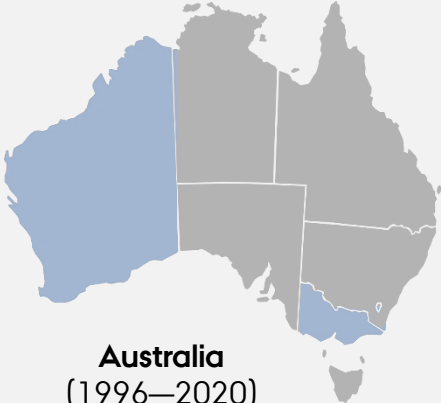# Optimising data analytics by peeking inside compressed data

ECE R-Day 2023, October 13

Aaron Hurst
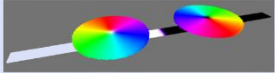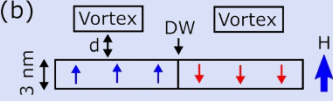
An Unexpected Journey

Australia
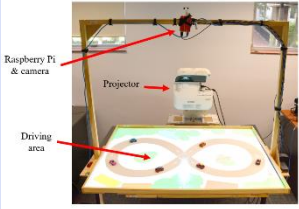(1996—2020)

BSc & M.Eng
(2013—2017)

STUDENT RESEARCH

(a)

(b)

Vortex   DW   Vortex

d

3 nm   H

Research project
(2015)

Raspberry Pi
& camera

Projector

Driving
area

Fig. 2.   Testbed as constructed.

Masters Thesis
(2017)

IBM

Data Scientist
(2018—2020)

Denmark
(2020—)

AU

UC San Diego

PhD
(2020—2023)

**Research
visit to USA**
(2023)

Postdoc
(2023—?)

Data Compression. Old News? New Needs.

Claude Shannon
(1916—2001)

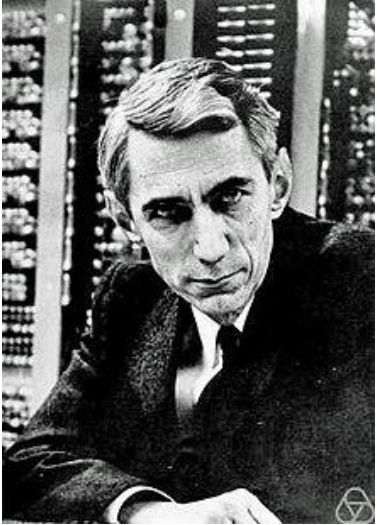**The value of data is in its use.** That is, frequent queries and analytics

Thus, don't just focus on efficient data transmission & storeage, but also **make data easy to use.**

Goal: Holistically optimise for data storage (compression) & utility (analytics)

Research during my PhD

**Goal:** Holistically optimise for data storage (compression) & utility (analytics)

1 Can we **do analytics** directly on compressed data?

2 Can we **optimise compression** for data analytics?

3 Can we use this to optimise **IoT / Edge Computing** systems?

4 Can we use this to optimise **approximate query processing**?

5 Can we apply this as lossy compression for **federated learning**?

Student Research Opportunities

Claude Shannon
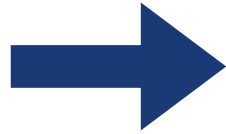
You?

Scope

- ✓ R&D Project
- ✓ Bachelor project
- ✓ Masters Thesis

Topics/types of projects

- ➢ Optimising compression algorithms (e.g. for column-wise databases)
- ➢ Investigating how to apply algorithms to new problems/domains (e.g. logs, streaming data)
- ➢ Low-latency analytics tasks on compressed data (e.g. table joins, machine learning)
- ➢ Etc.

... Come chat later for more info!

Network Computing, Communications and Storage

Daniel E. Lucani Rötter
Professor

Qi Zhang
Assoc. Professor

Hei Victor Cheng
Assistant Professor

Marcell Féher
Postdoc

Gajraj Kuldeep
Postdoc

Nan Li
PhD

Mads Antonsen
PhD

Francesco Taurone
PhD

Xiao Cia
PhD

Valdemar Tang
PhD

Guoyou Sun
PhD

Aaron Hurst
PhD

AARHUS
UNIVERSITY