Data Reduction using Singular Value Decomposition (SVD) Algorithm

Jing Zhang

jing.zhang@hlrs.de

High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

Overview

- Motivation and aim
- SVD algorithm
- Results by applying SVD
  - 2D steady-state test case (1)
  - 3D transient test cases (2 - 5)
- Summary and future work

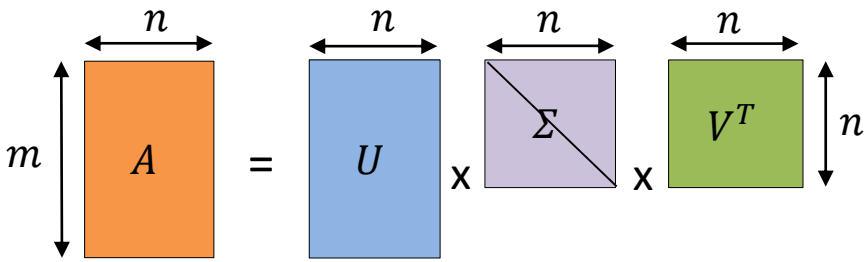
Motivation and aim

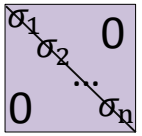
- Steady increase in computing power
  - to use more complex models
  - to simulate a lot of fluid-flow problems
  - to perform simulations with large amount of CFD data
- I/O is bottleneck
  - subsystem slow compared to other parts of computing system

- **Aim** of this study
  - to reduce amount of CFD data transferred from memory to disk
  - by using data-reduction algorithm SVD
  - to minimize impact of I/O bottleneck on computing performance

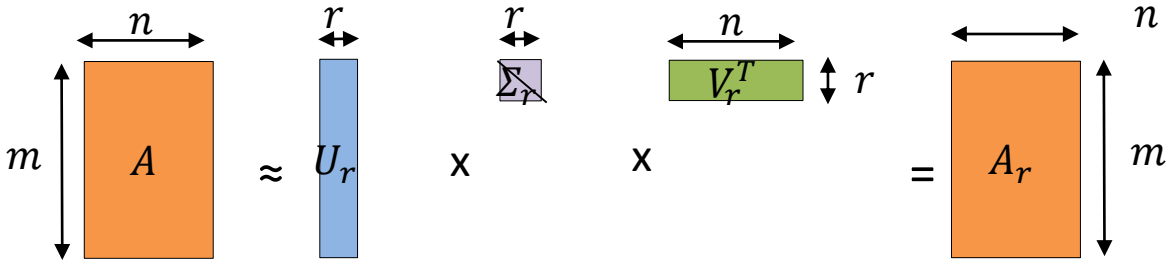
SVD algorithm

- $A = U \times \Sigma \times V^T$ (if $m > n$)



(Singular values: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ )

- $A = \sum_{i=1}^{n} u_i \sigma_i v_i^T \approx \sum_{i=1}^{r} u_i \sigma_i v_i^T = A_r \ (if\ \sigma_r \gg \sigma_{r+1})$



