# A COMPLETE AXIOMATIZATION FOR FUNCTIONAL AND MULTIVALUED DEPENDENCIES IN DATABASE RELATIONS

Catriel Beeri, Ronald Fagin, and John H. Howard

Proc. 1977 ACM SIGMOD, D.C.P. Smith (ed.), Toronto, 47-61.

# A COMPLETE AXIOMATIZATION FOR FUNCTIONAL AND MULTIVALUED DEPENDENCIES IN DATABASE RELATIONS

Catriel Beeri

Department of Electrical Engineering and Computer Science
Princeton University, Princeton, N.J. 08540

Ronald Fagin and John H. Howard

IBM Research Laboratory K53/282
San Jose, California 95193

ABSTRACT

We investigate the inference rules that can be applied to functional and multivalued dependencies that exist in a database relation. Three types of rules are discussed. First, we list the well known rules for functional dependencies. Then we investigate the rules for multivalued dependencies. It is shown that for each rule for functional dependencies the same rule or a similar rule holds for multivalued dependencies. There is, however, one additional rule for multivalued dependencies that has no parallel among the rules for functional dependencies. Finally, we present rules that involve functional and multivalued dependencies together. The main result of the paper is that the rules presented are complete for the family of functional and multivalued dependencies.

## 1. INTRODUCTION

In the relational model of data bases, the data is organized into relations. A relation can be viewed as a table where each row of the table describes an entity and each column corresponds to an attribute of the entity being described. The relations in the data base are time varying; entities are inserted, deleted and modified. However, the structure of the relations is invariant. This structure is described by means of a relational schema. Such a schema consists of a description of the structure of each one of the relations, e.g., its name, the attributes that appear in it, integrity constraints, etc.

It has been observed that the relationships between the attributes in a relation have an important role in determining its properties. Functional dependencies are an important example of such a relationship. Codd, in his first papers on the relational model [Codd1, Codd2], observed that relations in which certain patterns of functional dependencies occur exhibit undesirable behavior. This observation led him to the definition of second and third normal forms for relational schemas. These normal forms eliminate most of the problems related to the existence of functional dependencies in the relations.

Because of their importance for the understanding of the properties of relational schemas, functional dependencies have been extensively studied. Armstrong [Arm] did an exhaustive study of their properties and presented a set of inference rules (he called them axioms) for the family of functional dependencies. He proved that if F is a set of functional dependencies that is closed under his rules, then there is a relation R such that the set of functional dependencies which hold for R is exactly F. As shown by Fagin [Fag1], this result implies the completeness of Armstrong's rules for functional dependencies. Using Armstrong's rules, Bernstein and Beeri ([Bern1, Bern2]) have recently presented an efficient algorithm for synthesizing a relational schema from a given set of functional dependencies. Their work clarifies the connection between the user's view as expressed by a set of functional dependencies and the relational schema that embodies it.

It has been observed by workers in the area of data base semantics (see e.g. [SS]) that the concept of a functional dependency is not general enough to capture the semantics of the user's view of the data base. There are dependencies that are not functional.

Suppose, for example, that the children of employees are listed in the relation which contains the information about the employees. Clearly, the set of children names of an employee depends on the identifier of the employee only; however, since there may be more than one child name in the set this is not a functional dependency. It has also been observed that the existence of such 'generalized' dependencies in a relation may lead to redundancy in the representation of the data and, as a result, to undesirable properties similiar to those that were observed by Codd in relations that are not in third normal form. However, since third normal form is based on the concept of functional dependency, these properties exist even in third normal form schemas.

Recently, Fagin [Fag2] and, independently, Zaniolo [Zan] have presented a formal definition of a nonfunctional dependency, which Fagin calls a "multivalued dependency." It turns out that these dependencies are indeed a generalization of functional dependencies. Also, they have the important property that the existence of such dependencies in a relation is equivalent to the fact that the relation is the natural join of some of its projections. It seems therefore that the study of these dependencies may lead to a better understanding of the structure of relational schemas.

In [Fag2], Fagin demonstrated a number of properties of these multivalued dependencies. In particular, he showed that the existence of such dependencies in a relation may indeed result in undesirable redundancy. This observation led him to a definition of a new ('fourth') normal form for relational schemas. To facilitate further study of these dependencies and their influence on the structure of relations, we investigate in this paper the inference rules that can be applied to these dependencies. We present a set of such rules and prove that this set is complete. There are three types of inference rules to be discussed. The first type is inference rules for functional dependencies. The second type is inference rules for multivalued depend encies, that is, rules that allow us to infer from the existence of some multivalued dependencies in a relation the existence of additional multivalued dependencies in the relation. The third type is "mixed" inference rules by which, given the existence of functional and multivalued dependencies in a relation, one can infer the existence in the relation of additional dependencies, the existence of which cannot be inferred using rules of the previous types.

The rules for functional dependencies are well known and we present them without proofs. Certain inference rules of the second type were noted by Fagin [Fag2] and Zaniolo [Zan]. The existence of rules of the third type was first observed by Fagin, who noted a special case. Fagin's work was mainly meant to be an introduction to the concept of a multivalued dependency and some of the rules were presented in restricted forms. Our purpose in this paper is to state and prove the rules in what we believe is their most general form and to show that the rules that we present are complete, that is, no additional rules are needed.

There are quite practical reasons for obtaining a complete axiomatization (that is, a complete set of inference rules). It is well known that consideration of dependencies that exist among attributes in the data base is a valuable tool for the database designer. Assume that a database designer has noted certain functional and multivalued dependencies that hold for a relation schema that is being analyzed. He can use these dependencies as input to a 'box' which (using the complete set of rules) can determine for any other dependency whether it is a logical consequence of the input dependencies. (See [Beer1] or [Bern1] for a description of such a 'box' for functional dependencies and [Beer2] for a description of a 'box' for multivalued dependencies.) If the set of rules is not known to be complete then there may exist dependencies that are logical consequences of the input dependencies but cannot be so labeled by the 'box'. Thus, the completeness of the rules ensures the schema designer that he has complete knowledge about the input dependencies and all their logical consequences. For further discussion of this subject see [Fag3].

It is interesting to note that the rules for multivalued dependencies presented here are very similar to the rules for functional dependencies. Specifically, we show that for each rule for functional dependencies, either the same rule or a very similar rule applies to multivalued dependencies and, to obtain a complete set of rules for multivalued dependencies, only one additional rule is needed. This similarity has been explored by Beeri [Beer2] to extend the results of [Bern1, Bern2] to multivalued dependencies.

In this paper we deal only with inference rules for dependencies in a fixed given relation. The problems that arise when the relation is not fixed are of a different nature. For example, one may want to know, given a set of dependencies in a relation, which dependencies will be valid in a projection of the relation or in a join of the relation with some other relation. These problems are outside the scope of this paper. (The projection problem was discussed bv Fagin [Fag2].)

The paper is organized as follows. In Section 2 we introduce relations and define the concept of a constraint on a set of relations. We view both functional and multivalued dependencies as special types of constraints. We also explain what is an inference rule and what is a complete set of inference rules. In Section 3 we define functional dependencies and list their inference rules. Multivalued dependencies are discussed in Section 4. They are defined in Section 4.1, their inference rules (of the second type) are discussed and proved in Section 4.2 and the inference rules of the third type are proved in Section 4.3. In Section 5 we prove that the rules listed in Sections 3 and 4 are complete for the family of functional and multivalued dependencies.

