1114-05-330 **Glenn Tesler*** (gptesler@math.ucsd.edu), Department of Mathematics, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0112. *Computing probabilities of repeated words in a string using the string correlation lattice.* Preliminary report.

The distribution of the number of occurrences of a word w over random strings depends not just on the length of w, but also on how occurrences of w may overlap. This is characterized by the *string autocorrelation* structure of w (Guibas and Odlyzko, 1981). We introduce the *string correlation lattice*: a partially ordered set with a lattice structure, whose elements characterize overlaps among collections of words. We use it to study:

(1) What is the probability that a random string of length n over an alphabet of size q has some k-mer (length k word) occurring at least m times?

(2) We consider the same problem with occurrences of each k-mer counted in both forwards and reverse directions.

(3) In DNA sequences over the alphabet $\{A, C, G, T\}$, we may count occurrences of each k-mer in both the forwards and *reverse complement* directions: reverse the letters and substitute $A \leftrightarrow T$, $C \leftrightarrow G$; e.g., $AACTC \rightarrow GAGTT$. We also generalize reverse complements to other alphabets.

(4) We also consider these problems for random strings generated by a biased q-sided die (e.g., modeling GC content in a DNA sequence). The results are in terms of symmetric functions of probabilities of symbols in the alphabet. (Received September 01, 2015)