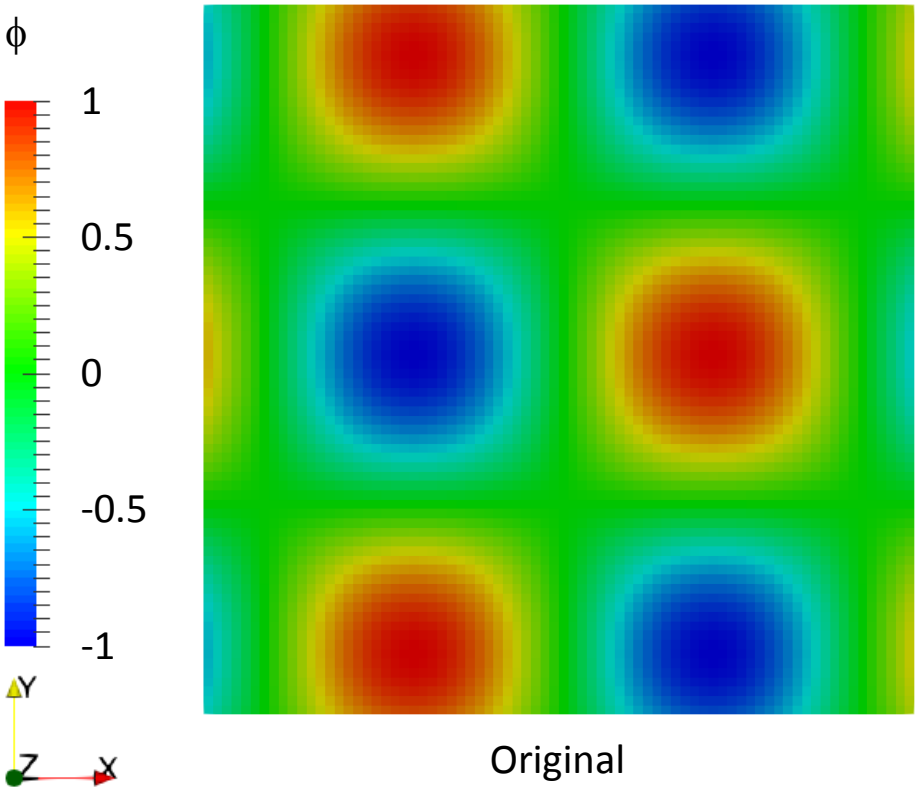
- SVD subroutine PDGESVD in ScaLAPACK library

Parameters for data-reduction algorithm SVD

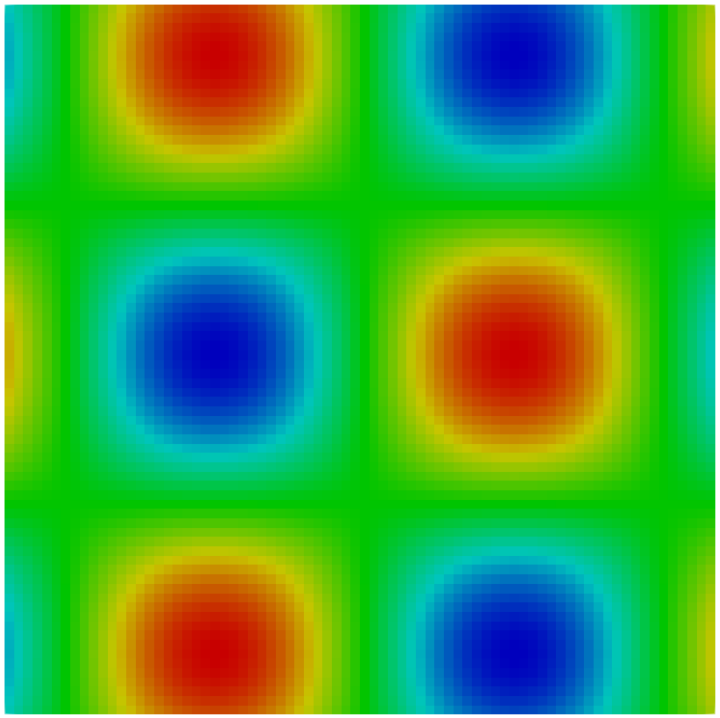
- Compression ratio: $CR = \frac{m \times n}{m \times r + r + r \times n} = \frac{m \times n}{r \times (m + 1 + n)}$
- Requirement of compression: $\mathrm{r} < \frac{m \times n}{m + 1 + n}$
- Mean Squared Error: $MSE = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} |A(i,j) - A_r(i,j)|^2}{m \times n}$
- Peak Signal to Noise Ratio: $PSNR = 10 log_{10} \frac{max[A(i,j)]^2}{MSE}$
  - Good quality for SVD algorithm: $PSNR \geq 35$ dB

2D steady-state test case 1 ($A_{76\times76}$)

- Original matrix is square and symmetric
- Original and reconstructed matrix are similar



