The construction is completed by means of one large "garbage collection" component G, involving internal elements $g_1[k] \in X$ and $g_2[k] \in Y$, $1 \leqslant k \leqslant m(n-1)$, and external elements of the form $u_i[j]$ and $\bar{u}_i[j]$ from W. It consists of the following set of triples:

$$G = \{(u_i[j], g_1[k], g_2[k]), (\bar{u}_i[j], g_1[k], g_2[k]): \\ 1 \leqslant k \leqslant m(n-1), 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$$

Thus each pair $g_1[k], g_2[k]$ must be matched with a unique $u_i[j]$ or $\bar{u}_i[j]$ that does not occur in any triples of $M' - G$. There are exactly $m(n-1)$ such "uncovered" external elements, and the structure of G insures that they can always be covered by choosing $M' \cap G$ appropriately. Thus G merely guarantees that, whenever a subset of $M - G$ satisfies all the constraints imposed by the truth-setting and fan-out components, then that subset can be extended to a matching for M.

To summarize, we set

$$W = \{u_i[j], \bar{u}_i[j]: 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$$

$$X = A \cup S_1 \cup G_1$$

where

$$A = \{a_i[j]: 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$$
$$S_1 = \{s_1[j]: 1 \leqslant j \leqslant m\}$$
$$G_1 = \{g_1[j]: 1 \leqslant j \leqslant m(n-1)\}$$

$$Y = B \cup S_2 \cup G_2$$

where

$$B = \{b_i[j]: 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$$
$$S_2 = \{s_2[j]: 1 \leqslant j \leqslant m\}$$
$$G_2 = \{g_2[j]: 1 \leqslant j \leqslant m(n-1)\}$$

and

$$M = \left(\bigcup_{i=1}^{n} T_i\right) \cup \left(\bigcup_{j=1}^{m} C_j\right) \cup G$$

Notice that every triple in M is an element of $W \times X \times Y$ as required. Furthermore, since M contains only

$$2mn + 3m + 2m^2n(n-1)$$

triples and since its definition in terms of the given 3SAT instance is quite direct, it is easy to see that M can be constructed in polynomial time.

From the comments made during the description of M, it follows immediately that M cannot contain a matching unless C is satisfiable. We now must show that the existence of a satisfying truth assignment for C implies that M contains a matching.

Let $t: U \rightarrow \{T,F\}$ be any satisfying truth assignment for C. We construct a matching $M' \subseteq M$ as follows: For each clause $c_j \in C$, let $z_j \in \{u_i, \bar{u}_i: 1 \leqslant i \leqslant n\} \cap c_j$ be a literal that is set true by t (one must exist since t satisfies c_j). We then set

$$M' = \left[\bigcup_{t(u_i)=T} T_i^t\right] \cup \left[\bigcup_{t(u_i)=F} T_i^f\right] \cup \left[\bigcup_{j=1}^{m} \{(z_j[j], s_1[j], s_2[j])\}\right] \cup G'$$

where G' is an appropriately chosen subcollection of G that includes all the $g_1[k], g_2[k]$, and remaining $u_i[j]$ and $\bar{u}_i[j]$. It is easy to verify that such a G' can always be chosen and that the resulting set M' is a matching. ■

In proving NP-completeness results, the following slightly simpler and more general version of 3DM can often be used in its place:

EXACT COVER BY 3-SETS (X3C)

INSTANCE: A finite set X with $|X| = 3q$ and a collection C of 3-element subsets of X.

QUESTION: Does C contain an *exact cover* for X, that is, a subcollection $C' \subseteq C$ such that every element of X occurs in exactly one member of C'?

Note that every instance of 3DM can be viewed as an instance of X3C, simply by regarding it as an unordered subset of $W \cup X \cup Y$, and the matchings for that 3DM instance will be in one-to-one correspondence with the exact covers for the X3C instance. Thus 3DM is just a restricted version of X3C, and the NP-completeness of X3C follows by a trivial transformation from 3DM.