## 2. RELATIONS AND CONSTRAINTS

The word relation is used in the literature to denote a set of tuples and also to denote a structural description of sets of tuples. It is important to distinguish between these two denotations. In this paper we deal with sets of tuples and the word relation is used only to denote a set of tuples.

The functional and multivalued dependencies treated in this paper are special types of constraints on relations. In this section we explain what a constraint is and discuss inference rules and completeness of sets of inference rules for families of constraints.

### 2.1 Relations and Relational Operations

Attributes are symbols taken from a finite set $\{A_1, A_2 \ldots\}$. Each attribute A has associated with it a domain, denoted by DOM(A), which is the set of possible values for that attribute. For a set of attributes X, an X-value is an assignment of values to the atrributes of X from their domains. We will use the letters A, B, ... for single attributes and the letters X, Y, ... for sets of attributes. Following common practice in papers on relational data bases, we will not distinguish between the attribute A and the set $\{A\}$. Also, if X and Y are sets of attributes (not necessarily disjoint), then we write XY for the union of X and Y whenever that union appears in a dependency or in an expression denoting a projection of a relation.

A relation on the set of attributes $\{A_1, \ldots, A_n\}$ is a subset of the cross product $DOM(A_1) \times \ldots \times DOM(A_n)$. The elements of the relation are called tuples or rows. Since a relation is a set, the order of the rows is unimportant. Also, since the attributes are distinct, the order of the columns is unimportant. A relation R on $\{A_1, \ldots, A_n\}$ will be denoted by $R(A_1, \ldots, A_n)$. Similarly, if R is a relation over the union of the sets X,Y,..., then we use the notation R(X, Y, ...). The letters u, v, ... will be used to denote single tuples.

There are two operations on relations that are of interest to us - projection and the natural join.

If u is a tuple in a relation R(X) and A is an attribute in X, then u[A] is the A-component of u. Similarly, if Y is a subset of X, then u[Y] is the tuple (of size $|Y|$) containing the components of u corresponding to the elements of Y. The projection of R on Y, denoted by R[Y], is defined by

$$R[Y] = \{u[Y] \mid u \in R\}$$

That is, R[Y] is a relation on Y containing all tuples generated from the tuples of R by omitting the components corresponding to attributes not in Y. Note that R[Y] is a set, so each tuple occurs only once even if it is generated by several tuples of R. We will use the term projection also for single tuples; e.g., u[Y] is called the projection of u on Y.

Let R(X,Y) and S(Y,Z) be relations where X, Y and Z are disjoint sets of attributes (i.e., Y is the set of attributes that are common to R and S). The natural join of R and S is the relation T(X,Y,Z) defined by

$$T(X,Y,Z) = \{(x,y,z) \mid (x,y) \in R \text{ and } (y,z) \in S\}$$

That is, the natural join is created by gluing together tuples of R with tuples of S that have the same values for all attributes that are common to the two relations.

## 2.2 Constraints and Inference rules

It is convenient to discuss functional and multivalued dependencies and their inference rules in the context of constraints on relations. As we will see, these dependencies are indeed special types of constraints.

A constraint involving the set of attributes $\{A_1, \ldots, A_n\}$ is a predicate on the collection of all relations on this set. A relation $R(A_1, \ldots, A_n)$ obeys the constraint if the value of the predicate for R is 'TRUE'. If R obeys the constraint, then we also say that the constraint is valid in R. A constraint is defined by giving a notation for expressing it and the condition under which a relation obeys it.

As an example, suppose that employees in a company are to be described by a relation EMP(EMP-NO, EMP-SAL, EMP-AGE,...). A priori, any relation of this form can exist in the data base. However, if one specifies a constraint that EMP-AGE should be between 18 and 65 then only relations in which this constraint is valid can exist in the data base. Similarly, the specification that EMP-NO is a key is actually a constraint on the relations. A relation obeys this constraint only if no two different tuples in it have the same value of the attribute EMP-NO. Note that to specify that EMP-NO is a key is the same as to require that all the other attributes of the relation are functionally dependent on it. Thus, functional dependencies are constraints. The constraints which we consider in this paper are functional dependencies and multivalued dependencies.

It is often possible, given a set of constraints, to prove that some other constraints are implied by the given set. Given a set of constraints $\Gamma = \{C_1, \ldots, C_n\}$, we say that the constraint C is implied by $\Gamma$ if C is valid in every relation which obeys all the constraints in $\Gamma$. In other words, C is implied by $\Gamma$ if there exists no counterexample relation that obeys all the constraints of $\Gamma$ but does not obey C. An inference rule for a family of constraints is a rule by which, given some constraints from the family, one can infer that some other constraint is implied by them.

For an example of an inference rule let us look at the well known transitivity rule for functional dependencies. (Functional dependencies are defined formally in the next section.) the rule states that if the functional dependencies $A \rightarrow B$ and $B \rightarrow C$ are valid in a relation then the functional dependency $A \rightarrow C$ is also valid in it. Since the rule enables us to infer the validity of $A \rightarrow C$ in a relation from the validity of $A \rightarrow B$ and $B \rightarrow C$ in the relation, it enables us to infer that $A \rightarrow C$ is implied by $A \rightarrow B$ and $B \rightarrow C$.

A set of inference rules is complete for a family of constraints if for each set $\Gamma$ of constraints from the family, the constraints that are implied by $\Gamma$ are exactly those that can be derived from it using these inference rules. That is, the rules present us with an effective way of finding all the constraints that are implied by any given set of constraints in the family. We note the following important consequence of the definition of completeness: A set of rules is complete for a family of constraints if and only if, for each set $\Gamma$ of constraints in the family and for each constraint C that can not be inferred from $\Gamma$ by using the rules in the set, there exists a relation in which $\Gamma$ is valid but C is not valid. This observation is the basis of our completeness proof in Section 5. (Actually, in addition to proving completeness in the sense defined here, we also prove there completeness of our rules for a slightly stronger concept of completeness.)

The concept of completeness of a set of rules is of prime importance in any system where inference rules are used. As we noted in the Introduction, only if a complete set of rules is used, can the database designer be assured that he has complete knowledge of all dependencies that hold in a given database. Indeed, the fact that Armstrong's rules for functional dependencies are complete is an inherent and basic assumption in the work on functional dependencies reported in [Bern1, Bern2]. Similarly, a study of multivalued dependencies and their influence on the structure of relations must be based on a complete set of rules for them. This need for a complete set of rules was the motivation for the work reported in this paper.

## 3. FUNCTIONAL DEPENDENCIES AND THEIR INFERENCE RULES

Functional dependencies form a family of constraints that has been treated extensively in the literature (see, e.g., [Arm, Bern1, Bern2] and the references given there). Therefore we omit the proofs in this Section.

A functional dependency (abbr. FD), f, is a statement f: $X \to Y$ where X and Y are sets of attributes. If R(X, Y, ...) is a relation on a set of attributes that contains X and Y, then R obeys the FD f if every two tuples of R which have the same projection on X also have the same projection on Y. Given f:$X \to Y$, we say that f is a functional dependency from X to Y, that Y is functionally dependent on X or that X functionally determines Y. From the definition it follows that for each pair of sets X and Y there is at most one functional dependency from X to Y. Therefore, we usually omit the name of the FD and write $X \to Y$.

Note that even though an FD $X \to Y$ involves only the attributes of the sets X and Y, it can be interpreted as a constraint on relations whose attributes include the attributes in X and Y and perhaps more. Indeed, it follows from the definition that the validity of the FD for a particular relation depends only on the values in the columns named by the attributes from X and Y.