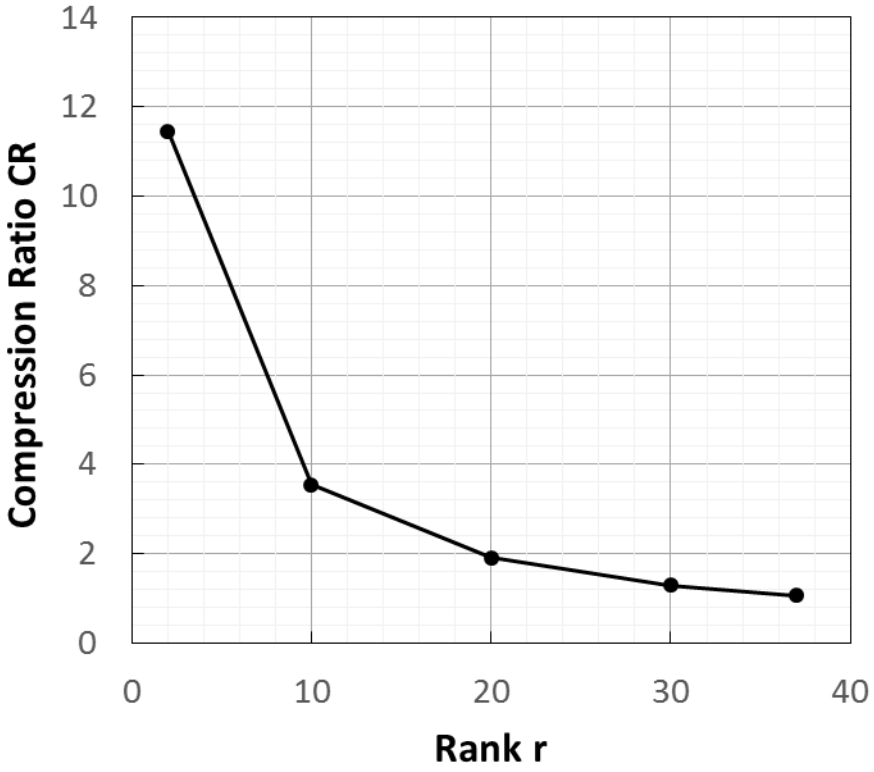
Original

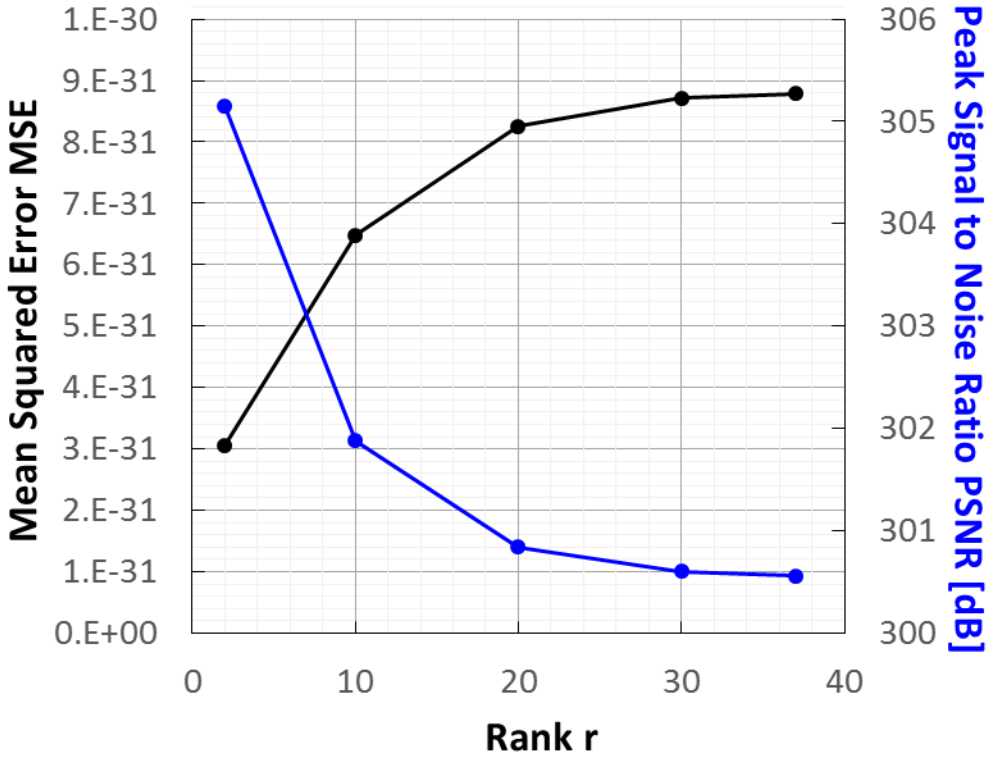


Reconstructed (rank=2)

2D steady-state test case 1 ($A_{76\times76}$)

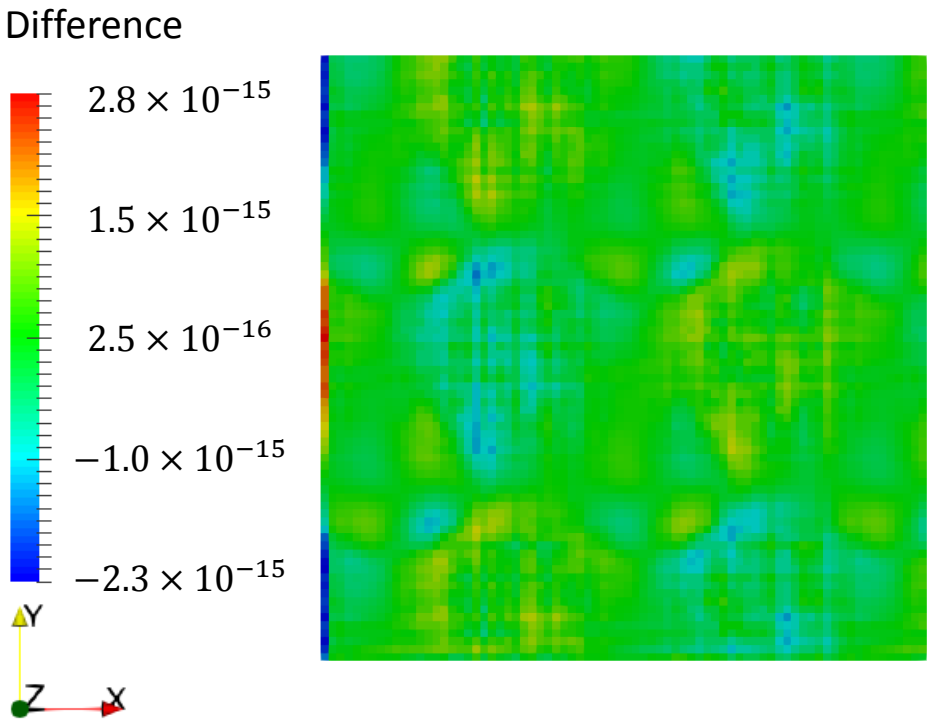
- Limit on rank number 37
- Variation of CR, MSE and PSNR according to rank value up to 37





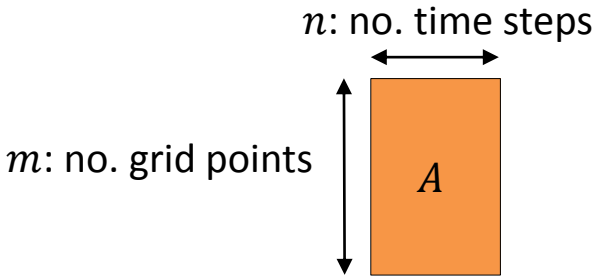
2D steady-state test case 1 ($A_{76\times76}$)

