key* for R with respect to F (or simply a key for R whenever F is understood from context), if X is a superkey for R with respect to F and, in addition, for no proper subset Y $\subset$ X, is it the case that Y is a superkey for R with respect to F. A key X is *simple* if |X|= 1. We denote the set of all keys for R with respect to F by KEYS(F).

An attribute A $\in$ sch(R) is *prime* with respect to F (or simply prime whenever F is understood from context) if A $\in$ X for some X $\in$ KEYS(F); otherwise A is *nonprime* with respect to F.

An *antikey* for R with respect to F (or simply an antikey for R whenever F is understood from context) is a maximal subset X of sch(R) such that X is not a superkey for R. We denote the set of all antikeys for R with respect to F by ANTI(F).

We now investigate the structure of KEYS(F), when F is a monodependent and optimum set of FDs.

Definition 20. The set of *extreme* attributes with respect to a set of FDs F, denoted by EXT(F), is defined by

$$\text{EXT(F)} = \{A \mid \nexists W \rightarrow Z \in F \text{ such that } A \in Z - W\}.$$

It can easily be verified that for all keys, X for R with respect to F, EXT(F) $\subseteq$ X holds.

Theorem 31. Let F be monodependent and optimum, and let E = $\{E_F(X_1), \ldots, E_F(X_k)\}$ be the set of all nonempty equivalence classes of F. Then X $\subseteq$ sch(R) is a key for R if and only if X = EXT(F) $\cup \bigcup_{i=1}^{k} Y_i$, where $\forall i \in \{1, \ldots, k\}, Y_i \rightarrow Z_i \in E_F(X_i)$.

Proof. If. Let V = sch(R)$-$X and suppose that Y = $\bigcup_{i=1}^{k} Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$. We first show that X $\rightarrow$ V $\in F^+$ and thus X $\rightarrow$ sch(R) $\in F^+$. Let A $\in$ V and thus $\exists$ W $\rightarrow$ Z $\in$ F such that A $\in$ Z. Moreover, by the definition of Y, E_F(W) = $E_F(Y_i)$ for some $Y_i \subseteq$ Y $\subseteq$ X. It follows that $Y_i \rightarrow$ A $\in F^+$ and thus X $\rightarrow$ A $\in F^+$ as required.

It remains to show that $\forall$ A $\in$ X, (X$-$A) $\rightarrow$ sch(R) $\notin F^+$. There are two cases to consider. In the first case, when A $\in$ EXT(F), the result follows by the definition of EXT(F), since (X$-$A) $\rightarrow$ A $\notin F^+$. In the second case, when A $\in$ Y$-$EXT(F), assume to the contrary that (X$-$A) $\rightarrow$ A $\in F^+$. Thus A $\in Y_i$ for some $Y_i \subseteq$ Y $\subseteq$ X, with $Y_i \rightarrow Z_i \in E_F(Y_i)$. It follows that (X$-$A) $\rightarrow Z_i \in F^+$ and, since F satisfies the intersection property, it also follows that $(Y_i-$ A$) \rightarrow Z_i \in F^+$, due to the fact that (X$-$A) $\cap\, Y_i = Y_i-$A. (Note that X$-$A $\subset Y_i$ is not possible, otherwise we can derive a contradiction of the fact that F is optimum.)

By Lemma 10, $\forall$ B $\in Z_i$, $Y_i \rightarrow Z_i \in$ F is the unique FD such that B $\in Z_i$. Therefore, by Proposition 4 $(Y_i-$A$) \rightarrow Y_i \in$ (F $-$ $\{Y_i \rightarrow Z_i\})^+$. It follows that F is not optimum, since we can replace $Y_i \rightarrow Z_i \in$ F by $(Y_i-$ A$) \rightarrow Z_i$ without changing F^+. The result that (X$-$A) $\rightarrow$ A $\notin F^+$ follows as required.

Only if. Suppose that X is a key for R. Then obviously EXT(F) $\subseteq$ X. There are three cases to consider. Firstly, if X is a proper superset of EXT(F) $\cup \bigcup_{i=1}^{k} Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$, then by the *if* part of the proof X is a superkey for R but X is not a key for R. Secondly, if X is a proper subset of EXT(F) $\cup \bigcup_{i=1}^{k} Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$, then on using a similar argument to that made in the *if* part we can prove that $(Y_i-$ A$) \rightarrow Z_i \in F^+$, for some A $\in Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$. This leads to a contradiction of the fact that F is optimum. Thirdly, we need to consider the case when X is incomparable with all sets of attributes of the form EXT(F) $\cup \bigcup_{i=1}^{k} Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$. Thus there is some equivalence class $E_F(X_i) \in$ E, such that for all FDs $Y_i \rightarrow Z_i \in E_F(X_i)$, $Y_i \not\subseteq$ X.

We claim that $\forall Y_i \rightarrow Z_i \in E_F(X_i)$, $Z_i \subseteq$ X. Assume to the contrary that for some $Y_i \rightarrow Z_i \in E_F(X_i)$, $Z_i \not\subseteq$ X. Let A $\in Z_i-$X and W = X $\cap Y_i$ and thus |W| $< |Y_i|$. Since F satisfies the intersection property, it follows that W $\rightarrow$ A $\in F^+$ is a nontrivial FD. This leads to a contradiction of the fact that F is optimum on using a similar argument to that used in the *if* part of the proof; in this case we can either replace $Y_i \rightarrow Z_i \in$ F by $Y_i \rightarrow Z_i-$A or we can replace $Y_i \rightarrow Z_i$ by W $\rightarrow Z_i$. (We utilise Lemma 10 and Proposition 4.) Therefore the claim follows.

We conclude that for any FD $Y_i \rightarrow Z_i \in E_F(X_i)$, there exists an attribute C $\in Y_i-$X such that C does not appear on the right-hand side of any FD in $E_F(X_i)$, due to the fact that $Y_i \not\subseteq$ X. Furthermore, by Lemma 29 it follows that C does not appear on the right-hand side of any FD in F. Therefore A $\in$ EXT(F) and a contradiction arises since EXT(F) $\subseteq$ X. The result that X = EXT(F) $\cup \bigcup_{i=1}^{k} Y_i$, where $Y_i \rightarrow Z_i \in E_F(X_i)$, now follows. □

The next two examples show that we cannot relax Theorem 31 so that F satisfies only the intersection property or only the split-freeness property.

EXAMPLE 4. Let F = {AB → C, C → A}, with sch(R) = {A, B, C}. It can be verified that F satisfies the intersection property but F does not satisfy the split-freeness property. Furthermore, it can be verified that the result of Theorem 31 does not hold for F, since AB and BC are the only keys for R and thus ABC is not a key for R.

EXAMPLE 5. Let F = {A → B, C → A}, with sch(R) = {A, B, C}. It can be verified that F satisfies the split-freeness property but F does not satisfy the intersection property. Furthermore, it can be verified that the result of Theorem 31 does not hold for F, since C is the only key for R and thus AC is not a key for R.

COROLLARY 32. Let F be monodependent and optimum, and let E = $\{E_F(X_1), \ldots, E_F(X_k)\}$ be the set of all nonempty equivalence classes of F. Then the number of keys for R is given by

$$| KEYS(F) | = \prod_{i=1}^{k} | E_F(X_i) | . \quad \Box$$

The following corollary, which follows immediately from Corollary 32, characterises exactly when there is a unique key for R with respect to F.

COROLLARY 33. Let F be monodependent and optimum, and let E = $\{E_F(X_1), \ldots, E_F(X_k)\}$ be the set of all nonempty equivalence classes of F. Then |KEYS(F)|= 1 if and only if $\forall E_F(X_i) \in$ E, $|E_F(X_i)| = 1$. □

7.1 The superkey of cardinality k problem for monodependent sets of FDs

The superkey of cardinality k problem, defined below, is known to be NP-complete in the general case [Lucchesi and Osborn 1978]. Herein we show that this problem can be solved in polynomial time in the size of the input set of FDs, F, when F satisfies the split-freeness property but is still NP-complete when F satisfies the intersection property but not necessarily the split-freeness property.

DEFINITION 21. The superkey of cardinality k problem is that of finding a superkey for R, with respect to F, whose cardinality is less than or equal to a natural number k.

The next auxiliary lemma will be useful in proving the subsequent theorem.

LEMMA 34. Let $\mathcal{A}$ be a family of subsets of sch(R) such that $\forall$ X $\in \mathcal{A}$, |X| = type(R)−2. Then a set of FDs F over R can be constructed in polynomial time in the size of F such that ANTI(F) = $\mathcal{A}$ and F satisfies the intersection property.

Proof. Let A $\in$ sch(R) and let lhs(A) denote the set of attributes, {sch(R)$-$XA | X $\in \mathcal{A}$ and A $\notin$ X}, noting that sch(R)$-$XA is a single attribute. By Lemma 13.3 in [Mannila and Räihä 1992] it follows that F = {lhs(A) $\rightarrow$ A | A $\in$ sch(R)} is a reduced set of FDs such that ANTI(F) = $\mathcal{A}$; we observe that if lhs(A) = $\emptyset$, then $\emptyset \rightarrow$ A $\in$ F. It can be verified that the construction of F can be done in polynomial time in the size of F, since there are at most $\binom{type(R)}{\mathrm{type(R)}-2}$ subsets of sch(R) in $\mathcal{A}$.