Using the definition it is easy to prove the validity of the following inference rules (or axioms) for FD's. Fagin [Fag1] showed that it follows from Armstrong's result [Arm] that these rules are complete for FD's. In the rules, X, Y, Z and W are arbitrary sets of attributes.

FD1 (Reflexivity): If $Y \subseteq X$ then $X \to Y$.

FD2 (Augmentation): If $Z \subseteq W$ and $X \to Y$ then $XW \to YZ$.

FD3 (Transitivity): If $X \to Y$ and $Y \to Z$ then $X \to Z$.

(Strictly speaking, FD1 is an axiom schema; the "reflexive" FD's exist in every relation no matter what other FD's exist in it. However, this distinction is not significant for the purpose of this paper and we ignore it.)

Even though this set of rules is complete, it is convenient to introduce additional rules (which, of course, are consequences of FD1 - FD3).

FD4 (Pseudo-transitivity): If $X \to Y$ and $YW \to Z$ then $XW \to Z$.

FD5 (Union): IF $X \to Y$ and $X \to Z$ then $X \to YZ$.

FD6 (Decomposition): If $X \to YZ$ then $X \to Y$ and $X \to Z$.

The rules FD5, FD6 state that for a given set of attributes we can combine and decompose the sets that depend on it arbitrarily. In particular, it follows that the FD $X \to A_1 \ldots A_n$ is equivalent to the set of FD's $X \to A_1, \ldots, X \to A_n$. This property is very useful for the manipulation of FD's and for proving their properties. (See, for example, the extensive use of these properties in [Bern1, Bern2].)

Note: The set of rules FD1, FD3, FD5, is also complete for FD's.

For a set F of FD's the closure of F, denoted by $F^+$, is the set of all FD's derivable from F using the rules FD1 - FD3. Armstrong's result can be stated as follows: For every set F of FD's, there exists a relation such that the set of FD's that are valid in it is exactly $F^+$.

## 4. MULTIVALUED DEPENDENCIES AND THEIR INFERENCE RULES

### 4.1 Multivalued Dependencies

The concept of functional dependency is not general enough to capture the various types of dependencies that exist in relations. It is possible that in a relation the values of the attributes in a set Y depend only on the values of the attributes in a set X, but there is more than one Y-value for a given X-value. Such a dependency is not functional. The concept of multivalued dependency was introduced by Fagin [Fag2] and Zaniolo [Zan] to describe such dependencies. The reader is referred to their papers for detailed discussion and examples.

Let R(U) be a relation, and let Y be a subset of U. For each subset X of U

(X and Y are not necessarily disjoint), and for each X-value, x, we define

$Y_R(x) = \{y \mid$ For some tuple $u \in R$, $u[X] = x$ and $u[Y] = y\}$

$Y_R$ is a function that gives, for each X and for each X-value, the set of Y-values that appear with this X-value in tuples of the relation. In terms of relational operations, the value of $Y_R(x)$ is computed by first selecting from R the tuples (if any) that have x as their x-projection and then projecting the resulting set of tuples on Y. The set $Y_R(x)$ is nonempty if and only if x appears in at least one tuple of R. The definition can be generalized to arguments that are sets of X-values in the obvious way. In particular, it is possible to consider composition, e.g., $Z_R(Y_R(x))$.

A multivalued dependency (abbr. MVD), g, on a set of attributes U is a statement g:X→→Y, where X and Y are subsets of U. Let Z be the complement of the union of X and Y in U. A relation R(U) obeys the MVD g:X→→Y if for every XZ-value, xz, that appears in R, we have $Y_R(xz) = Y_R(x)$. In words, the MVD g is valid in R if the set of Y-values that appears in R with a given x appears with every combination of x and z in R. Thus, this set is a function of x alone and does not depend on the Z-values that appear with x. Given g:X→→Y, we say that g is a multivalued dependency from X to Y (in the set U). As we do for FD's, here also we usually omit the name g of the MVD.

Remark: We allow X or Y to be the empty set. If Y is the empty set, then we get the MVD X→→∅; we agree that this MVD is valid in all relations. If X is the empty set, we get ∅→→Y which is valid in a relation if and only if the set of Y-values in the relation is independent of the values of all the other attributes in the relation. Let R(Y,Z) be a relation where Y and Z are disjoint. Intuitively, the MVD ∅→→Y is valid in R if and only if R is the Cartesian product of its projections R[Y] and R[Z]. This is in accordance with an alternative characterization of MVD's given by Fagin (see Proposition 2 below). For further discussion of these special cases, see [Fag2].

From the definitions it follows that if X→Y is valid in a relation R then X→→Y is also valid in R. Thus, each FD is also an MVD. The converse is not true. An MVD X→→Y is an FD only if for each x the set $Y_R(x)$ contains at most one element, which is not always the case.

There is a fundamental difference between the definitions of FD's and MVD's. An FD X→Y is defined in terms of the sets X and Y alone and can be interpreted as a constraint involving any set of attributes that contains X and Y. Indeed, to check if an FD X→Y is valid in a relation R(U) we only have to check its validity in the projection R[XY]; the validity does not depend on the values of the other attributes. On the other hand, the validity of X→→Y in R(U) depends on the values of all attributes in U and cannot be checked in R[XY]. It is possible that X→→Y is not valid in R(U) but that X→→Y is valid in the projection R[U'], where $U' \subset U$. Thus, MVD's are sensitive to "context" while functional dependencies are not. (See [Fag2] for further discussion of this issue). We will see later that this "context dependence" gives rise to an inference rule for MVD's that has no parallel among the rules for FD's. It follows that the specification of U is an intergral part of the MVD. It would be more appropriate, perhaps, to use the notation X→→Y(U) to stress the fact that the MVD involves the set U. However, in this paper we assume that U is a fixed given set of attributes. For this reason, we omit the reference to U and write simply X→→Y.

In Fagin's paper [Fag2], he required that the left and right sides of an MVD be disjoint. That is, for X→→Y to be defined, it is required that X and Y be disjoint. The reason for this restriction is that transitivity does not always hold if the restriction is lifted. We follow Zaniolo [Zan] in that we do not require this restriction. As we show in the next subsection, the use of the more general definition leads to a set of inference rules that are simple to state, easy to use and all but one of which (complementation) are similar to corresponding rules for FD's. This similarity is explored in [Beer2] to extend the results of [Bern1, Bern2] to multivalued dependencies.

Fagin's results for the restricted form of MVD's are related to our more general form by the following proposition, which is a direct consequence of the definition.

Proposition 1: For all sets of attributes X,Y and U such that $X,Y \subseteq U$ and for each relation R(U), the MVD X→→Y (on the set U) is valid in R if and only if X→→Y-X is valid in R. (Y-X is the set difference of Y and X.) □

In other words, the MVD X→→Y is equivalent to the restricted MVD X→→Y-X.

The following important proposition gives an alternative characterization of MVD's.

Proposition 2: Let X, Y and Z be sets such that their union is U and Y ∩ Z ⊆ X. For each relation R(U), the MVD X→→Y is valid in R if and only if R is the natural join of its projections R[XY] and R[XZ].

Proof: Fagin [Fag2] has proved the proposition for the case that X, Y and Z form a partition of U. It is straightforward to verify that the claim follows from this fact and Proposition 1. □

The most interesting special case of Proposition 2 is given by the following corollary.

Corollary: Let X and Y be subsets of U and let Z be the complement (in U) of the union of X and Y. For each relation R(U), the MVD X→→Y is valid in R if and only if R is the natural join of its projections R[XY] and R[XZ].