- Maximal absolute value from original matrix: 1.00
- Maximal absolute difference between original and reconstructed matrix (rank=2): 4×10^{-15}



3D transient test cases

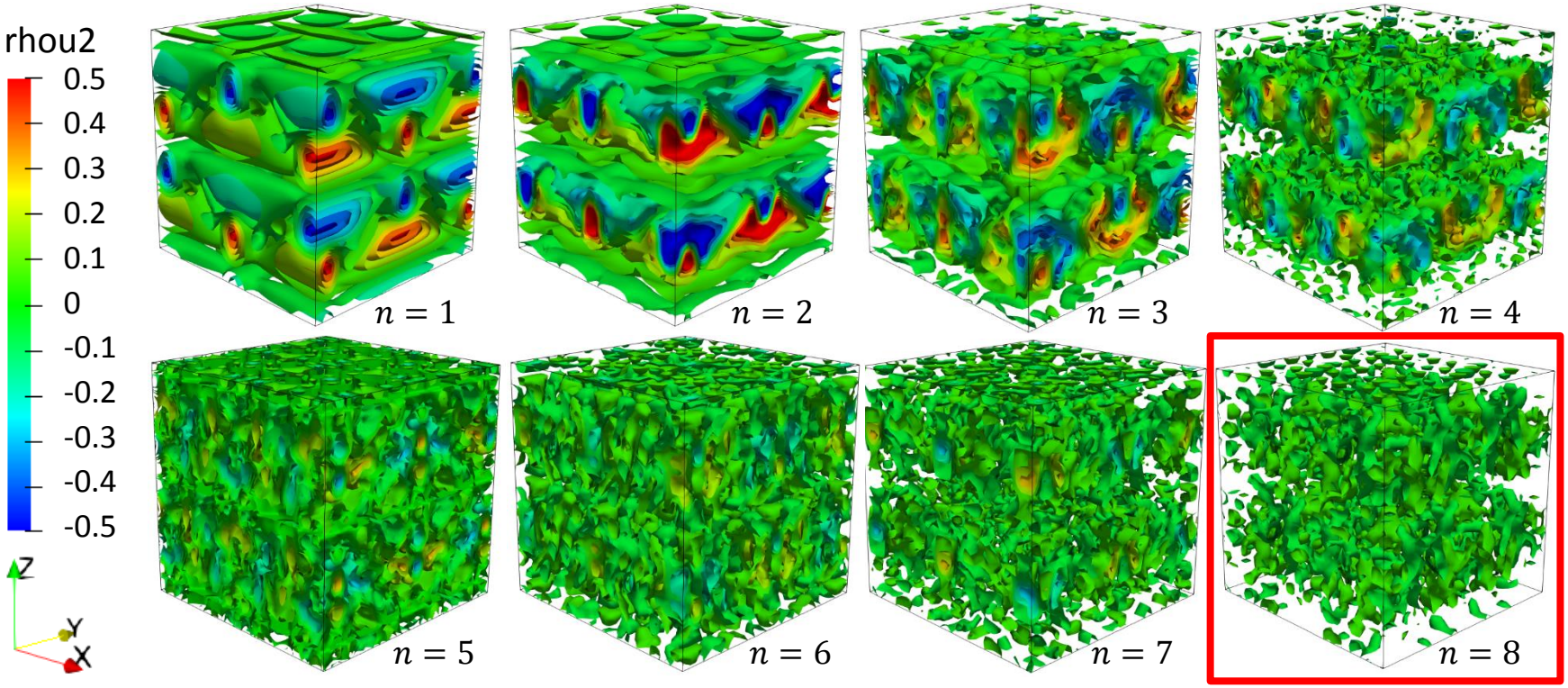
- Only one dataset Rhou2 (i.e. Rho × u2) is analysed
- Transformed data from 3D transient test cases into 2D dataset



- Data-transformation process
  - Convert many HDF5 files at every time step into many single binary files using HDF5 utility h5dump
  - Merge all binary files into one binary file including all data
  - Reconvert the merged binary file into one 2D HDF5 file for applying SVD using HDF5 utility h5import

