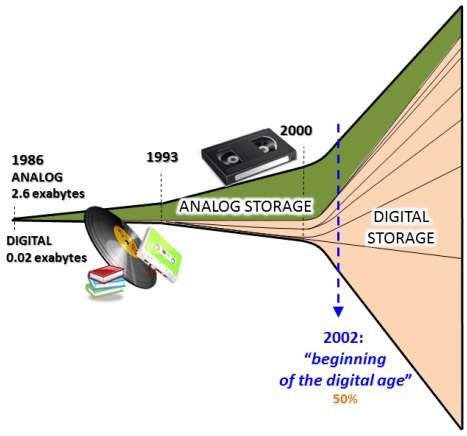
COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Algorithms for Data Science

Instructor: *Michał Dereziński* **EECS 498-005 Special Topics, Winter 2024**

Overview:
This course will introduce algorithmic and theoretical foundations of data science. With the emergence of machine learning and data science, as well as the ever-increasing data sizes, providing theoretical foundations for the computational complexity of data manipulation algorithms is becoming increasingly important. The course will cover several important algorithms in data science and demonstrate how their performances can be analyzed. While fundamental ideas covered in EECS 376 (e.g., design and analysis of algorithms) will be important, some topics will introduce new concepts and ideas, including randomized dimensionality reduction, sketching algorithms, and optimization algorithms (e.g., for training machine learning models).



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

The course will delve into concepts in linear algebra, such as matrix multiplication and singular value decomposition, as well as in probability theory, such as expectation, independence and concentration of random variables.

Prerequisites: EECS 376 (advisory), solid background in linear algebra and probability

A tentative list of topics:
Randomized methods for big data (dimensionality reduction, sketching), computational linear algebra (fast matrix multiplication, singular value decomposition, linear regression, principal component analysis), convex optimization (gradient descent, Newton's method, stochastic gradient), and spectral methods (spectral graph theory and spectral clustering).