The Mathematics of Error-Correcting Codes
An Outline for an Oberwolfach Seminar

1 Background

Shannon's ground-breaking work in which *Information Theory* was born is generally considered one of the greatest scientific achievements of the 20th century. This theory comes hand in hand with the *Theory of Communication* and the problems one faces in attempting to communicate reliably over noisy communication channels. The study of such problems has evolved into the field of **Error-Correcting Codes**. In the most basic scenario one transmits n-bit long messages. However, in order to deal with noise that may arise during transmission, some efficiency must be given up. From among all 2^n possible n-bit strings we consider only a subset $C \subseteq \{0,1\}^n$ ("a code book") and restrict all our messages to n-bit strings from C. In this theory a subset $C \subseteq \{0,1\}^n$ is called a **code** (or a *binary* code when we want to stress that our alphabet is the set $\{0,1\}$.) Consider a message $y \in \{0,1\}^n$ that reaches the receiver. The simplest situation is when $y \in C$, in which case we assume that the message y is indeed what the transmitter had sent. If, however, $y \notin C$, we assume that y is a noisy version of some $x \in C$. We pick that $x \in C$ that is the most likely original message, namely, the one that differs from y in the smallest possible number of bits, i.e., that $x \in C$ which minimizes $d_H(x,y)$. Here $d_H(u,v)$ is the *Hamming distance* between the two strings $u, v \in \{0,1\}^n$, namely, the number of coordinates in which they differ $d_H(u,v) = |\{i|u_i \neq v_i\}|$. It follows that if every two distinct members $u, v \in C$ are at distance $d_H(u,v) \geq r$, then our conclusion is correct whenever no more than $\lfloor \frac{r}{2} \rfloor$ errors occur. In this case we say that the code $C \subseteq \{0,1\}^n$ is $\lfloor \frac{r}{2} \rfloor$*-error-correcting*.

Thus we understand that a major issue in the theory is to find codes which meet two inherently conflicting goals

- The cardinality $|C|$ is large - and so we make good use of the communication channel. This objective is usually quantified in terms of C's *rate*, namely $R(C) = \frac{1}{n} \log_2 |C|$.
- All pairwise distances $d_H(x,y)$ for $x \neq y \in C$ are large - and so the code can correct many errors. The main parameter is C*'s distance* which is $d(C) = \min d_H(x,y)$ over all $x \neq y \in C$. Often we refer to a normalized version $\delta(C) = \frac{d(C)}{n}$.

Understanding this tradeoffs is a mystery that has baffled mathematicians for many years. This fundamental problem is one of the main focal points of this school. A main specific challenge is to understand how large $R(C)$ can be if $C \subseteq \{0,1\}^n$ satisfies $\delta(C) \geq x$ for some $1 \geq x \geq 0$ and n is large. This maximum is usually denoted by $R(x)$,

2 New developments

We intend to mention some fascinating new developments in the mathematical aspects of coding theory. Here are some of the possible topics.

- A harmonic-analytic approach to the problem of the asymptotic problem and the study of the function $R(x)$.
- The theory of association schemes has been a major ingredient in the study of codes. We will introduce some of the basics of this theory.
- We will survey some connections between the construction of good codes and problems pertaining to expander graphs.
- It should be clear that the very same problems can be posed for underlying spaces other than $\{0,1\}^n$. For example, when considered for a Euclidean n-space, these problems become questions on the existence of good sphere packings. We will survey some of these developments and some of the recent breakthroughs in this area.
- On the other hand, analogous questions where the underlying space is the infinite k-regular trees is tightly related to the problem of constructing graphs of high girth. We will point out this connection and the open problems to which it leads.
- Fascinating connections have emerged in recent years between theoretical computer science and the theory of error-correcting codes, e.g., in the context of the PCP theory and derandomization. These lead to new and widely open problems, some of which we will survey. Among those are list decoding and locally decodable codes.