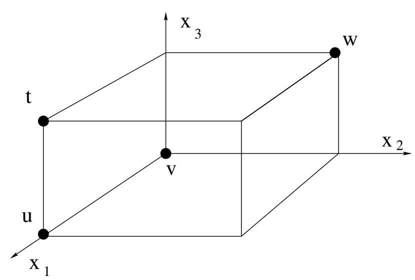
A HYPERCUBE PROBLEM

Charles Pinter
Mathematics Dept, Bucknell University, Lewisburg PA, *cpinter@bucknell.edu*

Many problems of combinatorics are easy to state and difficult to solve. This paper concerns such a problem: Consider the n-dimensional hypercube whose sides are of length 1. The set of vertices is then the set $\{0,1\}^n$ consisting of all n-tuples of 0s and 1s. The cube lies in a coordinate system having axes $x_1, x_2, \ldots, x_n$. We shall be concerned here with an arbitrary subset S of the vertices, that is, with $S \subseteq \{0,1\}^n$.

Any vertex $v \in S$ is characterized by its n-tuple of coordinates. But if $v \in S$, and S is a *proper* subset of $\{0,1\}^n$, it may happen that v can be unambiguously identified — that is, distinguished from all the other vertices in S — by giving only a few of its coordinates.

For example, in the subset $S = \{u, v, w, t\}$ of the cube shown below, u is completely identified by providing the information that $x_1 = 1$ and $x_3 = 0$. We shall say that $\{x_1, x_3\}$ is a "specifying set" for u.



Let the symbol $u(x_i)$ denote the value (0 or 1) of the x_i-coordinate of the vertex u. If S is an arbitrary subset of $\{0,1\}^n$ and $v \in S$, the following definition should be clear: The set of variables $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ is called a *specifying set* for v if the following is true $\forall u \in S$:

$$[u(x_{i_j}) = v(x_{i_j}) \quad \text{for} \quad j = 1, \ldots, k] \quad \Rightarrow \quad u = v.$$

The *specification number* of v is the cardinality of the smallest specifying set of v in S. The *mean specification number* of S is the average of the specification numbers of the elements of S.

One is tempted to conjecture that if S has m elements, then the mean specification number for S is no greater than $\log_2 m$, irrespective of the dimension of the cube. The problem of proving this conjecture is referred to here as the Specification Bound Problem. There is an ancillary problem, not as glamorous but equally important, to be called the Specification Coding Problem: What is the *least* possible value of the mean specification number of any m-element set $S \subseteq \{0,1\}^n$? Another perspective on this problem: How should m vertices be configured in the n-cube so as to minimize their average specification number? Both problems arise from (and have applications in) the theory of machine learning.

In this paper we construct a conceptual framework for the study of these questions within the framework of extremal combinatorics. We solve the Specification Bound Problem, using a version of the Kruskal-Katona Theorem. The Coding Problem is more difficult, but we offer several conjectures.