It remains to show that F satisfies the intersection property. Let Z $\rightarrow$ A, Y $\rightarrow$ B $\in$ F be distinct nontrivial FDs. It is sufficient to show that if A $\in$ Y then B $\in$ Z, since then the only nontrivial FDs of the form W $\rightarrow$ A $\in F^+$ satisfy Z $\subseteq$ W. Now, due to the fact that A $\in$ Y, we have that for some X $\in \mathcal{A}$, B $\notin$ X and sch(R)$-$XB = A. It follows that A $\notin$ X and that sch(R)$-$XA = B. Therefore, by the construction of lhs(A), B $\in$ Z. □

THEOREM 35. The superkey of cardinality k problem is NP-complete, when F is a set of FDs that satisfies the intersection property.

Proof. The problem is known to be in NP [Lucchesi and Osborn 1978]. It remains to show that the problem is NP-hard.

Let $\mathcal{A}$ be a family of incomparable subsets of sch(R). By the results in [Thi 1986] there always exists a set of FDs F over R such that ANTI(F) = $\mathcal{A}$. In addition, $\mathcal{A}$ uniquely determines another family $\mathcal{K}$ of incomparable subsets of sch(R) such that $\mathcal{K}$ = KEYS(F). Moreover, for any set of FDs F such that ANTI(F) = $\mathcal{A}$, ANTI(F) uniquely determines KEYS(F) and vice versa. If X is a superset of a member of $\mathcal{K}$ then we say that X is a superkey for R with respect to $\mathcal{A}$.

By Lemma 2.4 in [Demetrovics and Thi 1988] the vertex cover problem, which is known to be NP-complete [Garey and Johnson 1979], can be reduced to the following problem, which we call the restricted superkey of cardinality k problem for antikeys. Given a family $\mathcal{A}$ of subsets of sch(R) such that $\forall$ X $\in \mathcal{A}$, |X| = type(R)$-$2, does there exist a superkey for R with respect to $\mathcal{A}$ of cardinality k or less. (Intuitively, $\mathcal{A}$ can encode a graph whose vertex set is sch(R), and each X $\in \mathcal{A}$ encodes the edge {A, B}, where sch(R)$-$X = {A, B}.)

By Lemma 34 given $\mathcal{A}$ such that $\forall$ X $\in \mathcal{A}$, |X| = type(R)$-$2, we can construct a set of FDs F in polynomial time in the size of F such that ANTI(F) = $\mathcal{A}$ and F satisfies the intersection property. Furthermore, we now have that X is a superkey for R with respect to F if and only if X is a superkey for R with respect to $\mathcal{A}$. The result now follows, since we have reduced the restricted superkey of cardinality k problem for antikeys to the problem of finding a superkey of cardinality k or less for R with respect F. □

We next show that when F satisfies that split-freeness property all of its keys have equal cardinalities.

THEOREM 36. Let F be a set of FDs which satisfies the split-freeness property. Then the cardinalities of all the keys in KEYS(F) are equal. Moreover, this cardinality can be found in polynomial time in the size of F.

Proof. Let X and W be distinct keys in KEYS(F). Then by Definition 19 X and W are incomparable. The result that |X| = |W| follows by an induction similar to the one employed in the proof of Lemma 20, using the fact that for no proper subset V of X or W, is V a key for R. It follows that the cardinalities of all the keys in KEYS(F) are the same. Next it is easy to find a key in KEYS(F) in polynomial time in the size of F using Algorithm 3.3 in [Atzeni and De Antonellis 1993] (see [Kundu 1985] for another such algorithm); let X be such a key. It follows that the cardinality of all the keys in KEYS(F) is |X|. □

The next example shows that Theorem 36 does not hold if F satisfies the intersection property but not the split-freeness property.

Example 6. Let F = {A → BC, BC → A}, with sch(R) = {A, B, C}. It can be verified that F satisfies the intersection property but F does not satisfy the split-freeness property. It can also be verified that the result of Theorem 36 does not hold for F, since A and BC are the two keys for R.

The following result is immediate from Theorem 36.

Corollary 37. The superkey of cardinality k problem can be solved in polynomial time in the size of F when F satisfies the split-freeness property. □

Although the superkey of cardinality k problem can be solved in polynomial time in the size of F when F satisfies the split-freeness property, as the next example shows KEYS(F) may contain an exponential number of keys.

Example 7. Let R be a relation schema, with type(R) = $2n$ for some natural number n, and where sch(R) = $\{A_1, A_2, \ldots, A_{2n-1}, A_{2n}\}$. In addition, let F = $\{A_i \rightarrow A_{i+1}, A_{i+1} \rightarrow A_i \mid i \in \{1, 3, \ldots, 2n-1\}\}$. It can easily be verified that F is monodependent and optimum. It follows by Corollary 32 that the cardinality of KEYS(F) is 2^n and by Theorem 36 that the cardinality of each of these keys is n.

7.2 The prime attribute problem for monodependent sets of FDs

The prime attribute problem, defined below, is known to be NP-complete in the general case [Lucchesi and Osborn 1978]. Herein we show that this problem can be solved in polynomial time in the size of the input set of FDs, F, when F either satisfies the intersection property or the split-freeness property.

Definition 22. The prime attribute problem is that of deciding whether an attribute A ∈ sch(R) is prime with respect to F.

The following result which shows the connection between the set of nonprime attributes and the set of antikeys was proved in [Demetrovics and Thi 1987, Theorem 3.8] and [Mannila and Räihä 1992, Theorem 13.7].

Lemma 38. An attribute A ∈ sch(R) is nonprime with respect to a set of FDs F if and only if A is a member of the intersection of all the antikeys in ANTI(F). □

The next auxiliary lemma will be used in the proof of Theorem 40.

LEMMA 39. If F is a set of FDs that satisfies the intersection property then $\forall$ X $\in$ ANTI(F), |X|$\geq$ type(R)$-$2.

Proof. Suppose to the contrary that for some X $\in$ ANTI(F), |X|$<$ type(R)$-$2 and let $A_1, A_2, A_3 \in$ sch(R)$-$X. Then, by Definition 19, for any $i, j \in \{1, 2, 3\}$, X$A_i \rightarrow A_k \in F^+$ and also X$A_j \rightarrow A_k \in F^+$ are incomparable FDs, where $i \neq j$ and $k \in \{1, 2, 3\} - \{i, j\}$. By the intersection property and the union rule it follows that X $\rightarrow$ sch(R). Therefore X is a superkey for R contradicting the fact that X is an antikey. The result that |X|$\geq$ type(R)$-$2 follows as required. □

THEOREM 40. The prime attribute problem can be solved in polynomial time in the size of F, when F is a set of FDs that satisfies the intersection property.

Proof. By Lemma 39 and Definition 19 the set of antikeys ANTI(F) for R can be computed in polynomial time in the size of F (see Lemma 13.4 in [Mannila and Räihä 1992]). The result follows by Lemma 38. □

THEOREM 41. The prime attribute problem can be solved in polynomial time in the size of F, when F is a set of FDs that satisfies the split-freeness property.

Proof. Assume without loss of generality that F is an annual set of FDs; if F is not annual, then by Lemma 24 an annual cover of F can be computed in polynomial time in the size of F.

Let ACY(F) denote the set of attributes $\bigcup$ {ACY(X) | X $\subseteq$ sch(R) and E_F(X) $\neq \emptyset$}. In addition, let NPR(F) be the smallest subset of attributes of sch(R) that contains ACY(F) and satisfies the following condition, for some X $\subseteq$ sch(R): CYC(X) $\subseteq$ NPR(F) if $\exists$ W $\in$ LHS(X) such that W $\cap$ CYC(X) $\subseteq$ NPR(F).

The set of attributes NPR(F) can be computed in polynomial time in the size of F, since its cardinality is at most type(R). It remains to show that A is nonprime if and only if A $\in$ NPR(F).

If. Suppose that A $\in$ NPR(F). We need to show that A is nonprime. There are two cases to consider. Firstly, A $\in$ ACY(X) for some X $\subseteq$ sch(R). Assume to the contrary that A is prime. Then by Lemma 38 $\exists$ Y $\in$ ANTI(F) such that A $\notin$ Y. It follows that YA is a superkey for R. Furthermore, $\forall$ V $\in$ LHS(X), V $\not\subseteq$ Y, otherwise Y is a superkey for R and therefore it is not an antikey for R. Let U $\in$ LHS(X) and B $\in$ U such that B $\notin$ Y; A $\neq$ B, since ACY(X) $\cap$ U $= \emptyset$. Now, Y $\rightarrow$ B $\notin F^+$, since Y is an antikey for R, but YA $\rightarrow$ B $\in F^+$. It follows that for some subset Y' of Y, Y'A $\rightarrow$ B $\in F^+$ is reduced. By Lemma 23 we arrive at a contradiction. Thus A is nonprime as required.

Secondly, assume that A $\in$ NPR(F)$-$ACY(F) and that ACY(F) $\subseteq$ Z $\subset$ NPR(F), for some Z $\subseteq$ sch(R). We prove the result by induction on the number of attributes in Z, the hypothesis being that the result holds when Z $\subset$ NPR(F). Now, assume that for some X $\subseteq$ sch(R), with E_F(X) $\neq \emptyset$, $\exists$ W $\in$ LHS(X) such that W $\cap$ CYC(X) $\subseteq$ Z but (CYC(X)$-$W) $\not\subseteq$ Z. It remains to show that if A $\in$ CYC(X)$-$W, where A $\notin$ Z, then A is nonprime. Suppose to the contrary that A is prime. Then by Lemma 38 $\exists$ Y $\in$ ANTI(F) such that A $\notin$ Y. It follows that YA is a superkey for R. Furthermore, by Lemma 38 Z $\subseteq$ Y and thus W $\cap$ CYC(X) $\subseteq$ Y also holds. In order to conclude the proof we show that W$-$CYC(X) $\subseteq$ Y. Let B $\in$ W$-$CYC(X). There are three subcases to consider.

