instantaneous codes, so one expects to achieve a lower expected codeword length if \( L \) is minimized over all uniquely decodable codes. In this section we prove that the class of uniquely decodable codes does not offer any further possibilities for the set of codeword lengths than do instantaneous codes. We now give Karush’s elegant proof of the following theorem.

**Theorem 5.5.1 (McMillan)** The codeword lengths of any uniquely decodable \( D \)-ary code must satisfy the Kraft inequality

\[
\sum D^{-l_i} \leq 1. \tag{5.48}
\]

Conversely, given a set of codeword lengths that satisfy this inequality, it is possible to construct a uniquely decodable code with these codeword lengths.

**Proof:** Consider \( C^k \), the \( k \)th extension of the code (i.e., the code formed by the concatenation of \( k \) repetitions of the given uniquely decodable code \( C \)). By the definition of unique decodability, the \( k \)th extension of the code is nonsingular. Since there are only \( D^n \) different \( D \)-ary strings of length \( n \), unique decodability implies that the number of code sequences of length \( n \) in the \( k \)th extension of the code must be no greater than \( D^n \). We now use this observation to prove the Kraft inequality.

Let the codeword lengths of the symbols \( x \in \mathcal{X} \) be denoted by \( l(x) \). For the extension code, the length of the code sequence is

\[
l(x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k} l(x_i). \tag{5.49}
\]

The inequality that we wish to prove is

\[
\sum_{x \in \mathcal{X}} D^{-l(x)} \leq 1. \tag{5.50}
\]

The trick is to consider the \( k \)th power of this quantity. Thus,

\[
\left( \sum_{x \in \mathcal{X}} D^{-l(x)} \right)^k = \sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} \cdots \sum_{x_k \in \mathcal{X}} D^{-l(x_1)} D^{-l(x_2)} \cdots D^{-l(x_k)} \tag{5.51}
\]

\[
= \sum_{x_1, x_2, \ldots, x_k \in \mathcal{X}} D^{-l(x_1)} D^{-l(x_2)} \cdots D^{-l(x_k)} \tag{5.52}
\]

\[
= \sum_{x^k \in \mathcal{X}^k} D^{-l(x^k)}. \tag{5.53}
\]
by (5.49). We now gather the terms by word lengths to obtain
\[ \sum_{x^k \in \mathcal{X}^k} D^{-l(x^k)} = \sum_{m=1}^{k_{\text{max}}} a(m) D^{-m}, \quad (5.54) \]
where \( l_{\text{max}} \) is the maximum codeword length and \( a(m) \) is the number of source sequences \( x^k \) mapping into codewords of length \( m \). But the code is uniquely decodable, so there is at most one sequence mapping into each code \( m \)-sequence and there are at most \( D^m \) code \( m \)-sequences. Thus, \( a(m) \leq D^m \), and we have
\[ \left( \sum_{x \in \mathcal{X}} D^{-l(x)} \right)^k = \sum_{m=1}^{k_{\text{max}}} a(m) D^{-m} \quad (5.55) \]
\[ \leq \sum_{m=1}^{k_{\text{max}}} D^m D^{-m} \quad (5.56) \]
\[ = k l_{\text{max}} \quad (5.57) \]
and hence
\[ \sum_j D^{-l_j} \leq (k_{\text{max}})^{1/k}. \quad (5.58) \]
Since this inequality is true for all \( k \), it is true in the limit as \( k \to \infty \). Since \( (k_{\text{max}})^{1/k} \to 1 \), we have
\[ \sum_j D^{-l_j} \leq 1, \quad (5.59) \]
which is the Kraft inequality.

Conversely, given any set of \( l_1, l_2, \ldots, l_m \) satisfying the Kraft inequality, we can construct an instantaneous code as proved in Section 5.2. Since every instantaneous code is uniquely decodable, we have also constructed a uniquely decodable code. \( \Box \)

**Corollary** A uniquely decodable code for an infinite source alphabet \( \mathcal{X} \) also satisfies the Kraft inequality.

**Proof:** The point at which the preceding proof breaks down for infinite \( |\mathcal{X}| \) is at (5.58), since for an infinite code \( l_{\text{max}} \) is infinite. But there is a