3.1.3 VERTEX COVER and CLIQUE

Despite the fact that VERTEX COVER and CLIQUE are independently useful for proving NP-completeness results, they are really just different ways of looking at the same problem. To see this, it is convenient to consider them in conjunction with a third problem, called INDEPENDENT SET.

An *independent set* in a graph $G = (V,E)$ is a subset $V' \subseteq V$ such that, for all $u,v \in V'$, the edge $\{u,v\}$ is *not* in E. The INDEPENDENT SET problem asks, for a given graph $G = (V,E)$ and a positive integer $J \leqslant |V|$, whether G contains an independent set V' having $|V'| \geqslant J$. The following relationships between independent sets, cliques, and vertex covers are easy to verify.

Lemma 3.1 For any graph $G=(V,E)$ and subset $V' \subseteq V$, the following statements are equivalent:

(a) V' is a vertex cover for G.
(b) $V-V'$ is an independent set for G.
(c) $V-V'$ is a clique in the *complement* G^c of G, where $G^c=(V,E^c)$ with $E^c=\{\{u,v\}: u,v \in V \text{ and } \{u,v\} \notin E\}$.

Thus we see that, in a rather strong sense, these three problems might be regarded simply as "different versions" of one another. Furthermore, the relationships displayed in the lemma make it a trivial matter to transform any one of the problems to either of the others.

For example, to transform VERTEX COVER to CLIQUE, let $G=(V,E)$ and $K \leq |V|$ constitute any instance of VC. The corresponding instance of CLIQUE is provided simply by the graph G^c and the integer $J=|V|-K$.

This implies that the NP-completeness of all three problems will follow as an immediate consequence of proving that any one of them is NP-complete. We choose to prove this for VERTEX COVER.

Theorem 3.3 VERTEX COVER is NP-complete.
Proof: It is easy to see that VC $\in$ NP since a nondeterministic algorithm need only guess a subset of vertices and check in polynomial time whether that subset contains at least one endpoint of every edge and has the appropriate size.

We transform 3SAT to VERTEX COVER. Let $U=\{u_1,u_2,\ldots,u_n\}$ and $C=\{c_1,c_2,\ldots,c_m\}$ be any instance of 3SAT. We must construct a graph $G=(V,E)$ and a positive integer $K \leq |V|$ such that G has a vertex cover of size K or less if and only if C is satisfiable.

As in the previous proof, the construction will be made up of several components. In this case, however, we will have only truth-setting components and satisfaction testing components, augmented by some additional edges for communicating between the various components.

For each variable $u_i \in U$, there is a truth-setting component $T_i=(V_i,E_i)$, with $V_i=\{u_i,\bar{u}_i\}$ and $E_i=\{\{u_i,\bar{u}_i\}\}$, that is, two vertices joined by a single edge. Note that any vertex cover will have to contain at least one of u_i and $\bar{u}_i$ in order to cover the single edge in E_i.

For each clause $c_j \in C$, there is a satisfaction testing component $S_j=(V'_j,E'_j)$, consisting of three vertices and three edges joining them to form a triangle:

$$V'_j = \{a_1[j],a_2[j],a_3[j]\}$$
$$E'_j = \{\{a_1[j],a_2[j]\},\{a_1[j],a_3[j]\},\{a_2[j],a_3[j]\}\}$$

Note that any vertex cover will have to contain at least two vertices from V'_j in order to cover the edges in E'_j.