In the first subcase B $\in$ ACY(V) for some V $\subseteq$ sch(R) in which case B $\in$ Y, since B $\in$ Z.

In the second subcase B $\in$ CYC(V) for some V $\subseteq$ sch(R), with V $\notin$ LHS(X). Assume that B is prime, otherwise B $\in$ Y holds by Lemma 38. If B $\notin$ Y then for some $Y' \subseteq$ Y, Y'A $\rightarrow$ B $\in F^+$ is reduced. Now, A $\notin$ W, since A $\in$ CYC(X)$-$W, and thus W $\rightarrow$ A $\in F^+$ is a reduced FD on using an argument similar to the one made in the proof of Lemma 9. Therefore, W $\rightarrow$ A and Y'A $\rightarrow$ B are reduced and cyclic FDs and thus Y'A $\in$ LHS(X) implying that B $\in$ CYC(X). This fact contradicts the assumption that B $\in$ W$-$CYC(X).

In the third subcase B $\in$ EXT(F), in which case if B $\notin$ Y then YA $\rightarrow$ B $\notin F^+$, contradicting the fact that Y is an antikey. The result now follows, since W $\subseteq$ Y implies that Y $\rightarrow$ A $\in F^+$ contradicting the fact that Y is an antikey for R. Thus A is nonprime as required.

Only if. Suppose that A $\notin$ NPR(F). We need to show that A is prime. The proof is obtained in three steps. Firstly by showing that $\emptyset \rightarrow$ A $\notin F^+$, secondly by showing that Y $\rightarrow$ A $\notin F^+$ and thirdly by showing that YA is a superkey for R.

If A $\in$ EXT(F) then A is prime, so assume that A $\notin$ EXT(F) and thus for some X $\subseteq$ sch(R), A $\in$ CYC(X)$-$NPR(F); in both cases we have $\emptyset \rightarrow$ A $\notin F^+$. It follows that CYC(X) $\not\subseteq$ NPR(F) and thus by the definition of NPR(F) we have that $\forall$ W $\in$ LHS(X), $\exists$ B $\in$ (W $\cap$ CYC(X))$-$NPR(F). Therefore it must be the case that $\mid$CYC(X)$-$NPR(F)$\mid > 1$, since otherwise $\forall$ W $\in$ LHS(X), A $\in$ W contradicting the fact that A $\in$ CYC(X).

By Lemma 38 it remains to exhibit an antikey Y $\in$ ANTI(F) such that A $\notin$ Y. In fact, for A to be prime it is sufficient to exhibit Y such that Y $\rightarrow$ A $\notin F^+$ and YA is a superkey for R [Kundu 1985]. Consider the following algorithm, designated by ANTIKEY(F, A), which returns a set of attributes Y $\subset$ sch(R).

ALGORITHM 1.

1. **begin**
2. Y := sch(R)$-$A;
3. **while** Y $\rightarrow$ A $\in F^+$ **do**
4. **find** X $\subseteq$ Y and B $\in$ X that satisfies B $\in$ (X $\cap$ CYC(X))$-$NPR(F);
5. Y := Y$-$B;
6. **end while**
7. **return** Y;
8. **end.**

We first show that Y $\rightarrow$ A $\notin F^+$. If A $\in$ EXT(F) then the result follows. So suppose that Z is a state of Y during the computation of ANTIKEY(F, A) such that Z $\rightarrow$ A $\in F^+$. In order to derive Z $\rightarrow$ A $\in F^+$, by Lemma 23 it must be the case that for some X $\subseteq$ Z, $\exists$ C $\in$ ((sch(R)$-$Z) $\cap$ CYC(X)), since by lines 2 and 4 NPR(F) $\subseteq$ Z holds. Therefore, CYC(X) $\not\subseteq$ NPR(F), otherwise we have CYC(X) $\subseteq$ Z. It follows by the definition of NPR(F) that $\forall$ W $\in$ LHS(X), $\exists$ B $\in$ (W $\cap$ CYC(X))$-$NPR(F); in particular, this last statement is true when W is X. Thus each time the while loop of Algorithm 1, beginning at line 3 and ending at line 6, is executed the cardinality of the current state Z of Y is reduced by one. The result follows, since $\emptyset \rightarrow$ A $\notin F^+$.

Algorithm 1 is nondeterministic and thus its output may vary from one execution to another. Let **S** = {Y | Y is returned from a computation of ANTIKEY(F, A)} and let MAX(**S**) be the family of all maximal sets of **S**. It remains to show that YA is a superkey for R, where Y $\in$ MAX(**S**). We conclude the result by induction on the number of times, say k, the while loop beginning at line 3 and ending at line 6 was executed when Y was returned from the computation of ANTIKEY(F, A).

(*Basis*): If k = 0 then A $\in$ EXT(F), and thus the result holds, since YA = sch(R).

(*Induction*): Assume the result holds when the number of times the while loop is executed is k, with k $\geq$ 0; we then need to prove that the result holds when the number of times the while loop is executed is k+1.

Let Z be the penultimate state of Y prior to the (k+1)*th* execution of the while loop. Furthermore, assume that X $\subseteq$ Z is the set of attributes during the last execution of line 4 satisfying B $\in$ (X $\cap$ CYC(X))$-$NPR(F), with Z = YB. It follows that for some C $\in$ (X $\cap$ CYC(X))$-$NPR(F), there are FDs, VB $\rightarrow$ C, VC $\rightarrow$ B $\in F^+$, which are reduced and cyclic FDs such that X = VB. Therefore V $\subseteq$ Y. In addition, since Y is maximal in **S**, the FD VB $\rightarrow$ C is used nonredundantly in a derivation of Z $\rightarrow$ A $\in F^+$. It follows that C $\notin$ Z and thus C $\notin$ Y, otherwise, if C $\in$ Z, then on using VC $\rightarrow$ B $\in F^+$ and YB $\rightarrow$ A $\in F^+$ we derive Y $\rightarrow$ A $\in F^+$, which leads to a contradiction of our earlier result.

We claim that Z is a superkey for R on utilising the induction on k. If k = 1 then ZA = sch(R) and thus Z is a superkey for R, since Z $\rightarrow$ A $\in F^+$. Otherwise, let W be the previous state of Z, and thus by induction hypothesis W is a superkey for R. There are two cases to consider. Firstly, if C $\in$ W then W = ZC, and Z is a superkey for R, since Z $\rightarrow$ C $\in F^+$ due to the fact that VB $\subseteq$ Z. Secondly, if C $\notin$ W then W = ZD for some D $\in$ W, with D $\neq$ C. Moreover, there is no redundant execution of the while loop in Algorithm 1, since we have assumed that Y is a maximal set in **S**. Therefore, the order in which the attributes are removed from the current state of Y, when line 5 is executed, is immaterial and thus for some computation of ANTIKEY(F, A) we have that (W$-$D)C is the previous state of Z. The result now follows as in the first case.

We finally claim that we can assume without loss of generality that C = A. Now, there is a reduced FD U $\rightarrow$ A $\in F^+$, for some U $\subseteq$ sch(R), with A $\in$ CYC(U)$-$NPR(F). Thus U $\not\subseteq$ Y and for some D $\in$ CYC(U), D $\notin$ Y, otherwise Y $\rightarrow$ A $\in F^+$ leading to a contradiction of our earlier result. Furthermore, there is no redundant execution of the while loop in Algorithm 1, since we have assumed that Y $\in$ **S** is maximal. Therefore, the order in which the attributes are removed from the current state of Y, when line 5 is executed, is immaterial and thus for some computation of ANTIKEY(F, A) we have that U $\rightarrow$ A is the FD VB $\rightarrow$ C.

The result now follows, since VA $\rightarrow$ B $\in F^+$ and V $\subseteq$ Y imply that YA $\rightarrow$ B $\in F^+$. Moreover, this implies that YA $\rightarrow$ Z $\in F^+$ and thus YA is a superkey for R as required. □

8. NORMAL FORMS AND MONODEPENDENCE

Herein we investigate the effect that monodependence has on relation schemas that are in a particular normal form. We first formalise the concepts of 2NF, 3NF and BCNF [Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993].

Definition 23. Let F be a set of FDs over R. R is in *Second Normal Form* (2NF) with respect to F (or simply in 2NF whenever F is understood from context) if for every nontrivial FD X $\rightarrow$ A $\in$ F, either X is *not* a proper subset of a key for R or A is prime.

R is in *Third Normal Form* (3NF) with respect to F (or simply in 3NF whenever F is understood from context) if for every nontrivial FD X $\rightarrow$ A $\in$ F, either X is a superkey for R or A is prime.

R is in *Boyce-Codd Normal Form* (BCNF) with respect to F (or simply in BCNF whenever F is understood from context) if for every nontrivial FD X $\rightarrow$ A $\in$ F, X is a superkey for R.

Lemma 42. Let R be in 2NF with respect to F, which satisfies the intersection property, and assume that $|$KEYS(F)$|> 1$. Then all the attributes in sch(R) are prime.