Proposition 2 (especially in the form given in the corollary) is probably the most important single property of multivalued dependencies. First, as shown in [Fag2] and in the next section, it can be used to prove various other properties of MVD's. Even more important is the relationship it establishes between MVD's and decompositions of relations. Essentially, the proposition means that an MVD is an alternative way of specifying that a relation can be expressed as the natural join of some of its projections, that is, that the relation can be decomposed into relations on smaller sets of attributes. Decomposition is a basic component of the theory of relational data bases. A well known example of its use is Codd's normalization process. More generally, decomposition can be viewed as part of the process of schema design. A possible approach to the problem of designing a relational schema is to consider all attributes to be initially contained in one big relation. The schema is the result of decomposing this relation into a set of relations on smaller sets of attributes. A suitable decomposition is determined by the relationships (e.g., dependencies) that exist among the attributes. (See [Fag3] for further discussion of decomposition.) Because of this close connection between the concept of an MVD and the concept of decomposition, we expect that MVD's will have an important role in the theory of relational data bases.

### 4.2 Inference Rules for MVD's

In this subsection we discuss "pure" inference rules for MVD's, that is, rules that deal with the implication of new MVD's from given MVD's. We will discuss "mixed" rules that involve both FD's and MVD's in the next subsection. In the following we assume that U is a given set of attributes and that all other sets of attributes are contained in it. All MVD's are to be interpreted in the context of U.

In Proposition 2, the sets Y and Z have symmetric roles. As a corollary we obtain the following inference rule.

MVD0 (Complementation): Let X, Y and Z be sets such that their union is U and Y ∩ Z ⊆ X. Then X→→Y if and only if X→→Z.

The rule is called the complementation rule since it is usually applied when Z is the complement (in U) of either Y or of the union of X and Y. It is the only rule for MVD's in which the consequense of applying the rule to some MVD depends (to some extent) on the underlying set U. In all the other rules only the sets that appear in the given dependencies participate in forming the sides of the result dependency. As an example of the use of the rule, we note that Proposition 1 is a consequence of the rule. Indeed, it follows by two applications of MVD0 that X→→Y iff X→→U-(Y-X) iff X→→Y-X.

We now proceed to present the additional inference rules that together with MVD0 constitute a complete set of inference rules for MVD's.

MVD1 (Reflexivity): If Y ⊆ X then X→→Y.

MVD2 (Augmentation): If Z ⊆ W and X→→Y then XW→→YZ.

MVD3 (Transitivity): If X→→Y and Y→→Z then X→→Z-Y.

The validity of the first two rules follows directly from the definition and we omit the proofs. The most interesting case of MVD2 occurs when Z = ∅. Fagin [Fag2] has proved transitivity, using

Proposition 2, for the special case where X, Y and Z are pairwise disjoint. (Without these restrictions, some of the MVD's in the rule are undefined according to his definition of MVD's.) Note that when Y and Z are disjoint, MVD3 gives us classical transitivity: if $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$, then $X \rightarrow\rightarrow Z$. We present here a direct proof for the general case.

We look first at the simple situation of FD's. If $X \rightarrow Y$ and $Y \rightarrow Z$ then the proof of $X \rightarrow Z$ is almost immediate. It is true for all sets X, Y and Z that $Z_R(x) \subseteq Z_R(Y_R(x))$. If $X \rightarrow Y$ then $Y_R(x)$ contains a single element; if also $Y \rightarrow Z$ then $Z_R(Y_R(x))$ contains a single element and so is equal to $Z_R(x)$. The fact that $Z_R(x)$ contains a single element means that $X \rightarrow Z$. For MVD's, however, the situation is more complicated. Even if $Z_R(x) = Z_R(Y_R(x))$, the condition for $X \rightarrow\rightarrow Z$ may not be satisfied.

By Proposition 1 we know that $Y \rightarrow\rightarrow Z$ is equivalent to $Y \rightarrow\rightarrow Z\text{-}Y$. Therefore, without loss of generality, we assume that Y and Z are disjoint. Let W be the complement in U of the union of X and Z. Let R(U) be a relation in which $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$ are valid. To prove that $X \rightarrow\rightarrow Z$ we have to show that $Z_R(x) = Z_R(xw)$ for each XW-value, xw, that appears in R.

First we show that, for each y in $Y_R(x)$, we have $Z_R(x) = Z_R(xy)$. Clearly, $Z_R(xy) \subseteq Z_R(x)$. If there exists some z in the set difference $Z_R(x) - Z_R(xy)$ then a) the combination xz appears in some tuples of R, and b) there is no tuple u' in R such that u'[X] = x, u'[Z] = z and u'[Y] = y hold for u'. It follows that $Y_R(xz)$ does not contain y, although $Y_R(x)$ does contain y. Since Y and Z are disjoint, this contradicts the given MVD $X \rightarrow\rightarrow Y$.

Now, since $Y \rightarrow\rightarrow Z$, we have immediately that $Z_R(xy) = Z_R(y)$ and, by the previous paragraph, $Z_R(x) = Z_R(y)$. This is true for each y in $Y_R(x)$. Since we want to show that $Z_R(xw) = Z_R(x)$, we need only show that $Z_R(xw) = Z_R(y)$ for some y in $Y_R(x)$. But, since Z and Y are disjoint, we know that $Y \subseteq XW$. Therefore xw has a projection y' on Y. This y' belongs to $Y_R(x)$ and, since $Y \rightarrow\rightarrow Z$, it follows that $Z_R(xw) = Z_R(y')$. This concludes the proof.

The rule MVD3 cannot be generalized. If Y and Z are not disjoint then it is not always true that $X \rightarrow\rightarrow Z$. See [Fag2] for counterexamples. In particular, if Z is a subset of Y then $Y \rightarrow\rightarrow Z$; however, there are many examples of MVD's $X \rightarrow\rightarrow Y$ such that there is no MVD from X to a subset of Y.

In the next section we prove that the rules MVD0 - MVD3 are complete for the family of MVD's. However, a minimal complete set of rules is not necessarily convenient to use. As we did for FD's, we introduce here some additional rules that we have found to be useful for the manipulation of MVD's.

MVD4 (Pseudo-transitivity):
If $X \rightarrow\rightarrow Y$ and $YW \rightarrow\rightarrow Z$
then $XW \rightarrow\rightarrow Z\text{-}YW$.

This rule is a consequence of MVD2 and MVD3. Using augmentation on $X \rightarrow\rightarrow Y$ we obtain $XW \rightarrow\rightarrow YW$, then $XW \rightarrow\rightarrow Z\text{-}YW$ is obtained by transitivity.

MVD5 (Union): If $X \rightarrow\rightarrow Y_1$ and $X \rightarrow\rightarrow Y_2$
then $X \rightarrow\rightarrow Y_1Y_2$.

MVD6 (Decomposition): If $X \rightarrow\rightarrow Y_1$ and $X \rightarrow\rightarrow Y_2$
then $X \rightarrow\rightarrow Y_1 \cap Y_2$,
$X \rightarrow\rightarrow Y_1 - Y_2$ and
$X \rightarrow\rightarrow Y_2 - Y_1$.

While MVD5 is exactly parallel to FD5, rule MVD6 is more restricted then FD6. We cannot decompose a set that appears on the right side of an MVD in an arbitrary manner. The only decomposition allowed is the result of Boolean operations on several sets that depend on a common set.

The rules MVD5 and MVD6 are very useful for the manipulation of MVD's. The corresponding rules for FD's, namely FD5 and FD6, are heavily relied upon in essentially all the papers that deal with FD's (e.g., [Beer1, Bern1, Bern2]). The rules MVD5 and MVD6, though somewhat restricted, are equally useful. For example, they are used in the completeness proof of the next section. They are also used in [Beer2], e.g., in the construction of an efficient algorithm to decide if a given MVD can be derived from a given set of MVD's.

