

# Baler: A tool for machine learning based data compression

---

Alexander Ekman for the Baler collaboration

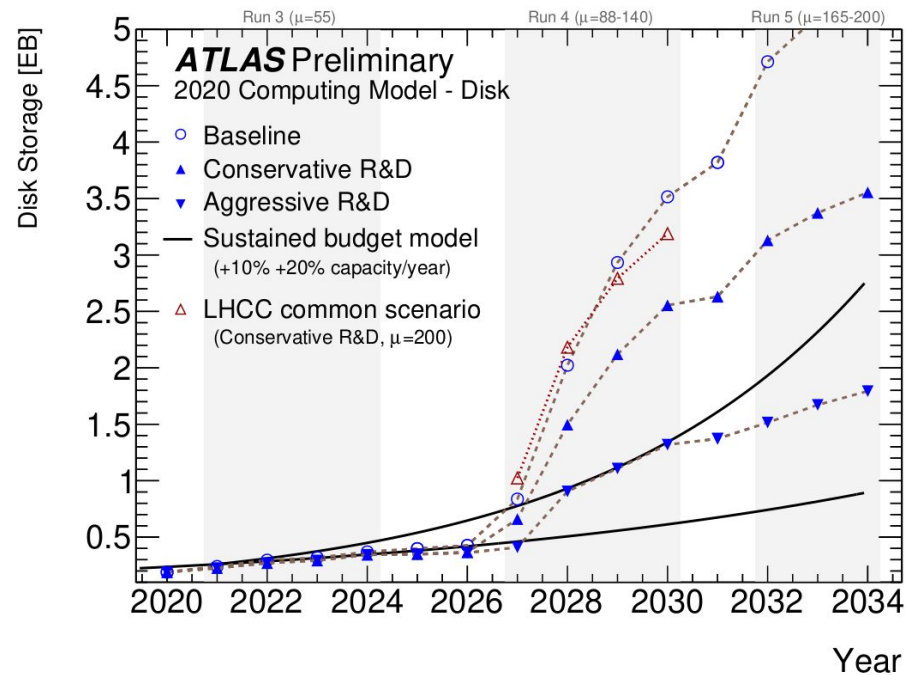
# Problem: More data than storage



KUNGL. FYSIOGRAFISKA SÄLLSKAPET I LUND  
AKADEMI FÖR NATURVETENSKAP, MEDICIN OCH TEKNIK



- Collection and generation of data is overwhelming processing and storage capacity in science and industry
- High demand for greater compression than traditional lossless and lossy methods



ATLAS HL-LHC Computing Conceptual Design Report  
Calafiura, P ; Catmore, J ; Costanzo, D ; Di Girolamo, A  
<http://cds.cern.ch/record/2729668/>

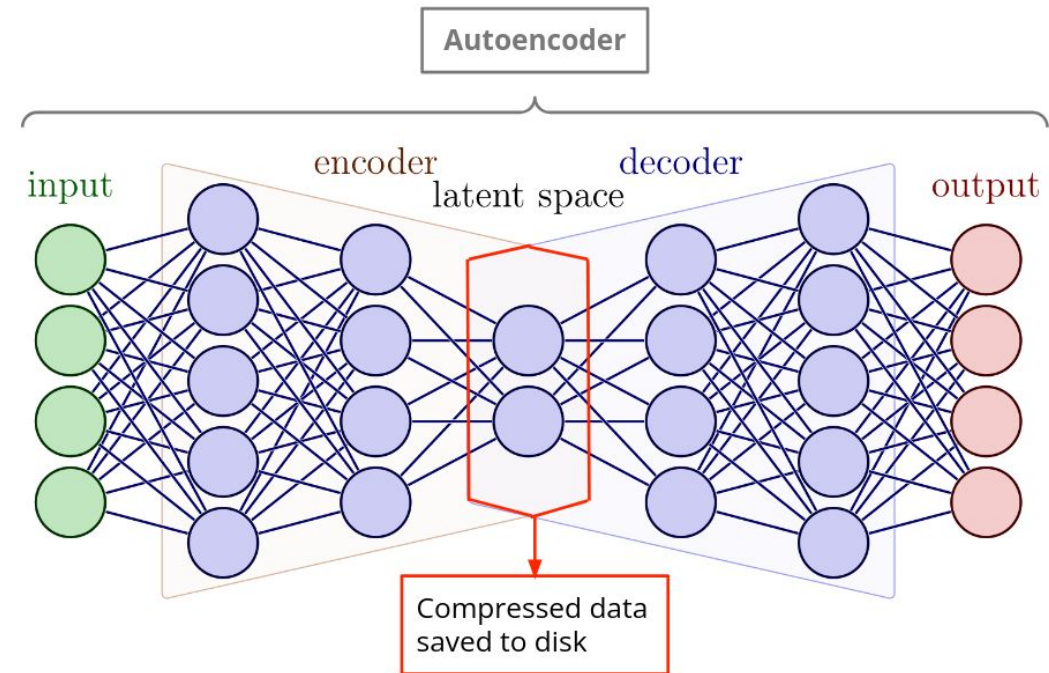
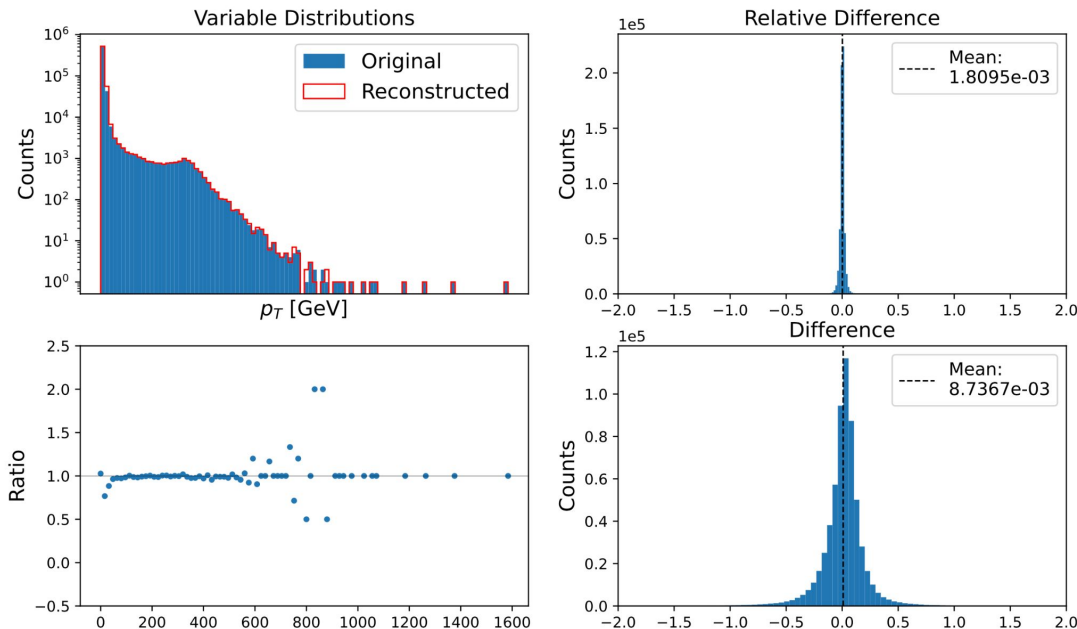