The only part of the construction that depends on which literals occur in which clauses is the collection of communication edges. These are best viewed from the vantage point of the satisfaction testing components. For each clause $c_j \in C$, let the three literals in c_j be denoted by x_j, y_j, and z_j. Then the communication edges emanating from S_j are given by:

$$E''_j = \{\{a_1[j],x_j\},\{a_2[j],y_j\},\{a_3[j],z_j\}\}$$

The construction of our instance of VC is completed by setting $K=n+2m$ and $G=(V,E)$, where

$$V = (\bigcup_{i=1}^{n} V_i) \cup (\bigcup_{j=1}^{m} V'_j)$$

and

$$E = (\bigcup_{i=1}^{n} E_i) \cup (\bigcup_{j=1}^{m} E'_j) \cup (\bigcup_{j=1}^{m} E''_j)$$

Figure 3.3 shows an example of the graph obtained when $U=\{u_1,u_2,u_3,u_4\}$ and $C=\{\{u_1,\bar{u}_3,\bar{u}_4\},\{\bar{u}_1,u_2,\bar{u}_4\}\}$.

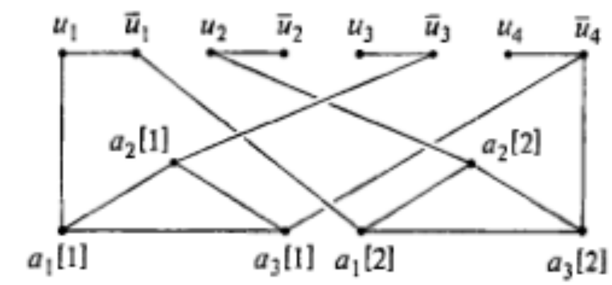

Figure 3.3 VERTEX COVER instance resulting from 3SAT instance in which $U=\{u_1,u_2,u_3,u_4\}$, $C=\{\{u_1,\bar{u}_3,\bar{u}_4\},\{\bar{u}_1,u_2,\bar{u}_4\}\}$. Here $K=n+2m=8$.

It is easy to see how the construction can be accomplished in polynomial time. All that remains to be shown is that C is satisfiable if and only if G has a vertex cover of size K or less.

First, suppose that $V' \subseteq V$ is a vertex cover for G with $|V'| \leq K$. By our previous remarks, V' must contain at least one vertex from each T_i and at least two vertices from each S_j. Since this gives a total of at least $n+2m=K$ vertices, V' must in fact contain *exactly* one vertex from each T_i and *exactly* two vertices from each S_j. Thus we can use the way in which V' intersects each truth-setting component to obtain a truth assignment $t: U \to \{T,F\}$. We merely set $t(u_i)=T$ if $u_i \in V'$ and $t(u_i)=F$ if

$\bar{u}_i \in V'$. To see that this truth assignment satisfies each of the clauses $c_j \in C$, consider the three edges in E_j''. Only two of those edges can be covered by vertices from $V_j' \cap V'$, so one of them must be covered by a vertex from some V_i that belongs to V'. But that implies that the corresponding literal, either u_i or $\bar{u}_i$, from clause c_j is true under the truth assignment t, and hence clause c_j is satisfied by t. Because this holds for every $c_j \in C$, it follows that t is a satisfying truth assignment for C.

Conversely, suppose that $t: U \to \{T,F\}$ is a satisfying truth assignment for C. The corresponding vertex cover V' includes one vertex from each T_i and two vertices from each S_j. The vertex from T_i in V' is u_i if $t(u_i) = T$ and is $\bar{u}_i$ if $t(u_i) = F$. This ensures that at least one of the three edges from each set E_j'' is covered, because t satisfies each clause c_j. Therefore we need only include in V' the endpoints from S_j of the other two edges in E_j'' (which may or may not also be covered by vertices from truth-setting components), and this gives the desired vertex cover. ■

3.1.4 HAMILTONIAN CIRCUIT

In Chapter 2, we saw that the HAMILTONIAN CIRCUIT problem can be transformed to the TRAVELING SALESMAN decision problem, so the NP-completeness of the latter problem will follow immediately once HC has been proved NP-complete. At the end of the proof we note several variants of HC whose NP-completeness also follows more or less directly from that of HC.