We prove here only MVD5. It is an easy exercise to prove MVD6 from MVD5 and MVD0. Starting with $X \rightarrow\rightarrow Y_2$, we augment both sides by X to obtain $X \rightarrow\rightarrow XY_2$.

Similarly, augmenting $X \to\to Y_1$ by $Y_2$, we obtain $XY_2 \to\to Y_1Y_2$. If we were dealing with FD's, we could apply transitivity to terminate the proof. However, $XY_2$ and $Y_1Y_2$ are not disjoint, and applying MVD3 we get only $X \to\to Y_1$. Instead, we first apply complementation to obtain $XY_2 \to\to U-X-Y_1Y_2$. Now the sides are disjoint and we apply transitivity and then complementation once more to obtain the desired result $X \to\to Y_1Y_2$.

Currently, we know of no way to derive MVD5 or MVD6 from MVD1 - MVD3, that is, without using complementation. We conjecture that complementation must be used. This is not just a theoretical problem. In [Beer2] it is shown that for the synthesis of relational schemas from MVD's, one needs to consider a set of inference rules that does not contain MVD0. It is therefore interesting to know if in such a set MVD5 and MVD6 are independent of the other rules.

### 4.3 Mixed Inference Rules

In the previous subsection we dealt with the following problem: Given a set of MVD's, what are the additional MVD's implied by the set? We now turn our attention to the more general problem. Suppose we are given a set F of FD's, and a set G of MVD's. We can apply rules FD1-FD6 to F to obtain additional FD's; we can also apply MVD0 - MVD6 to G to obtain additional MVD's. Are there any additional dependencies that are implied by F ∪ G? What are the inference rules that can be used to derive them? One such rule has already been mentioned implicitly, namely, that each FD is also an MVD. Thus we can apply MVD0 - MVD6 to F ∪ G, not only to G. It also turns out that certain combinations of FD's and MVD's imply additional FD's that cannot be derived by the use of the above rules. These observations lead us to introduce here three additional rules.

FD-MVD1: If $X \to Y$ then $X \to\to Y$.

FD-MVD2: If $X \to\to Z$ and $Y \to Z'$ ($Z' \subseteq Z$), where Y and Z are disjoint, then $X \to Z'$.

FD-MVD1 follows from the definitions. We prove FD-MVD2. Let R(U) be a relation in which $X \to\to Z$ and $Y \to Z'$ are valid. Since Y and Z are disjoint, it follows from the definition of an MVD that $Z_R(x) = Z_R(xy)$ for each xy that appears in the relation. Since $Z' \subseteq Z$, it follows that also $Z'_R(x) = Z'_R(xy)$. But from $Y \to Z'$ it follows that $XY \to Z'$. So $Z'_R(x)$ which is equal to $Z'_R(xy)$ contains a single element and $X \to Z'$.

Note that if we are given a set F of FD's and a set G of MVD's, the use of the rule FD-MVD2 can derive new FD's that possibly cannot be derived from the set F alone. Similarly, using rule FD-MVD1 on these new FD's, we can derive additional MVD's that possibly cannot be derived from G and F without using rule FD-MVD2. As an example, suppose that we have the attributes A, B, C, D, E and we are given the dependencies $D \to C$ and $A \to\to BC$. Using rule FD-MVD2 we can derive the dependency $A \to C$. We leave it to the reader to check that this dependency cannot be derived from the given two dependencies using all rules except FD-MVD2.

The next rule we introduce is not independent; rather, it can be derived from the rules for MVD's and the rules FD-MVD1, FD-MVD2. We present it for the same reason we introduced additional rules for FD's and MVD's - to obtain a set of useful and flexible rules. For more details on the use of this rule see [Beer2].

FD-MVD3: If $X \to\to Y$ and $XY \to Z$, then $X \to Z-Y$.

To prove this rule, we first augment $X \to\to Y$ by X to obtain $X \to\to XY$. Applying transitivity we obtain $X \to\to Z-XY$. Now, $XY \to Z$ is equivalent to $XY \to Z-XY$ so by applying FD-MVD2 to $X \to\to Z-XY$ and $XY \to Z-XY$, we obtain $X \to Z-XY$. Clearly, this implies $X \to Z-Y$.

## 5. COMPLETENESS OF THE RULES

In this section we prove that the set of rules introduced in Sections 3 and 4 is complete for the family of functional and multivalued dependencies.

Let F and G be sets of FD's and MVD's (on a set U), respectively. The closure of F ∪ G, denoted by $(F,G)^+$, is the set of all FD's and MVD's that can be derived from F ∪ G by repeated applications of the rules in the set {FD1, FD2, FD3, MVD0, MVD1, MVD2, MVD3, FD-MVD1, FD-MVD2}. By the results of the previous sections and those of [Arm, Fag1], each of these rules is a valid inference rule for the family of dependencies. Therefore, each dependency in $(F,G)^+$ is implied by F ∪ G, i.e., it is valid in each relation that obeys all the dependencies in F and G. To prove completeness of this set of rules, it remains to show that the converse is also true, that is, that each dependency that is implied by $F \cup G$ belongs to $(F,G)^+$.

Recall that a dependency f is implied by $F \cup G$ if there is no counterexample relation such that all dependencies in $F \cup G$ are valid in it but f is not. To show completeness of the rules, we have to show that for each dependency not in $(F,G)^+$ such a counterexample relation does exist.

In the following we assume that sets F and G of FD's and MVD's, respectively, are given. Before we present the completeness theorem we need a few more concepts.

Let X be a subset of U. There are several sets Y such that the MVD $X \twoheadrightarrow Y$ is in $(F,G)^+$, (e.g., $X \twoheadrightarrow U-X$ is always in $(F,G)^+$). Following Fagin, we use the notation $X \twoheadrightarrow Y_1 | Y_2 | \ldots | Y_k$ to denote the collection of MVD's $X \twoheadrightarrow Y_1$, $X \twoheadrightarrow Y_2$, ..., $X \twoheadrightarrow Y_k$. From now on, when this notation is used, we assume that none of the sets $Y_1,\ldots,Y_k$ is the empty set.

Let us denote by DEP(X) the family of all sets Y for which $X \twoheadrightarrow Y$. (DEP(X) is, of course, a function of the given sets of dependencies F and G.) We have seen that DEP(X) is closed under Boolean operations ($MVD^c$, MVD6). Therefore, it contains a unique subfamily with the following properties:

a) The sets in the subfamily are nonempty.

b) Each pair of sets in the subfamily is disjoint.

c) Each set in DEP(X) is a union of sets from the subfamily.

This subfamily consists of all nonempty minimal sets in DEP(X), i.e., those sets that do not contain any other nonempty set of DEP(X). We call this subfamily the <u>dependency basis</u> of X. If $Y_1,\ldots,Y_k$ are the sets in the dependency basis of X, then as Fagin noted [Fag2], the "generalized" MVD $X \twoheadrightarrow Y_1 | \ldots | Y_k$ contains all the information about MVD's that have X as their left side.

Let X* denote the set of all attributes that are functionally dependent on X (by functional dependencies in $(F,G)^+$). Clearly $X \subseteq X^*$. X* has the same role for FD's as DEP(X) has for MVD's. For each $A \in X^*$ we have $X \rightarrow A$. Thus, each element of X* appears as a singleton set in the dependency basis of X. The dependency basis contains other sets if and only if X* is a proper subset of U. These remaining sets cover U-X*. We note that since $X \rightarrow X^*$ and $X^* \rightarrow X$, it follows that DEP(X) = DEP(X*) and the dependency basis of X is the same as the dependency basis of X*.

