The Linus sequence

Amites Sarkar

The Linus sequence was first described in 1968. Our motivation for studying it came from ergodic theory, although no knowledge of ergodic theory is required in order to understand my talk. Indeed, all the proofs I'll present are purely combinatorial in nature.

The definition of the Linus sequence L_n is that it is a 0-1 sequence which starts with $L_1 = 0$, and for $n > 1$, L_n is chosen so as to avoid a long repeated word. More precisely, define the *terminal repeat length* of a sequence $L_1 L_2 \dots L_n$ as the largest $r \geq 0$ such that the last r digits $L_{n-r+1} \dots L_n$ are the same as the immediately preceding r digits $L_{n-2r+1} \dots L_{n-r}$. We define L_n for $n > 1$ so as to minimize the terminal repeat length of $L_1 \dots L_n$. The first few terms of the Linus sequence are as follows.

$$L = 0\,1\,0\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\,0\,1\,0\,0\,0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,0\,1\,0\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\,0\,1\,0\,0\,0\cdots$$

For example, $L_9 = 0$ since a 1 would cause a terminal repeat length of 3 (repeated block 011), while a 0 would cause a terminal repeat length of only 2 (repeated block 10).

This sequence is fantastically tantalizing because there are many symmetries in it which elude proof. We have many conjectures that are not only backed by numerical evidence but are quite understandable intuitively, yet elude proof. For example it is clear that the frequency of a word, the frequency of the reverse word and the frequency of the word obtained by interchanging 0s and 1s are all the same. We can't prove that. We can't even prove that the frequency of 1s is $\frac{1}{2}$, or that the frequency of any single word even exists at all.

In my talk I'll describe some techniques we (eventually!) developed to analyze this sequence, including the proof of a related combinatorial result (on *justified sequences*) which is of interest in its own right. This is all joint work with Paul Balister and Steve Kalikow.