

Seminar Announcement

Idoia Ochoa
Ph.D. Candidate
Electrical Engineering
Stanford University



Thursday, March 3, 2016
Green Hall, Room 0120
10:10 AM

Genomic Data Compression, Processing, and Analysis

Abstract: Recent technological advances have led to a drastic reduction in the cost of genome sequencing, resulting in the accumulation of an unprecedented amount of highly distributed and heterogenous genomic data. Our research is dedicated to identifying and addressing the challenges arising in the context of such data. This undertaking - which combines tools from information and coding theory, statistics, and machine learning - includes the design and deployment of new algorithms for coping with the distribution and storage of the data, for facilitating its access and queries, and for improving the analysis and inference performed on it. The talk will focus on some of my work in this area, geared toward alleviating the requirements for storage of the data, and improving its analysis. I will present some of our lossless compressors of whole genomes and raw sequencing reads, as well as lossy compressors and denoisers of the quality scores that come with the sequencing data. In some cases lossy compression results not only in substantial storage savings, but also boosted performance when inference is done on the reconstructed data. Conversely, denoising the quality scores can result in substantial compression gains. Empirical manifestations of these phenomena will be exhibited, and some of the theory that explains them will be discussed.

Bio: Idoia Ochoa is currently a Ph.D. candidate in the Electrical Engineering department at Stanford University, advised by Prof. Tsachy Weissman. She received her M.Sc. from the same department in 2012. Prior to Stanford, she graduated with B.Sc. and M.Sc. degrees in Telecommunication Engineering (Electrical Engineering) from the University of Navarra, Spain, in 2009. She has conducted internships at Google and Genapsys, and served as technical consultant for the HBO show "Silicon Valley". Her research interests include compression and coding, bioinformatics, information theory, machine learning, communications and signal processing. She is recipient of the Stanford Graduate Fellowship, La Caixa Graduate Fellowship, and an award for excellence from the Basque Government.