<u>Theorem 1</u> (<u>Completeness theorem for FD's and MVD's</u>): Let F and G be sets of FD's and MVD's (on a set U), respectively. For each functional or multivalued dependency that does not belong to $(F,G)^+$ there exists a relation R(U) such that all the dependencies in $(F,G)^+$ are valid in R but the given dependency is not valid in R.

<u>Proof</u>: Let X be the left side of the dependency that is not in $(F,G)^+$. The set X* is a proper subset of U since otherwise every (functional or multivalued) dependency with left side X belongs to $(F,G)^+$. Let $W_1,\ldots,W_m$ $(m \geq 1)$ be the sets in the dependency basis of X that cover U-X*. Thus, X*, $W_1,\ldots,W_m$ form a partition of U. The MVD $X \twoheadrightarrow X^* | W_1 | \ldots | W_m$ is in $(F,G)^+$ and, furthermore, if an MVD in $(F,G)^+$ has X as its left side then its right side is a union of a subset of X* and some of the sets $W_1,\ldots,W_m$.

The relation R(U) is constructed as follows: We choose the set {0,1} as the domain of each of the attributes in U. The relation R has $2^m$ rows, one row for each sequence of zeros and ones of length m. In the row corresponding to a sequence $\langle a_1,\ldots,a_m \rangle$ (where $a_i \in \{0,1\}$), each of the attributes in $W_i$ is assigned the value $a_i$ $(i=1,\ldots,m)$. Each attribute in X* is assigned the value 1 in all the rows of the relation. For example, if m=3, then the row corresponding to the sequence <0,1,1> has all 1's in the X* columns, all 0's in the $W_1$ columns, all 1's in the $W_2$ columns and all 1's in the $W_3$ columns.

We now want to prove that R satisfies the condition of the theorem. In what follows, we use the inference rules presented in the previous sections for two different purposes. First, we sometimes show that if some given dependencies are in $(F,G)^+$ then $(F,G)^+$ also contains some other dependency. That we can use the rules for this purpose follows directly from the definition of $(F,G)^+$. Second, we also show that if some dependencies are valid in R then there is another dependency that is valid in R. We can use the rules for this purpose since we have proved that they are valid inference rules for the family of dependencies, i.e., their application to dependencies that are valid in a relation always produces dependencies valid in that relation. We will indicate our intention each time we use the rules.

We now prove the following three claims about the relation R we have just constructed.

(1) If the right side of an FD is a nonempty subset of $W_i$ then the FD is valid in R if and only if its left side intersects $W_i$ (for $i=1,\ldots,m$).

(2) Each MVD that has $W_i$ as its right side is valid in R (for $i=1,\ldots,m$).

(3) If the right side of an MVD is a nonempty proper subset of $W_i$ then the MVD is valid in R if and only if its left side intersects $W_i$ (for $i=1,\ldots,m$).

We first prove one direction of claim 1. For each fixed row of R, all the attributes in $W_i$ have the same value. It follows that every attribute in $W_i$ is functionally dependent in R on every other attribute in $W_i$ (and, by augmentation, on every set that contains such an attribute). Thus we have proved that if the left side of the FD intersects $W_i$ then the FD is valid in R. From this also follows the corresponding direction of claim 3, since every FD is also an MVD.

We now prove claim 2 and the other directions of claims 1 and 3. The relation R is the Cartesian product of its projections $R[W_i]$ and $R[U-W_i]$. It follows immediately that the MVD $\emptyset \rightarrow\rightarrow W_i$ is valid in R and, by augmentation, $Y \rightarrow\rightarrow W_i$ is valid in R, for every set Y. This proves claim 2. Now let Y and Z be sets such that Y is disjoint from $W_i$ and Z is a nonempty subset of $W_i$. It follows from the above factorization of R that for each Y-value y, the set $Z_R(y)$ contains two Z-values - a 0-assignment and a 1-assignment to the attributes of Z. In particular, the FD $Y \rightarrow Z$ is not valid in R; this concludes the proof of claim 1. If Z is a proper subset of $W_i$ let A be an attribute in $W_i - Z$. Then $Z_R(ya)$, where ya is a YA-value, contains only a single Z-value since the attributes of Z must be assigned the same value as the attribute A. Therefore $Z_R(y) \neq Z_R(ya)$ and it follows that the MVD $Y \rightarrow\rightarrow Z$ is not valid in R. This concludes the proof of claim 3.

We now show that R satisfies the condition of the theorem. First, let f be an FD in $(F,G)^+$. We show that f is valid in R. By FD6 we can assume that f is of the form $Y \rightarrow B$ where B is a single attribute. Now, if B is in $X^*$ then f is clearly valid in R since in R every attribute of $X^*$ assumes a single value and is, there fore, functionally dependent on any other attribute. If B is not in $X^*$ then it belongs to some $W_i$. If Y is disjoint from this $W_i$ then from the MVD $X \rightarrow\rightarrow W_i$ and the FD $Y \rightarrow B$ which are both in $(F,G)^+$ it would follow by rule FD-MVD2 that $X \rightarrow B$ is in $(F,G)^+$. This is impossible since B is not in $X^*$. Therefore Y must intersect $W_i$ and $Y \rightarrow B$ is valid in R by claim 1.

Let now $g: Y \rightarrow\rightarrow Z$ be an MVD in $(F,G)^+$. We show that g is valid in R. We note that $Y \rightarrow Z \cap X^*$ is valid in R. We will show that, for each i, $Y \rightarrow\rightarrow Z \cap W_i$ is also valid in R. First, suppose that, for some i, the set $Z \cap W_i$ is either empty or all of $W_i$. By claim 2 above, $Y \rightarrow\rightarrow W_i$ is valid in R; as we have noted in Section 4, $Y \rightarrow\rightarrow \emptyset$ is always valid. Next, suppose that, for some i, $Z \cap W_i$ is a nonempty proper subset of $W_i$. If Y does not intersect $W_i$ we can use augmentation on $Y \rightarrow\rightarrow Z$ to obtain that $U-W_i \rightarrow\rightarrow Z$ is in $(F,G)^+$. Since $X \rightarrow\rightarrow U-W_i$ is also in $(F,G)^+$ it follows by transitivity (MVD3) that $X \rightarrow\rightarrow Z-(U-W_i)$, that is $X \rightarrow\rightarrow Z \cap W_i$ is in $(F,G)^+$. This is a contradiction to the assumption that $W_i$ is a member of the dependency basis of X. Thus Y must intersect $W_i$ and, by claim 3, $Y \rightarrow\rightarrow Z \cap W_i$ is valid in R. We have now shown that, for each i, $Y \rightarrow\rightarrow Z \cap W_i$ is valid in R and so is also $Y \rightarrow\rightarrow Z \cap X^*$. By taking the union (MVD5) it follows that $Y \rightarrow\rightarrow Z$ is valid in R.