Proof. We prove the result by contradiction. Let X, Y $\in$ KEYS(F) be two distinct keys for R and let A $\in$ sch(R) be a nonprime attribute. It follows that X $\cap$ Y is a proper subset of both X and Y. Furthermore, X $\rightarrow$ A, Y $\rightarrow$ A $\in F^+$. Thus X $\cap$ Y $\rightarrow$ A $\in F^+$, since F satisfies the intersection property. This contradicts the fact that R is in 2NF, since X $\cap$ Y is a proper subset of a key for R. □

In [Jou and Fischer 1982] it was shown that, in general, testing whether R is in 3NF is NP-complete. The next corollary shows that when F satisfies the intersection property testing for 3NF can be done in polynomial time in the size of F.

Corollary 43. Testing whether R is in 3NF with respect to a set of FDs F, which satisfies the intersection property, can be done in polynomial time in the size of F.

Proof. Due to the fact that testing whether a set of attributes X is a superkey for R can be done in polynomial time in the size of F [Beeri and Bernstein 1979], it only remains to show that testing whether an attribute A is prime can be done in polynomial time in the size of F.

There are two cases to consider. Firstly, if $|$KEYS(F)$|= 1$, then A is prime if and only if A $\in$ X, where KEYS(F) $= \{$X$\}$. The result follows, since testing whether $|$KEYS(F)$|= 1$ can be done in polynomial time in the size of F [Biskup et al. 1991]. Secondly, if $|$KEYS(F)$|> 1$, then the result follows, since by Lemma 42 all of the attributes A $\in$ sch(R) are prime and if R is in 3NF then it is also in 2NF. □

The following theorem shows that when F satisfies the intersection property then 2NF and 3NF are equivalent.

Theorem 44. If F satisfies the intersection property, then R is in 2NF if and only if R is in 3NF.

Proof. If R is in 3NF then it is in 2NF [Codd 1972; Ullman 1988; Atzeni and De Antonellis 1993].

We prove the only if part by contradiction. If R is in 2NF and |KEYS(F)|> 1 then the result that R is in 3NF follows from the definitions of 2NF and 3NF on using Lemma 42. So assume that |KEYS(F)| = 1, with KEYS(F) = {X}, and that R is in 2NF but that R is not in 3NF. It follows that ∃ YA ⊂ sch(R) such that Y → A ∈ F^+ and A is nonprime, and ∃ B ∈ Y such that B is nonprime, where Y is not a superkey for R. Thus X and Y are incomparable and therefore it must be the case that (X ∩ Y) → A ∈ F^+, since F is satisfies the intersection property. However, this contradicts the fact that R is in 2NF, since (X ∩ Y) ⊂ X. □

The next example shows that, in general, when F is monodependent then 3NF does not imply BCNF.

Example 8. Let F = {A → B, B → A}, with sch(R) = {A, B, C}. It follows that KEYS(F) = {AC, BC} and thus R is not in BCNF. On the other hand, F is monodependent and, in addition, R is in 3NF, since all of A, B and C are prime.

We next present two special cases when 2NF implies 3NF and 3NF implies BCNF. The first special case follows from Theorem 44 and Theorem 3 in [Biskup et al. 1991].

Corollary 45. Let R be a relation schema and F be a set of FDs that satisfies the intersection property and suppose that |KEYS(F)|= 1. Then R is in 2NF if and only if R is in 3NF if and only if R is in BCNF. □

The second special case follows from Theorem 44 and Lemma 4.2 in [Date and Fagin 1992].

Corollary 46. Let R be a relation schema and F be a set of FDs that satisfies the intersection property and suppose that all the keys in KEYS(F) are simple. Then R is in 2NF if and only if R is in 3NF if and only if R is in BCNF. □

The following lemma is an immediate consequence of Lemma 27 and Corollary 28. We observe that due to the fact that R is in BCNF, the optimum cover G of F has only one equivalence class E_G(X), where X is a key for R.

Lemma 47. Let R be a relation schema that is in BCNF with respect to a monodependent set of FDs F and suppose that |KEYS(F)|> 1. Then |KEYS(F)| ≤ type(R) and ∀ Y ∈ KEYS(F), |Y|= type(R)−1. □

The next theorem characterises schemas which are in BCNF with respect to a monodependent set of FDs, in terms of a set of keys having a simple structure and whose cardinality is at most type(R).

Theorem 48. Assume that R is in BCNF with respect to a set of FDs F. Then F is monodependent if and only if either |KEYS(F)|= 1 or ∀ Y ∈ KEYS(F), |Y|= type(R)−1.

Proof. The only if part follows directly by Lemma 47. For the if part let KEYS(F) = $\{X_1, \ldots, X_n\}$. There are two cases to consider.

Firstly, assume that |KEYS(F)|= 1. Then on letting Y $=X_1$ there are two subcases to consider. In the first subcase assume that W $\rightarrow$ A and Z $\rightarrow$ A are incomparable FDs in F^+. It follows that Y $\subseteq$ W, Z and therefore (W $\cap$ Z) $\rightarrow$ A $\in F^+$. In the second subcase assume that WB $\rightarrow$ A and ZA $\rightarrow$ B are cyclic FDs in F^+. It follows that Y $\subseteq$ WB, ZA and that A, B $\notin$ Y and thus both W $\rightarrow$ A and Z $\rightarrow$ B are in F^+. The result that F is monodependent follows.

Secondly, assume that |KEYS(F)|> 1 and thus $n \leq$ type(R), since $\forall$ Y $\in$ KEYS(F), |Y|= type(R)$-$1. There are two subcases to consider. In the first subcase W $\rightarrow$ A and Z $\rightarrow$ A $\in F^+$ cannot be incomparable FDs, since $\forall i \in I$, $|X_i|=$ type(R)$-$1. In the second subcase assume that WB $\rightarrow$ A and ZA $\rightarrow$ B are cyclic FDs in F^+, with A $\neq$ B. It follows that W = Z since WBA = ZAB = sch(R), and thus F is monodependent as required. □

We observe that by Example 7 when R is in 3NF with respect to a monodependent set of FDs, F, then |KEYS(F)| may have exponential cardinality.

9. LOSSLESS JOIN AND DEPENDENCY PRESERVING DECOMPOSITIONS ASSUMING MONODEPENDENCE

Database design can be viewed as the process of replacing a relation schema R together with a set of FDs F over R by a set of relation schemas $\mathbf{R} = \{R_1, \ldots, R_m\}$ such that $\bigcup_{i=1}^{m}$ sch(R_i) = sch(R); $\mathbf{R}$ is called a decomposition of R. With each $R_i \in \mathbf{R}$ we associate a set of FDs F_i such that F_i is the *projection* of F^+ onto R_i. Two fundamental desirable properties of decompositions are that they possess the *lossless join property* and that they be *dependency preserving* [Ullman 1988; Mannila and Räihä 1992; Atzeni and De Antonellis 1993]. Intuitively, the lossless join property [Aho et al. 1979] means that the original relation schema R can be faithfully reconstructed from the decomposition $\mathbf{R}$ with respect to F. On the other hand, dependency preservation [Beeri and Honeyman 1981] means that $\bigcup_{i=1}^{m} F_i$ is a cover of F, i.e. that the semantics of the set of FDs F are preserved in the decomposition $\mathbf{R}$. Thus a decomposition $\mathbf{R}$ of R that possesses the lossless join property is dependency preserving and, in addition, is in BCNF (i.e. such that each R_i is in BCNF with respect to F_i) is indeed desirable.

Assume that F is not necessarily monodependent. The problem that has confronted database researchers since the late 1970's is firstly, whether the desirable properties just mentioned can be obtained, and secondly, if they can, then at what computational cost. An early result [Beeri and Bernstein 1979, Theorem 3], shown by exhibiting a simple counterexample, is that it is not always possible to obtain a decomposition which is both dependency preserving and in BCNF. It was further shown in [Beeri and Bernstein 1979; Tsou and Fischer 1982] that the problem of deciding whether there exists a dependency preserving decomposition of R, which is in BCNF with respect to F, is NP-hard. An exponential time algorithm in the size of F, which decides whether there exists a decomposition of R which is dependency preserving and in BCNF, was given by Osborn [Osborn 1979]. On the other hand, Tsou and Fischer [Tsou and Fischer 1982] showed that a lossless join decomposition of R which is in BCNF can always be obtained in polynomial time in the size of F.

It was shown in [Biskup et al. 1979; Vardi 1984] that if a decomposition $\mathbf{R}$ of R is dependency preserving and there exists a relation schema, $R' \in \mathbf{R}$, such that sch(R')

is a superkey for R, then **R** also possesses the lossless join property with respect to F. An immediate consequence of this result is that if there exists a dependency preserving decomposition of R which is in BCNF, then there exists a decomposition of R which possesses the lossless join property, is dependency preserving, and is in BCNF. This is due to the fact that if **R** does not posses the lossless join property, then we can simply add a relation schema R' to **R** such that sch(R') is a key for R, noting that R' is in BCNF with respect to the projection of F^+ onto R'.

Later Nummenmaa and Thanisch [Nummenmaa and Thanisch 1990] gave a sufficient condition for the existence of a decomposition of R which possesses the lossless join property, is dependency preserving, and is in BCNF. Furthermore, this condition can be tested in polynomial time in the size of F. In the special case when the decomposition is *acyclic* [Beeri et al. 1983], Yuan and Ozsoyoglu [Yuan and Özsoyoğlu 1992] gave a necessary and sufficient condition for the existence of a decomposition of R which possesses the lossless join property, is dependency preserving, and is in BCNF. Furthermore, this condition can also be tested in polynomial time in the size of F.