For convenience in what follows, whenever $<v_1, v_2, \ldots, v_n>$ is a Hamiltonian circuit, we shall refer to $\{v_i, v_{i+1}\}$, $1 \leqslant i < n$, and $\{v_n, v_1\}$ as the edges "in" that circuit. Our transformation is a combination of two transformations from [Karp, 1972], also described in [Liu and Geldmacher, 1978].

Theorem 3.4 HAMILTONIAN CIRCUIT is NP-complete

Proof: It is easy to see that HC $\in$ NP, because a nondeterministic algorithm need only guess an ordering of the vertices and check in polynomial time that all the required edges belong to the edge set of the given graph.

We transform VERTEX COVER to HC. Let an arbitrary instance of VC be given by the graph $G = (V,E)$ and the positive integer $K \leqslant |V|$. We must construct a graph $G' = (V',E')$ such that G' has a Hamiltonian circuit if and only if G has a vertex cover of size K or less.

Once more our construction can be viewed in terms of components connected together by communication links. First, the graph G' has K "selector" vertices $a_1, a_2, \ldots, a_K$, which will be used to select K vertices from the vertex set V for G. Second, for each edge in E, G' contains a "cover-testing" component that will be used to ensure that at least one endpoint of that edge is among the selected K vertices. The component for

$e = \{u,v\} \in E$ is illustrated in Figure 3.4. It has 12 vertices,

$$V_e' = \{(u,e,i),(v,e,i): 1 \leqslant i \leqslant 6\}$$

and 14 edges,

$$E_e' = \{\{(u,e,i),(u,e,i+1)\},\{(v,e,i),(v,e,i+1)\}: 1 \leqslant i \leqslant 5\}$$
$$\cup \{\{(u,e,3),(v,e,1)\},\{(v,e,3),(u,e,1)\}\}$$
$$\cup \{\{(u,e,6),(v,e,4)\},\{(v,e,6),(u,e,4)\}\}$$

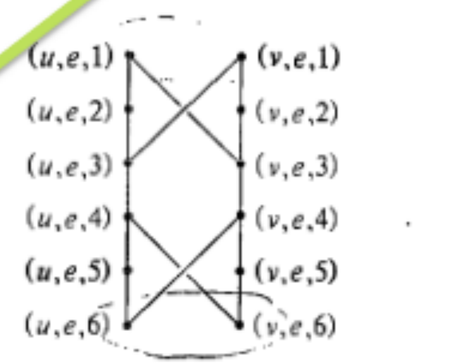

Figure 3.4 Cover-testing component for edge $e = \{u,v\}$ used in transforming VERTEX COVER to HAMILTONIAN CIRCUIT.

In the completed construction, the only vertices from this cover-testing component that will be involved in any additional edges are $(u,e,1)$, $(v,e,1)$, $(u,e,6)$, and $(v,e,6)$. This will imply, as the reader may readily verify, that any Hamiltonian circuit of G' will have to meet the edges in E_e' in exactly one of the three configurations shown in Figure 3.5. Thus, for example, if the circuit "enters" this component at $(u,e,1)$, it will have to "exit" at $(u,e,6)$ and visit either all 12 vertices in the component or just the 6 vertices (u,e,i), $1 \leqslant i \leqslant 6$.

Additional edges in our overall construction will serve to join pairs of cover-testing components or to join a cover-testing component to a selector vertex. For each vertex $v \in V$, let the edges incident on v be ordered (arbitrarily) as $e_{v[1]}, e_{v[2]}, \ldots, e_{v[deg(v)]}$, where $deg(v)$ denotes the *degree* of v in G, that is, the number of edges incident on v. All the cover-testing components corresponding to these edges (having v as endpoint) are joined together by the following connecting edges:

$$E_v' = \{\{(v,e_{v[i]},6),(v,e_{v[i+1]},1)\}: 1 \leqslant i < deg(v)\}$$

As shown in Figure 3.6, this creates a single path in G' that includes exactly those vertices (x,y,z) having $x = v$.