Finally, let us consider the dependency (with left side X) which is known not to be in $(F,G)^+$. If it is an FD $X \rightarrow Y$ then Y is not a subset of $X^*$, so Y intersects $W_i$ for some i. By FD6 if $X \rightarrow Y$ is valid in R so is $X \rightarrow Y \cap W_i$ and this contradicts claim 1. (Recall that $X^*$ is disjoint from each of the $W_i$.) Therefore $X \rightarrow Y$ is not valid in R. If the dependency is an MVD $X \rightarrow\rightarrow Y$ then, for some i, $Y \cap W_i$ must be a nonempty proper subset of $W_i$ (else since $X \rightarrow Y \cap X^*$ and $X \rightarrow\rightarrow Y \cap W_i$ for each i are in $(F,G)^+$ the MVD $X \rightarrow\rightarrow Y$ would be in $(F,G)^+$). Now, for this i, $X \rightarrow\rightarrow Y \cap W_i$ is not valid in R by claim 3. Since $X \rightarrow\rightarrow W_i$ is valid in R, if $X \rightarrow\rightarrow Y$ were also valid in R we could apply MVD6 (intersection) to obtain a contradiction. Thus $X \rightarrow\rightarrow Y$ is not valid in R. □

By the theorem, we know that any single dependency not in $(F,G)^+$ is not implied by $F \cup G$. One might ask whether it is possible that the set $F \cup G$ implies a statement like '$f_1$ or $f_2$', where $f_1$ and $f_2$ do not belong to $(F,G)^+$. The meaning of such an implication is that either $f_1$ or $f_2$ must be valid in each relation in which $F \cup G$ is valid though none of them is valid in all such relations. Our next theorem states that this is not the case. As we have noted at the end of Section 2, this means that we are proving completeness of the rules using a concept of completeness that is stronger than the one defined there. We note that Fagin [Fag1] presents an example of a system in which completeness holds but in which "strong completeness" fails.

Theorem 2 (Strong completeness theorem): Let F and G be sets of FD's and MVD's (on a set U), respectively. There exists a relation R(U) such that the set of dependencies valid in R is exactly $(F,G)^+$.

Proof: We want to show that there exists a relation in which all dependencies in $(F,G)^+$ are valid and in which no other dependency is valid. We have seen in the proof of Theorem 1 that for each dependency f not in $(F,G)^+$ there exists a relation $R_f(U)$ in which all dependencies in $(F,G)^+$ are valid, f is not valid and all attributes assume only values from the set $\{0,1\}$. Let us now use for each dependency f a distinct set of values $\{0_f,1_f\}$. The required relation R(U) is the union of these relations $R_f(U)$ over all dependencies f not in $(F,G)^+$. It is easy to see that R satisfies the condition of the theorem. □

Note that our approach to proving strong completeness is somewhat different from Armstrong's approach. Armstrong proved strong completeness for FD's. For that he showed how to directly construct the counterexample relation in which the dependencies in the closure of a given set are valid and no other dependency is valid. The resulting relation is quite cumbersome. In our approach, one first constructs a counterexample relation for each dependency that is not in the closure of the given set, as in the proof of Theorem 1; then these relations are glued together as in the proof of Theorem 2. The ensuing relation is more easily understood than Armstrong's relation.

In the previous theorems we stated the completeness of our rules for the family of all dependencies. We now present completeness results for the subfamily of FD's and for the subfamily of MVD's. We first show that the rules FD1 - FD3 are complete for FD's. This result means that every FD that is derivable from a given set of FD's by using all nine rules (FD1-FD3, MVD0-MVD3, FD-MVD1,FD-MVD2) presented in this paper, can be derived from it using only the three rules FD1-FD3. Similarly, we show that the rules MVD0-MVD3 are complete for MVD's. That is, every MVD derivable from a given set of MVD's can be derived from it using only these four rules.

The completeness result for FD's is not new, of course. The proof presented here is essentially the proof given in [Fag1]. We include it here for several reasons. Firstly, we want to present all completeness results in one place. Secondly, this is the first time that the completeness issue is discussed in the context of FD's and MVD's together. Until now, every known inference rule for FD's was either one of FD1-FD3 or could be shown to be provable from them (e.g., FD4, FD5 and FD6). Now we have two additional rules, namely FD-MVD2 and FD-MVD3, which are not logical consequences of FD1-FD3. We want to stress the fact that, when only FD's are given, the rules FD1-FD3 are sufficient to derive all derivable FD's. (The reader should note, however, that this fact does follow from the definition of completeness and the proofs of completeness of this set of rules in [Arm, Fag1]. We want to stress a known result, not to prove a new result.) Finally, there is an interesting corollary that follows from the proof.

Theorem 3 (Strong completeness theorem for FD's): The rules FD1, FD2, FD3 are strongly complete for the family of FD's.

Proof: We have to show that, for every set of FD's, there exists a relation in which all the FD's that are derivable from the given set by using the rules FD1-FD3 are valid and in which no other FD is valid. What we will show is that, given a set of FD's, if an FD cannot be derived from it by using these three rules, then it cannot be derived from it at all (that is, by using all nine rules). The desired result will then follow as a corollary of Theorem 2. We use the same technique that was used in the proof of Theorem 1 but the proof here is much simpler.

We first note that the rules FD5 and

and FD6 can be proved from FD1-FD3. **This means that an application of** FD5 or FD6 is equivalent to a sequence of applications of rules from the set {FD1, FD2, FD3}. Therefore we can assume in the following, without loss of generality, that all FD's have a single attribute on their right side.

Let F be a given set of FD's on a set of attributes U, and let f:X—>A be an FD that cannot be derived from F by using the rules FD1-FD3. Let us denote by $\bar{X}$ the set of attributes that are functionally dependent on X by some FD derivable from F by using the rules FD1-FD3. (We note that, because of FD1, $\bar{X}$ contains X. It is also clear that $\bar{X}$ is contained in X* - the set of attributes that are functionally dependent on X by some FD in $F^+$. A priori, the latter containment may be proper. However, it follows from the theorem that $\bar{X}$ = X*. We cannot, of course, use this equality until the theorem has been proven.) Let R(U) be a relation consisting of the following two tuples. In the first tuple of R all the attributes are assigned the value 1. In the second tuple of R all attributes of $\bar{X}$ are assigned the value 1 and all the attributes in U-$\bar{X}$ are assigned the value 0. Note that U-$\bar{X}$ is not empty since it contains the attribute A.

It is easy to see that a) every FD whose right side is contained in $\bar{X}$ is valid in R, and b) an FD whose right side intersects U-$\bar{X}$ is valid in R if and only if its left side also intersects U-$\bar{X}$. (Compare to claims 1 and 3 in the proof of Theorem 1.) Let us now consider an arbitrary FD Y—>B in the given set F. We will show that the FD is valid in R. If Y is a subset of $\bar{X}$ then X—>Y can be derived from F by the rules FD1-FD3; since the FD Y—>B is in F, we obtain by FD3 (transitivity) that B is in $\bar{X}$ and Y—>B is valid in R. If Y intersects U-$\bar{X}$ then, by the two claims above, Y—>B is valid in R. It follows that all the FD's in F are valid in R and therefore all FD's and MVD's in $F^+$ are valid in R.

Now let us consider the given FD f:X—>A. Since it cannot be derived from F by using the rules FD1-FD3, A belongs to U-$\bar{X}$. It follows from (b) above that f is not valid in R and therefore that f is not in $F^+$. □

In many papers dealing with FD's (e.g., [Arm, Bern1, Bern2, Fag1]), the closure of a set of FD's is defined to be the set of all FD's derivable from it by using the rules FD1-FD3. This definition is more restricted than our definition which includes in the closure all FD's and MVD's derivable from the set by using all nine rules presented in this paper. However, it follows from the theorem that, as far as FD's are concerned, the two definitions are equivalent. Clearly, in any work that deals only with FD's and in which the concept of MVD does not appear, no ambiguity can arise if the more restricted definition is used.