Figure modified from:  
[https://tikz.net/neural\\_networks/](https://tikz.net/neural_networks/)

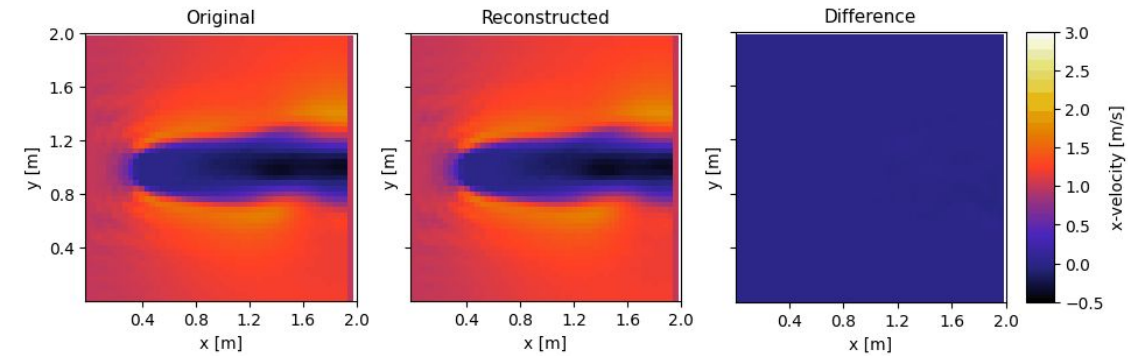
# Our solution: “Baler”

- Multidisciplinary tool to investigate the viability of this compression method
  - <https://github.com/baler-collaboration/baler>
- Simple to install as a pip package
  - `pip install baler-compressor`
- Promising performance for varying scientific fields

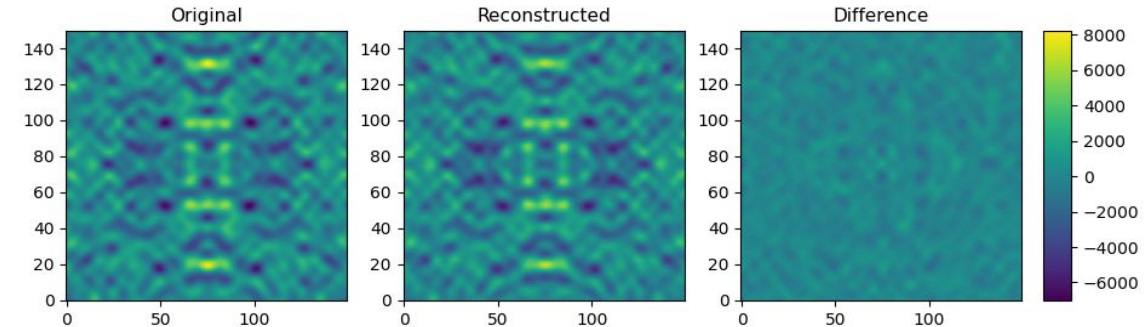
## High Energy Particle Physics



## Computational Fluid dynamics



## X-spectroscopy



# Future outlook

- We found a small demand for lossy compression of scientific data in final storage
- New focus on “online” compression and bandwidth compression using FPGA technology
- Draw inspiration from progress in machine learning based image and video compression