Our main result herein is that when F satisfies the intersection property, then F satisfying the split-freeness property, i.e. being monodependent, is both a necessary and sufficient condition for every decomposition of R, which possesses the lossless join property, also to be dependency preserving. In addition, as we have already mentioned after Definition 11, monodependence of a set of FDs F over R can be checked in polynomial time in the size of F. As a corollary of this result we are able to show that when F satisfies the intersection property, then F satisfying the split-freeness property, i.e. being monodependent, is both a necessary and sufficient condition for every decomposition of R, which possesses the lossless join property and is in BCNF, also to be dependency preserving. Therefore, assuming that F is monodependent, we can utilise the BCNF decomposition algorithm presented in [Tsou and Fischer 1982] in order to obtain, in polynomial time in the size of F, a decomposition of R possessing all these desirable properties. Moreover, we show that using an optimum cover G of F, which by Corollaries 16 and 26 can be computed in polynomial time in the size of F, we can obtain, in polynomial time in the size of G, a decomposition of R into BCNF with the smallest possible size, i.e. a decomposition which is *optimum*, possesses the lossless join property and is also dependency preserving. This optimum decomposition is the "best" possible decomposition a database designer can obtain, since we also show that it is unique.

Definition 24. A set $\mathbf{R} = \{R_1, \ldots, R_m\}$ is said to be a *decomposition* of a relation schema R (or simply a decomposition whenever R is understood from context), if $\forall i \in \{1, \ldots, m\}$, $\mathrm{sch}(R_i) \subseteq \mathrm{sch}(\mathrm{R})$, and $\bigcup_{i=1}^{m} \mathrm{sch}(R_i) = \mathrm{sch}(\mathrm{R})$.

We define the *size* of **R**, denoted by $||\mathbf{R}||$, to be the number of attributes appearing in **R** including repetitions.

From now on we let $\mathbf{R} = \{R_1, \ldots, R_m\}$ be a *decomposition* of R.

Definition 25. The *projection* of a set of FDs F onto a relation schema S, with $\mathrm{sch}(\mathrm{S}) \subseteq \mathrm{sch}(\mathrm{R})$, denoted by F[S], is given by

$$\mathrm{F[S]} = \{\mathrm{X} \to \mathrm{Y} \mid \mathrm{X} \to \mathrm{Y} \in \mathrm{F} \text{ and } \mathrm{XY} \subseteq \mathrm{sch(S)}\}.$$

The FDs in F[S] are said to be *embedded* in S. The set of FDs in F that are embedded in a decomposition $\mathbf{R}$ of R, denoted by F[$\mathbf{R}$], is given by F[$\mathbf{R}$] $= \bigcup_{i=1}^{m}$ F[R_i] [Beeri and Honeyman 1981]. If F = F[$\mathbf{R}$] then we say that F is embedded in $\mathbf{R}$.

DEFINITION 26. A decomposition $\mathbf{R}$ is a *dependency preserving decomposition* of R with respect to F (or simply dependency preserving whenever F is understood from context) [Beeri and Honeyman 1981], if for some cover G of F we have that G[$\mathbf{R}$]$^+ = F^+$.

DEFINITION 27. A decomposition $\mathbf{R}$ is a *lossless join decomposition* of R with respect to a set of FDs F (or simply $\mathbf{R}$ is lossless whenever R and F are understood from context) [Aho et al. 1979], if for all complete relations, r over R, such that r $\models$ F, the equation

$$r = \pi_{R_1}(r) \bowtie \ldots \bowtie \pi_{R_m}(r),$$

holds, where $\bowtie$ is the *natural join* operator.

DEFINITION 28. A decomposition $\mathbf{R}$ of R is in BCNF with respect to a set of FDs F (called a *BCNF decomposition* whenever F is understood from context), if each $R_i \in \mathbf{R}$ is in BCNF with respect to $F^+[R_i]$.

If $\mathbf{R}$ is a BCNF decomposition which is also lossless (and dependency preserving), we say that $\mathbf{R}$ is a lossless (and dependency preserving) BCNF decomposition.

We next define the concept of a derivation sequence, which is central to the developments that follow.

DEFINITION 29. A *derivation sequence* for some relation schema $R_0 \in \mathbf{R}$ with respect to a set of FDs F (or simply a derivation sequence whenever R_0 and F are understood from context) is a sequence of FDs, $\langle X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n \rangle$, satisfying the following conditions (cf. [Atzeni and Chan 1989]):

(1) $\forall i \in I$, $X_i \rightarrow A_i \in F^+$;
(2) $X_i \subseteq \text{sch}(R_0) \cup \bigcup_{j=1}^{i-1}\{A_j\}$; and
(3) $\text{sch}(R_0) \cup \bigcup_{i \in I}\{A_i\} = \text{sch(R)}$.

A derivation sequence $\langle X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n \rangle$ is said to be *embedded* in $\mathbf{R}$ if $\forall i \in I$, $X_i \rightarrow A_i$ is embedded in some relation schema S $\in \mathbf{R}$, i.e. $X_iA_i \subseteq$ sch(S).

LEMMA 49. Let $\langle X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n \rangle$ be a derivation sequence for some relation schema $R_0 \in \mathbf{R}$ with respect to a set of FDs F. Then sch(R_0) is a superkey for R.

Proof. The result follows by a straightforward induction on the number of FDs, n, in the derivation sequence, since $\text{sch}(R_0) \cup \bigcup_{i \in I}\{A_i\} = \text{sch(R)}$. □

LEMMA 50. Let $\langle X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n \rangle$ be a derivation sequence for some relation schema $R_0 \in \mathbf{R}$ with respect to a monodependent set of FDs F. If the sequence is embedded in $\mathbf{R}$, then $\mathbf{R}$ is a dependency preserving decomposition of R.

Proof. By Definition 29 $\langle X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n \rangle$ is a derivation sequence for R_0 if and only if it is a derivation sequence for R_0 with respect to any cover G of F. Thus we can assume that the derivation sequence for R_0 is with respect to G, where G is the optimum cover of F. We conclude the result by proving that G is embedded in $\mathbf{R}$ on using an induction on the number of FDs, n, in the derivation sequence.

(*Basis*): If $n = 0$, then the result is immediate, since in this case sch(R_0) = sch(R).

(*Induction*): Assume the result holds when the number of FDs in the derivation sequence is n, with $n \geq 0$; we then need to prove that the result holds when the number of FDs in the derivation sequence is $n + 1$.

Consider the $(n+1)th$ FD, i.e. $X_{n+1} \rightarrow A_{n+1}$, in the derivation sequence and let S be the relation schema, with sch(S) = sch(R) $-\{A_{n+1}\}$. By induction hypothesis G[S] is embedded in $\mathbf{R}$. If $A_{n+1} \in$ sch$(R_0) \cup \bigcup_{i=1}^{n}\{A_i\}$ then the result follows by induction hypothesis, since by Definition 29 $X_{n+1} \rightarrow A_{n+1}$ is redundant in the derivation sequence. So assume that $A_{n+1} \notin$ sch$(R_0) \cup \bigcup_{i=1}^{n}\{A_i\}$.

By assumption $X_{n+1} \rightarrow A_{n+1}$ is embedded in some $R_i \in \mathbf{R}$. Furthermore, by Theorem 31, on using Lemma 29, it must be the case that for a subset of X_{n+1}, say W, it is true that W $\rightarrow A_{n+1} \in$ G. Therefore, it must also be the case that W $\rightarrow$ Z $\in E_G$(W), with $A_{n+1} \in$ Z. Now, if $|E_G(\text{W})|= 1$, then the result follows, since by inductive hypothesis $\forall B_j \in$ Z$-A_{n+1}$, the FD W $\rightarrow B_j$ is embedded in $\mathbf{R}$. Thus we assume that $|E_G(\text{W})|> 1$. The result follows by Corollary 28, on using Lemma 27, since if W $\rightarrow A_{n+1}$ is embedded in R_i then all the FDs in E_G(W) are embedded in R_i. □

We assume that the reader is familiar with the *chase procedure* [Aho et al. 1979; Maier et al. 1979] but for completeness we give its definition.

DEFINITION 30. Let F be a set of FDs and $\mathbf{R}$ be a decomposition of R.

A *tableau* for $\mathbf{R}$, denoted by T($\mathbf{R}$), is a table having one column for each attribute in sch(R) and one row for each $R_i \in \mathbf{R}$. Each row, $t_i \in$ T($\mathbf{R}$) corresponding to $R_i \in \mathbf{R}$, is constructed from two types of variable as follows: $\forall$ A $\in$ sch(R), t_i[A] = a, if A $\in$ sch(R_i), otherwise t_i[A] = b_k; a is called a *distinguished variable* and is the same entry in each row where it appears, whereas b_k is called a *nondistinguished variable* and is a unique entry in each row where it appears.

The following transformation is applied to T($\mathbf{R}$) with respect to the individual FDs, X $\rightarrow$ Y $\in$ F, until no further changes are possible: if there are two rows t_1, t_2 in the current state of T($\mathbf{R}$) such that t_1[X] = t_2[X] but t_1[Y] $\neq$ t_2[Y] then $\forall$ A $\in$ Y, do the following. If t_1[A] = a and t_2[A] = b_i, or vice versa, then change all occurrences of b_i in the current state of T($\mathbf{R}$) to a, otherwise if t_1[A] = b_j and t_2[A] = b_i, with $j < i$, or vice versa, then change all occurrences of b_i in the current state of T($\mathbf{R}$) to b_j or vice versa.