As a corollary to the proof of Theorem 3 we can now obtain a simple characterization of the MVD's in the closure of a set F of FD's. Namely, each MVD in $F^+$ is either one of the FD's in $F^+$ or is the result of applying MVD0 (complementation) to such an FD. We note that this corollary can also be obtained from a result by Rissanen [Riss, The.1].

<u>Corollary</u>: Let F be a set of FD's on a set of attributes U. Let g:X—>—>Y be an MVD in $F^+$ and let $\bar{g}$ be the MVD $\bar{g}$:X—>—>U-Y (which is also in $F^+$). Then either g or $\bar{g}$ is the result of applying rule FD-MVD1 to an FD in $F^+$. (Recall that rule FD-MVD1 essentially states that every FD is also an MVD. Thus the corollary means that either g or its 'complement' $\bar{g}$ is essentially an FD in $F^+$.)

<u>Proof</u>: Let the set X be the left side of the given MVD g. We consider the relation R(U) constructed in the proof of Theorem 3, in which all attributes of X* (= $\bar{X}$) are assigned the value 1 in both tuples and the attributes of U-X* are assigned the value 1 in the first tuple and the value 0 in the second tuple. (For this X it is possible that U-X* is empty, since now no attribute A is given that is known to be in U-X*.) We want to show that either X—>Y or X—>U-Y is an FD in $F^+$. Since g is in $F^+$, we know that g is valid in R. Now, if Y is a subset of X* then X—>Y is an FD in $F^+$. If Y intersects U-X* then it is obvious that X—>—>Y is valid in R only if Y contains all attributes of U-X*. (Compare to claims 2 and 3 in the proof of Theorem 1.) But then U-Y is a subset of X* and X—>U-Y is an FD in $F^+$. □

We now present the completeness result for MVD's.

<u>Theorem 4 (Strong completeness theorem for MVD's)</u>: The rules MVD0, MVD1, MVD2, MVD3 are complete for the family of MVD's.

<u>Proof</u>: Let G be a given set of MVD's. We want to show that there exists a re-

lation in which all MVD's derivable from G by the use of the rules MVD0-MVD3 are valid and in which no other MVD is valid. Let R be the relation constructed in the proof of Theorem 2 where G is the given set of MVD's and F is the empty set of FD's. As we showed in the proof of Theorem 2, all MVD's derivable from the set F ∪ G (which is equal to G, since F is empty) by using the rules FD1-FD3, MVD0-MVD3, FD-MVD1 and FD-MVD2 are valid in R and no other MVD's are valid in R.

Now, the only way to derive new MVD's in $G^+$ without using the rules MVD0-MVD3 is by applying rule FD-MVD1 to some FD's. However, since F is the empty set, initially we have no FD's. The only way to generate new FD's when no FD's are given is to use rule FD1 which generates all the "reflexive" FD's (of the form X→Y where Y is a subset of X). It is easy to see that rule FD-MVD2 cannot be applied when only reflexive FD's are given and that the other FD-generating rules (FD2, FD3) cannot generate nonreflexive FD's from reflexive FD's. Thus, the closure of G contains only reflexive FD's and applying rule FD-MVD1 we get only the reflexive MVD's. However, these are also generated by rule MVD1. To conclude, all MVD's valid in R are generated from G by the rules MVD0-MVD3 as was to be shown. ❑

Note that now we also have a characterization of all FD's in the closure of a given set G of MVD's, namely, every FD in $G^+$ is a reflexive FD. It follows that when one is interested in MVD's, no ambiguity can arise if the closure of a given set of the MVD's is defined to be the set of MVD's derivable from it by using the rules MVD0-MVD3.

## 6. CONCLUSION

In this paper we have investigated the inference rules for functional and multivalued dependencies in a data base relation. A set of rules was presented and shown to be complete for the family of functional and multivalued dependencies. It was also shown that the subset of rules that apply to multivalued dependencies is a complete set of rules for these dependencies. Thus, the rules presented here are sufficient for the analysis of the properties of functional and multivalued dependencies.

In addition, we have shown an analogy between the well known rules for functional dependencies and the rules for multivalued dependencies. Specifically, it was shown that for each rule for functional dependencies, the same rule or a similar rule is valid for multivalued dependencies. There is, however, an additional rule (complementation) for multivalued dependencies that has no parallel among the rules for functional dependencies. Of particular importance are the rules for the manipulation of right sides of dependencies - the Union and the Decomposition rules. These two rules were already known to be very useful for the manipulation of functional dependencies. While the Decomposition rule for multivalued dependencies is slightly less general, these rules are still very useful for the manipulation of these dependencies since they allow us to perform Boolean operations on the right sides of dependencies. It turns out that actually, in many cases where these rules are used to manipulate functional dependencies, Boolean operations are all that is needed. Therefore, similar manipulations can be applied to multivalued dependencies. An example of the use of these rules is our proof of Theorem 1. Another example is given in [Beer2].

We conclude by mentioning some problems that merit further research. Now that we understand the properties of dependencies in a relation, we should try to clarify the influence of such dependencies on the structure of relational schemas. Some work in this direction has already been done ([Bern1, Bern2, Beer2, Fag2]). However, a general theory that ties together dependencies, relations and operations on relations is still lacking. A specific problem in this direction is to investigate what happens to dependencies when relations are joined. We hope that the results presented in this paper will serve as a basis for attacking these problems.

## REFERENCES

[Arm] Armstrong, W. W., "Dependency Structures of Data Base Relationships", Proc. of IFIP '74, North Holland, 1974, pp. 580-583.

[Beer1] Beeri, C. and Bernstein, P. A., "On the Algorithmic Properties of Functional Dependencies", in preparation.

[Beer2] Beeri, C., "Multivalued Dependencies and the Construction of Relational Schemas", in preparation.

[Bern1] Bernstein, P.A., "Synthesizing

Third Normal Form Relations from Functional Dependencies", ACM Trans. on Database Systems 1, No. 4 (Dec. 1976), pp. 277-298.

[Bern2] Bernstein, P. A. and Beeri, C., "An Algorithmic Approach to Normalization of Relational Database Schemas", Technical Report CSRG-73, Computer Systems Research Group, University of Toronto, Sept. 1976.

[Codd1] Codd, E. F., "A Relational Model for Large Shared Data Bases", CACM 13, No. 6 (June 1970), pp. 377-387.

[Codd2] Codd, E. F., "Further Normalization of the Data Base Relational Model", in Data Base Systems (Courant Computer Science Symposium 6, R. Rustin ed.), Prentice-Hall, 1972, pp. 33-64.

[Fag1] Fagin, R., "Functional Dependencies in a Relational Database and Propositional Logic", Revision of IBM Research Report RJ 1776, San Jose. California, 1976.

[Fag2] Fagin, R., "Multivalued Dependencies and a New Normal Form for Relational Databases", to appear in ACM Trans. on Database Systems.

[Fag3] Fagin, R., "The Decomposition vs. the Synthetic Approach to Relational Database Design", IBM Research Report RJ1976, San Jose, CAlifornia, 1977.

[Riss] Rissanen, J., "Independent Components of Relations", IBM Research Report RJ1899, San Jose, California 1977.

[SS] Schmid, H. A. and Swenson, J. R., "On the Semantics of the Relational Data Model", Proc. of ACM SIGMOD Conf., San Jose, California, May 1975, pp. 211-223.

[Zan] Zaniolo, C., "Analysis and Design of Relational Schemata for Database Systems", (Ph.D thesis), Computer Science Department, UCLA, Technical Report UCLA-ENG-7769, July 1976.