

# Transformations on Double Occurrence Words Motivated by DNA Rearrangement

Daniel Cruz<sup>1,\*</sup>, Margherita Maria Ferrari<sup>1</sup>, Nataša Jonoska<sup>1</sup>,  
Lukas Nabergall<sup>2</sup>, Masahico Saito<sup>1</sup>

<sup>1</sup>*Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620*

<sup>2</sup>*Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON N2L 3G1*

`dcruz@mail.usf.edu`

A double occurrence word (DOW) is a word in which every symbol appears exactly twice; two DOWs are equivalent if one is a symbol-to-symbol image of the other. By modeling DNA rearrangement processes using DOWs, it was discovered that over 95% of the scrambled genome of certain ciliates could be described using the so called repeat pattern ( $\alpha\alpha$ ) and the return pattern ( $\alpha\alpha^R$ ), with gaps allowed between the  $\alpha$ 's. These patterns generalize square and palindromic factors of DOWs, respectively. We introduce a notion of inserting repeat/return words into DOWs and study how two distinct insertions in the same word can produce equivalent DOWs. Given a DOW  $w$ , we characterize the structure of  $w$  which allows two distinct insertions to yield equivalent DOWs. In this case, we show that  $w$  belongs to one of three classes, and the class to which  $w$  belongs is only dependent on the locations of the insertions and on the length of the inserted repeat/return words. Further, we show that if one inserted word is a repeat word and the other is a return word, then both words must be trivial (i.e., have only one symbol).