The process of changing variables in the tableau is called *equating variables*. In addition, we call each transformation applied to the current state of T($\mathbf{R}$) a *chase* *step*, and we call the two rows t_1 and t_2, referred to above, the rows *affected* by the chase step. The resulting tableau is denoted by CHASE(F, $\mathbf{R}$). We call a row t $\in$ CHASE(F, $\mathbf{R}$) *winning* if $\forall$ A $\in$ sch(R), t[A] is a distinguished variable.

The next fundamental result was shown in [Aho et al. 1979].

LEMMA 51. A decomposition $\mathbf{R}$ of R is lossless with respect to F if and only if CHASE(F, $\mathbf{R}$) contains a winning row. $\square$

The following result is a consequence of Lemma 51 (cf. [Aho et al. 1979, Corollary 1]).

LEMMA 52. A decomposition $\{R_1, R_2\}$ of R is a lossless join decomposition with respect to F if and only if either X $\rightarrow$ sch(R_1) $\in F^+$ or X $\rightarrow$ sch(R_2) $\in F^+$, where X = sch(R_1) $\cap$ sch(R_2). $\square$

The next relevant result was shown in [Maier et al. 1979].

LEMMA 53. If $\mathbf{R}$ is a decomposition of R and G is a cover of F, then CHASE(F, $\mathbf{R}$) = CHASE(G, $\mathbf{R}$). $\square$

The next lemma is used in the subsequent theorem.

LEMMA 54. Let F be a monodependent set of FDs and let $\mathbf{R}$ be a lossless join decomposition of R with respect to F. Then there exists a derivation sequence with respect to F that is embedded in $\mathbf{R}$ for some relation schema $R_0 \in \mathbf{R}$.

Proof. Without loss of generality we assume that F is optimum, since by Definition 29 derivation sequences are cover invariant and by Lemma 53 losslessness is also cover invariant. By the main theorem in [Biskup et al. 1979; Vardi 1984] for some $R_0 \in \mathbf{R}$, sch(R_0) is a superkey for R, since $\mathbf{R}$ is lossless. It therefore remains to exhibit a derivation sequence for R_0 that is embedded in $\mathbf{R}$. We conclude the result by induction on the minimum number of chase steps, say k, that need to be applied to T($\mathbf{R}$) in order for CHASE(F, $\mathbf{R}$) to contain a winning row.
(*Basis*): If k = 0, then the result is immediate.

(*Induction*): Assume the result holds when the minimum number of chase steps that need to be applied to T($\mathbf{R}$) in order for CHASE(F, $\mathbf{R}$) to contain a winning row is, k, with k $\geq$ 0; we then need to prove that the result holds when the minimum number of chase steps that need to be applied to T($\mathbf{R}$) in order for CHASE(F, $\mathbf{R}$) to contain a winning row is k+1.

Observation 0. By Corollary 28 a sufficient condition for F to be embedded in $\mathbf{R}$ is that for each nonempty equivalence class E_F(X) of F, $\exists$ X $\rightarrow$ Y $\in E_F$(X) such that $\forall$ A $\in$ Y, X $\rightarrow$ A is embedded in some $R_i \in \mathbf{R}$.

Consider the last, i.e. the (k+1)*th*, chase step which is applied to the penultimate state, say T', of T($\mathbf{R}$) with respect to an FD X $\rightarrow$ Y $\in$ F. If $|E_F$(X)$|>$ 1 then by Lemma 27 the right-hand sides of all the FDs in E_F(X) are singletons and thus we can assume that Y = {A}; otherwise, if $|E_F$(X)$|=$ 1 then assume that A $\in$ Y. Let $t_1, t_2 \in T'$ be the two rows affected by the last chase step such that t_1[X] = t_2[X] but $(t_1[\mathrm{A}] = b') \neq (t_2[\mathrm{A}] = a)$, for some nondistinguished variable b'. If $|\mathrm{Y}| > 1$ then without any loss of generality assume that we split the last chase step into $|\mathrm{Y}|$ separate chase steps; the $(|\mathrm{Y}|)th$ such step now becomes the (k+1)th chase step being applied to T($\mathbf{R}$) with respect to X $\rightarrow$ A, with A $\in$ Y. By Lemma 51 it follows that t_1 is a winning row after the (k+1)th chase step is applied to T' and thus $\forall$ B $\in$ XA, $t_2[\mathrm{B}] = a$.

In order to simplify the notational aspects of this proof, we shall assume for the remainder of the proof that $|E_F(\mathrm{X})| > 1$. The argument can be appropriately modified, as was illustrated above, so as to deal with the case when $|E_F(\mathrm{X})| = 1$. In this case we replace any singleton equivalence class $E_F(\mathrm{X}) = \{\mathrm{X} \rightarrow \mathrm{Y}\}$ with the set of FDs $\{\mathrm{X} \rightarrow A_1, \ldots, \mathrm{X} \rightarrow A_p\}$, where $\mathrm{Y} = \{A_1, \ldots, A_p\}$.

We claim that in order to prove the result it is sufficient to show that X $\rightarrow$ A is embedded in some $R_i \in \mathbf{R}$. Let G = F $-E_F(\mathrm{X})$. Then, by Lemma 30, G is monodependent and optimum. Further, let S = ACY(X) $\cup$ CYC(X). We assume without loss of generality that G is a set of FDs over R', such that sch(R') = sch(R) $-$ S, since by Lemma 29 $\forall$ W $\rightarrow$ B $\in$ G, $\forall$ C $\in$ S, C $\notin$ WB. Let $\mathbf{R}'$ be the decomposition of R' resulting from $\mathbf{R}$ by removing all the attributes D $\in$ S from sch(R_i), $\forall R_i \in \mathbf{R}$. It follows that CHASE(G, $\mathbf{R}'$) contains a winning row and can be computed in k chase steps, since $\forall$ B $\in$ sch(R)$-$A, $t_1[\mathrm{B}] = a$. Therefore, by the inductive hypothesis it follows that G is embedded in $\mathbf{R}$.

By Theorem 31 we have that $\exists$ W $\in$ LHS(X) such that $\forall$ B $\in$ W, t[B] = a, where t $\in$ T($\mathbf{R}$) is the row corresponding to R_0. Furthermore, by Theorem 31 we can assume without loss of generality that t is the original state of $t_1 \in$ T($\mathbf{R}$). It remains to show that X $\rightarrow$ A is embedded in some $R_i \in \mathbf{R}$. Thus, in order to conclude the proof we show that $\exists t' \in$ T($\mathbf{R}$) such that $\forall$ B $\in$ XA, $t'[\mathrm{B}] = a$.

By Corollary 28, $\forall$ W $\rightarrow$ B $\in E_F(\mathrm{X})$, WB = XA. Consider the first chase step that applied an FD, W $\rightarrow$ B $\in E_F(\mathrm{X})$, to a state of T($\mathbf{R}$). (There must be at least one such chase step, since the (k+1)th chase step applies X $\rightarrow$ A.) It follows that one of the affected rows by this chase step, say t_3, is such that $\forall$ B $\in$ XA, $t_3[\mathrm{B}] = a$. It remains to show that $t_3[\mathrm{XA}] = t'[\mathrm{XA}]$, where $t' \in$ T($\mathbf{R}$) is the original state of t_3.

Suppose that this is not the case. Then some FD V $\rightarrow$ C $\notin E_F(\mathrm{X})$ was applied to a previous state of T($\mathbf{R}$) by a chase step affecting a previous state of t_3. However, by Lemma 29 C $\notin$ XA, implying that $t_3[\mathrm{XA}] = t'[\mathrm{XA}]$ as required. The result that $\exists R_i \in \mathbf{R}$ such that X $\rightarrow$ A is embedded in R_i follows, thus concluding the proof. □

An immediate consequence of Lemma 54 is that when F is monodependent then CHASE(F, $\mathbf{R}$) can be computed without equating nondistinguished variables. This is due to the fact that derivation sequences are a special case of *extension joins* [Honeyman 1980] (a straightforward induction on the number of FDs in the derivation sequence can be used to prove this fact). In addition, if F is optimum, then at most k chase steps are required to compute CHASE(F, $\mathbf{R}$), where k is the product of the number of nonempty equivalence classes of F multiplied by the maximum number of attributes on the right-hand side of any FD in F.

Theorem 55. A decomposition $\mathbf{R}$ of a relation schema R is a lossless join decomposition of R with respect to a monodependent set of FDs F if and only if there exists a derivation sequence that is embedded in $\mathbf{R}$ for some $R_0 \in \mathbf{R}$.

Proof. The only if part follows immediately from Lemma 54. The if part follows from the main theorem in [Biskup et al. 1979; Vardi 1984], since by Lemma 49 sch(R_0) is a superkey for R and by Lemma 50 $\mathbf{R}$ is a dependency preserving decomposition of R for $R_0 \in \mathbf{R}$. □

The next corollary follows immediately from the only if part of Theorem 55 and Lemma 50.

COROLLARY 56. Let F be a set of monodependent FDs. Then every lossless join decomposition $\mathbf{R}$ of R is dependency preserving. □

The next two counterexamples falsify the conjecture that the converse of Corollary 56 holds. In the theorem that follows we will prove the converse of Corollary 56 under the stronger assumption that F satisfies the intersection property.