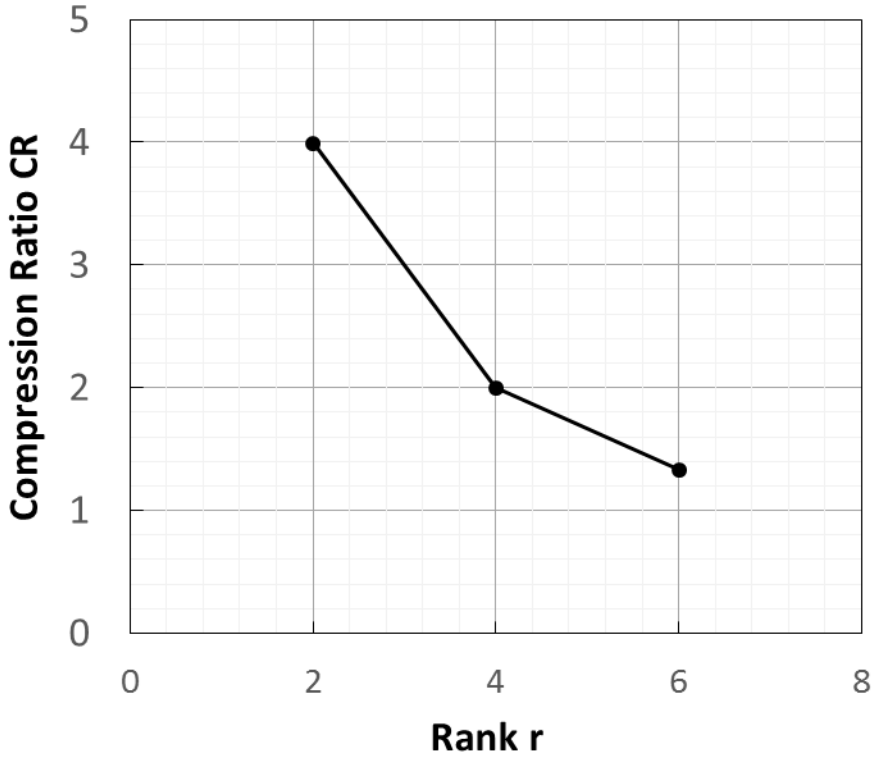
3D transient test case 2 (m=36^3, n=8)

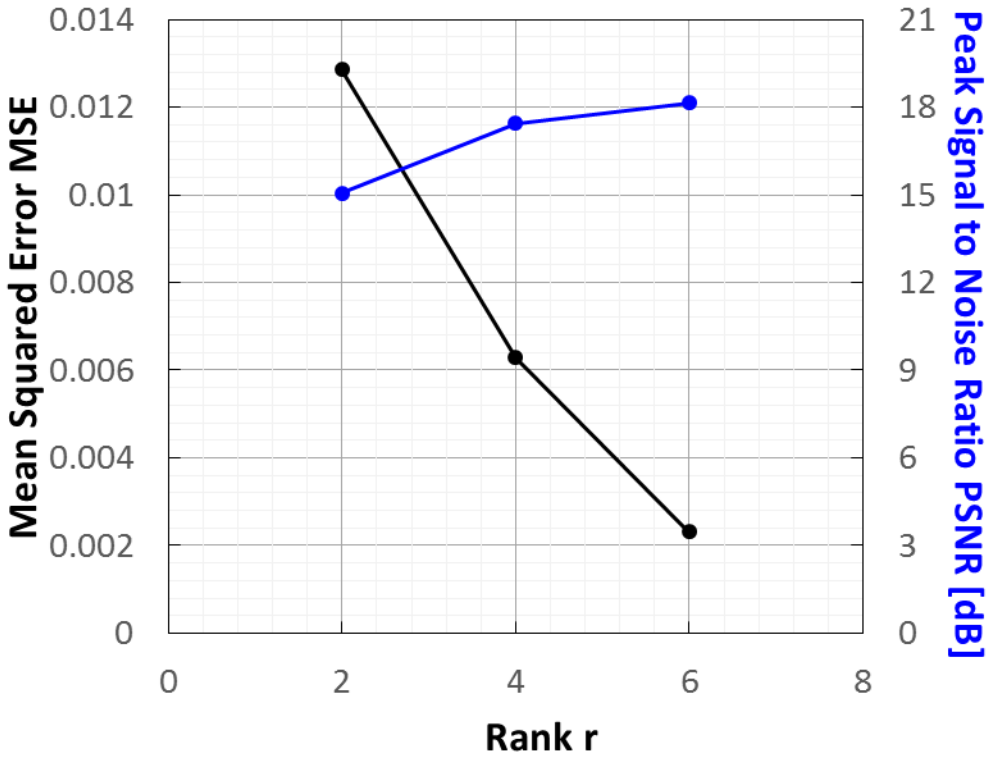
- Number of grid points: $m = 36^3 = 46656$
- Number of time steps: $n = 8$
- Non-dimensional simulation time: t = 0~20