EXAMPLE 9. Let F = {A → B, B → A, B → C}, with sch(R) = {A, B, C}. Then F is not monodependent since it violates the intersection property, but every lossless join decomposition of R is dependency preserving.

Also, let G = {A → B, B → A, AC → D, BD → C}, with sch(R) = {A, B, C, D}. Then G is not monodependent since it violates both the intersection and split-freeness properties, but every lossless join decomposition of R is dependency preserving.

THEOREM 57. Let F be a set of FDs that satisfies the intersection property. Then the following statements are equivalent:

(1) F satisfies the split-freeness property, i.e. F is monodependent.

(2) Every lossless join decomposition of R is dependency preserving.

Proof. (2 ⇒ 1.) We prove the result by contraposition. Assume that F violates the split-freeness property. We need to show that there exists a lossless join decomposition $\mathbf{R}$ of R such that $\mathbf{R}$ is not dependency preserving.

In view of our assumption there exist cyclic FDs, say XB → A, YA → B ∈ F^+, such that both Y → B $\notin F^+$ and (X ∩ Y)A → B $\notin F^+$. Let $\{R_1, R_2\}$ be a decomposition of R such that sch(R_1) = XBA and sch(R_2) = sch(R)−A. By Lemma 52, it follows that $\{R_1, R_2\}$ is a lossless join decomposition of R. Moreover, YAB $\not\subseteq$ sch(R_1), since this would imply that Y ⊆ X and thus (X ∩ Y)A → B ∈ F^+; also YAB $\not\subseteq$ sch(R_2), since A ∉ sch(R_2). Furthermore, YA → B $\notin F^+[\{R_1\}]$, since otherwise we would have that (X ∩ Y)A → B ∈ F^+; also YA → B $\notin F^+[\{R_2\}]$, since otherwise we would have that Y → B ∈ F^+. The only remaining possibility is that Y → X ∈ $F^+[\{R_2\}]$ and that XA → B ∈ $F^+[\{R_1\}]$ implying by the pseudo-transitivity inference rule that YA → B ∈ $F^+[\{R_1, R_2\}]$. However, by the intersection property we can deduce that (X ∩ Y)A → B ∈ F^+ leading to a contradiction. The result now follows, since we have shown that $\{R_1, R_2\}$ is *not* a dependency preserving decomposition of R.

(1 ⇒ 2.) It follows directly from Corollary 56. □

The next corollary specialises Theorem 57 to the case when the decomposition is in BCNF.

COROLLARY 58. Let F be a set of FDs that satisfies the intersection property. Then the following statements are equivalent:

(1) F satisfies the split-freeness property, i.e. F is monodependent.

(2) Every BCNF lossless join decomposition of R is also dependency preserving.

Proof. The proof that $2 \Rightarrow 1$ is essentially the same as that of Theorem 57 with the following proviso. If $\{R_1, R_2\}$ is a BCNF decomposition, then the result follows, otherwise decompose R_1 and R_2 further until a BCNF decomposition is attained using the standard BCNF decomposition algorithm [Mannila and Räihä 1992, Algorithm 12.2]. The assertion that $1 \Rightarrow 2$ follows directly from Theorem 57. □

The next example shows that it may be the case that F satisfies the intersection property and violates the split-freeness property property but there still exists a BCNF lossless join decomposition of R which is also dependency preserving.

EXAMPLE 10. Let F = {AB → CD, CD → AB}, with sch(R) = {A, B, C, D}. Then F satisfies the intersection property and violates the split-freeness property. It can easily be verified that **R** = {R} and **S** = {ABC, ABD} are both BCNF lossless join decompositions of R but that only **R** is also dependency preserving.

LEMMA 59. Let F be optimum and monodependent and let E_F(X) be an equivalence class of F. Then the relation schema S, with sch(S) = XY and X → Y ∈ E_F(X), is in BCNF with respect to F^+[S].

Proof. Let G = F^+[S]. Then by Definition 23 it remains to show that E_F(X) is a cover of G due to the following argument. By Corollary 28 ∀ W → Z ∈ E_F(X), XY = WZ holds. In addition, we have that sch(S) = XY. Therefore, W is a superkey for S with respect to E_F(X) implying that S is in BCNF with respect to F^+[S].

The containment $E_F(\mathrm{X})^+ \subseteq G^+$ follows by Corollary 28, so it remains to show that $G^+ \subseteq E_F(\mathrm{X})^+$. If $|E_F(\mathrm{X})| = 1$ then the result follows. So assume that $|E_F(\mathrm{X})| > 1$ and that there exists W → B ∈ $G^+ - E_F(\mathrm{X})^+$. By Lemma 27 |Y|= 1, so let Y = {A}. There are two cases to consider. Firstly, A = B and thus W ⊆ X. If W = X then W → B ∈ $E_F(\mathrm{X})^+$ leading to a contradiction. On the other hand, if W ⊂ X then X → A is not reduced leading to a contradiction of the fact that F is optimum.

Secondly, A ≠ B and thus B ∈ X. If W = XA−B, i.e. A ∈ W, then it must be the case that W → B ∈ $E_F(\mathrm{X})^+$, since by Lemma 29 B is not a member of the right-hand side of any other FD in F. On the other hand, if W ⊂ X, i.e. A ∉ W, then by the pseudo-transitivity inference rule we conclude that (X−B) → A ∈ F^+ leading to a contradiction of the fact that F is optimum.

It follows that W → B ∈ $E_F(\mathrm{X})^+$ and thus $G^+ \subseteq E_F(\mathrm{X})^+$ as required. Therefore S is in BCNF with respect to F^+[S]. □

A BCNF lossless join decomposition **R** of R with respect to F is *optimum* if there does not exist another BCNF lossless join decomposition of R with respect to F whose size is less than **R**.

Definition 31. A BCNF lossless join decomposition $\mathbf{R}$ of R is *optimum* if there does not exist a BCNF lossless join decomposition $\mathbf{S}$ or R such that $||\mathbf{S}|| < ||\mathbf{R}||$.

Theorem 60. There is a unique optimum BCNF lossless join and dependency preserving decomposition of R with respect to a monodependent set of FDs F, which can be obtained in polynomial time in the size of F.

Proof. Let G be the optimum cover of F, which by Corollary 16 or Corollary 26 can be computed in polynomial time in the size of F. Furthermore, let K $\subseteq$ sch(R) be a key for R, which by Algorithm 3.3 in [Atzeni and De Antonellis 1993] can be computed in polynomial time in the size of F. Moreover, let R' be a relation schema, with sch(R') = K.

Observation 1. By Theorem 36 all the keys in KEYS(F) have the same cardinality and thus there does not exist a key in KEYS(F) whose cardinality is less than |K|.

Let E = $\{E_G(X_1), \ldots, E_G(X_k)\}$ be the set of all nonempty equivalence classes of F, which can be computed in polynomial time in the size of F [Maier 1980]. In addition, let $\mathbf{R} = \bigcup_{i=1}^{k} R_i$, with sch$(R_i)$ = XY for some X $\rightarrow$ Y $\in E_G(X_i)$. Finally, if for no $R_i \in \mathbf{R}$ is it true that K $\subseteq$ sch(R_i), then add R' to $\mathbf{R}$.

Observation 2. By Corollary 28 and Lemma 29 for each $R_i \in \mathbf{R}$ there is exactly one equivalence class $E_G(X_i) \in$ E such that $\forall$ X $\rightarrow$ Y $\in E_G(X_i)$, sch(R_i) = XY.

It can easily be verified that $\mathbf{R}$ is dependency preserving and thus by the main theorem in [Biskup et al. 1979; Vardi 1984] $\mathbf{R}$ is also a lossless join decomposition of R, since K is a key for R. Furthermore, by Definition 28 and Lemma 59, $\mathbf{R}$ is in BCNF with respect to F. By Observation 1 at least |K| attributes are needed in order for some key for R to be a subset of sch(R_i) for some $R_i \in \mathbf{R}$. Moreover, by Observation 2 if $i \neq j$ then sch(R_i) and sch(R_j) are incomparable. Assume that $\mathbf{R}$ is not optimum and that $\mathbf{S}$ is a BCNF lossless join decomposition of R whose size is less than that of $\mathbf{R}$. The result now follows, since we arrive at a contradiction of the fact that G is optimum, due to the fact that in this case we can construct a cover H of F from the FDs embedded in $\mathbf{S}$ such that H has fewer attributes than G. The result that $\mathbf{R}$ is unique follows from Theorem 17. □

10. CONCLUDING REMARKS

Previously in [Levene and Loizou 1997; Levene and Loizou 1998] we have shown that monodependence provides a solution to the additivity problem and thus it is a fundamental property that a set F of FDs should possess, when considering incomplete information in relational database design. We have shown throughout the paper that monodependent sets of FDs possess many desirable properties that ease the process of database design. In particular, when F satisfies the intersection property, then by Corollary 16 the problem of finding an optimum cover of F can be solved in polynomial time in the size of F. Moreover, by Theorem 17 this optimum cover is unique. Furthermore, when F satisfies the split-freeness property, then by Theorem 21 F has optimum left-hand sides and by Corollary 26 an optimum cover of F can be obtained in polynomial time in the size of F.

We have shown in Theorem 35 that the superkey of cardinality k problem is NP-complete when F satisfies only the intersection property. On the other hand, by Corollary 37, this problem can be solved in polynomial time in the size of F when F satisfies the split-freeness property. By Theorem 40 the prime attribute problem can be solved in polynomial time in the size of F when F satisfies the intersection property and by Theorem 41 the prime attribute problem can be solved in polynomial time in the size of F when F satisfies the split-freeness property. With respect to normal forms when F satisfies the intersection property then by Theorem 44 2NF and 3NF are equivalent and by Theorem 48 BCNF has a simple characterisation, i.e. in this case F is monodependent if and only if either $|\mathrm{KEYS(F)}| = 1$ or $\forall$ $\mathrm{Y} \in \mathrm{KEYS(F)}$, $|\mathrm{Y}| = \mathrm{type(R)} - 1$. An investigation of the properties of the lattice of subsets of sch(R) induced by a monodependent set of FDs was carried out in [Levene 1995].

We also tackled a long standing problem in relational database theory by showing in Theorem 57 that when a set of FDs F over R satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every lossless join decomposition of R is also dependency preserving. In Corollary 58 we were able to conclude the tighter result that when F satisfies the intersection property, then it also satisfies the split-freeness property, i.e. is monodependent, if and only if every BCNF lossless join decomposition of R is also dependency preserving. It is still an open problem to determine necessary and sufficient conditions for every lossless join decomposition of R to be dependency preserving also when F may not satisfy the intersection property. Finally, in Theorem 60 we showed that there is a unique optimum BCNF lossless join decomposition of R with respect to a monodependent set of FDs F over R, that is also dependency preserving and, in addition, can be obtained in polynomial time in the size of F. This is ideal from a database designer's point of view, since such a decomposition cannot be bettered.

This paper provides strong evidence that when incomplete information is considered, then monodependent sets of FDs are of fundamental importance and possess many desirable properties. Due to our results the concept of monodependence is also relevant to database designers that are only considering complete relations in the database. For example, assume that the input specified to the database designer involves a set of FDs F over R which is *not* monodependent. It is an open problem whether a methodology to database design can be developed that transforms F into a monodependent set of FDs that preserves the semantics of the input set of FDs in some well defined manner. Such a methodology may involve adding attributes to sch(R) as is done in [Kandzia and Mangelmann 1980; Sciore 1981; Sciore 1983; Beeri and Kifer 1986].

ACKNOWLEDGMENTS

The authors would like to thank the referees for their constructive comments.

REFERENCES

Aho, A. V., Beeri, C., and Ullman, J. D. 1979. The theory of joins in relational databases. *ACM Transactions on Database Systems 4*, 297–314.

Armstrong, W. W. 1974. Dependency structures of data base relationships. In *Proceedings of IFIP Congress* (Stockholm, 1974), pp. 580–583.

Atzeni, P. and Chan, E. P. F. 1989. Efficient optimization of simple join expressions. *ACM Transactions on Database Systems 14*, 212–230.

Atzeni, P. and De Antonellis, V. 1993. *Relational Database Theory*. Benjamin/Cummings, Redwood City, Ca.

Atzeni, P. and Morfuni, N. M. 1986. Functional dependencies and constraints on null values in database relations. *Information and Control 70*, 1–31.

Beeri, C. and Bernstein, P. A. 1979. Computational problems related to the design of normal form relational schemas. *ACM Transactions on Database Systems 4*, 30–59.

Beeri, C., Fagin, R., Maier, D., and Yannakakis, M. 1983. On the desirability of acyclic database schemes. *Journal of the ACM 30*, 479–513.

Beeri, C. and Honeyman, P. 1981. Preserving functional dependencies. *SIAM Journal on Computing 10*, 647–656.

Beeri, C. and Kifer, M. 1986. An integrated approach to logical design of relational database schemes. *ACM Transactions on Database Systems 11*, 134–158.

Biskup, J., Dayal, U., and Bernstein, P. A. 1979. Synthesizing independent database schemes. In *Proceedings of ACM SIGMOD Conference on Management of Data* (Boston, Ma., 1979), pp. 143–151.

Biskup, J., Demetrovics, J., Libkin, L. O., and Muchnik, I. B. 1991. On relational database schemes having unique minimal key. *Journal of Information Processing and Cybernetics 27*, 217–225.

Codd, E. F. 1972. Further normalization of the data base relational model. In R. Rustin Ed., *Data Base Systems*, pp. 33–64. Englewood Cliffs, NJ: Prentice Hall.

Codd, E. F. 1979. Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems 4*, 397–434.

Date, C. J. 1995. *An Introduction to Database Systems* (Sixth ed.). Addison-Wesley, Reading, Ma.

Date, C. J. and Fagin, R. 1992. Simple conditions for guaranteeing higher normal forms in relational databases. *ACM Transactions on Database Systems 17*, 465–476.

Demetrovics, J. and Thi, V. D. 1987. Keys, antikeys and prime attributes. *Annales Univ. Sci. Budapest, Sect. Comp 8*, 35–52.

Demetrovics, J. and Thi, V. D. 1988. Relations and minimal keys. *Acta Cybernetica 8*, 279–285.

Garey, M. R. and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York.

Honeyman, P. 1980. Extension joins. In *Proceedings of International Conference on Very Large Data Bases* (Montreal, 1980), pp. 239–244.

Imielinski, T. and Lipski Jr., W. 1984. Incomplete information in relational databases. *Journal of the ACM 31*, 761–791.

Jou, J. H. and Fischer, P. C. 1982. The complexity of recognizing 3NF relation schemes. *Information Processing Letters 14*, 187–190.

Kandzia, P. and Mangelmann, M. 1980. On covering Boyce-Codd normal forms. *Information Processing Letters 11*, 218–223.

Kundu, S. 1985. An improved algorithm for finding a key of a relation. In *Proceedings of ACM Symposium on Principles of Database Systems* (Portland, Oregon, 1985), pp. 189–192.

Levene, M. 1995. A lattice view of functional dependencies in incomplete relations. *Acta Cybernetica 12*, 181–207.

Levene, M. and Loizou, G. 1997. The additivity problem for functional dependencies in incomplete relations. *Acta Informatica 34*, 135–149.

Levene, M. and Loizou, G. 1998. The additivity problem for data dependencies in incomplete relational databases. In B. Thalheim and L. Libkin Eds., *Semantics in Databases*, pp. 136–169. Berlin: Springer-Verlag.

Levene, M. and Loizou, G. 1999. *A Guided Tour of Relational Databases and Beyond*. Springer-Verlag, London. To appear.

LIEN, Y. E. 1982. On the equivalence of database models. *Journal of the ACM 29*, 333–362.

LUCCHESI, C. L. AND OSBORN, S. L. 1978. Candidate keys for relations. *Journal of Computer and System Sciences 17*, 270–279.

MAIER, D. 1980. Minimal covers in the relational database model. *Journal of the ACM 27*, 664–674.

MAIER, D. 1983. *The Theory of Relational Databases*. Computer Science Press, Rockville, Md.

MAIER, D., MENDELZON, A. O., AND SAGIV, Y. 1979. Testing implications of data dependencies. *ACM Transactions on Database Systems 4*, 455–469.

MANNILA, H. AND RÄIHÄ, K.-J. 1983. On the relationship of minimum and optimum covers for a set of functional dependencies. *Acta Informatica 20*, 143–158.

MANNILA, H. AND RÄIHÄ, K.-J. 1992. *The Design of Relational Databases*. Addison-Wesley, Reading, Ma.

NUMMENMAA, J. AND THANISCH, P. 1990. Conjectures and refutations in database design and dependency theory. In *Proceedings of International Conference on Database Theory* (Paris, 1990), pp. 215–228.

OSBORN, S. L. 1979. Testing for the existence of a covering Boyce-Codd normal form. *Information Processing Letters 8*, 11–14.

REITER, R. 1978. On closed world databases. In H. GALLAIRE AND J. MINKER Eds., *Logic and Data Bases*, pp. 55–76. New York: Plenum.

RISSANEN, J. 1977. Independent components of relations. *ACM Transactions on Database Systems 2*, 317–325.

SCIORE, E. 1981. Real world MVD's. In *Proceedings of ACM SIGMOD Conference on Management of Data* (Ann Arbor, 1981), pp. 121–132.

SCIORE, E. 1983. Improving database schemes by adding attributes. In *Proceedings of ACM Symposium on Principles of Database Systems* (Atlanta, 1983), pp. 379–383.

THI, V. D. 1986. Minimal keys and antikeys. *Acta Cybernetica 7*, 361–371.

TSOU, D.-M. AND FISCHER, P. C. 1982. Decomposition of a relation schema into Boyce-Codd normal form. *ACM SIGACT News 14*, 23–29.

ULLMAN, J. D. 1988. *Principles of Database and Knowledge-Base Systems*, Volume 1. Computer Science Press, Rockville, Md.

VARDI, M. Y. 1982. On decomposition of relational databases. In *Proceedings of IEEE Symposium on Foundations of Computer Science* (Chicago, Il., 1982), pp. 176–185.

VARDI, M. Y. 1984. A note on lossless database decompositions. *Information Processing Letters 18*, 257–260.

VINCENT, M. W. 1994. *The Semantic Justification for Normal Forms in Relational Database Design*. Ph. D. thesis, Department of Computer Science, Monash University.

YUAN, L.-Y. AND ÖZSOYOĞLU, M. 1992. Design of desirable relational database schemes. *Journal of Computer and System Sciences 45*, 435–470.