3D transient test case 2 ($m=36^3$, $n=8$)

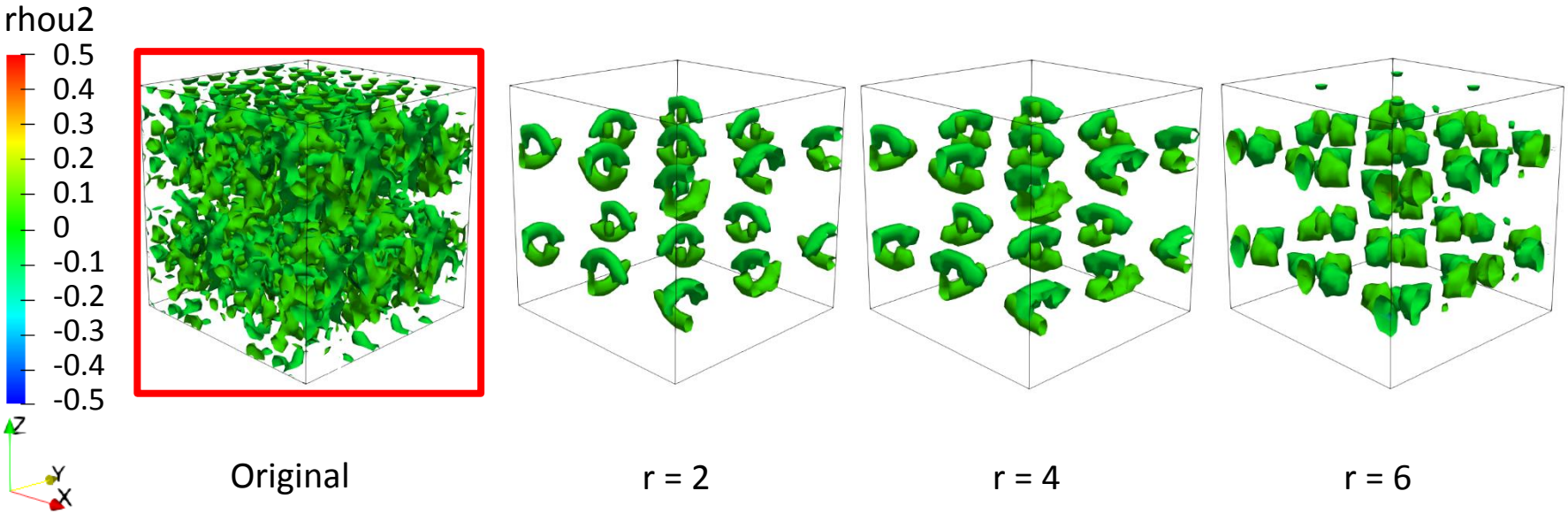
- Limit on rank number 7
- Variation of CR, MSE and PSNR according to rank value (2, 4, 6)





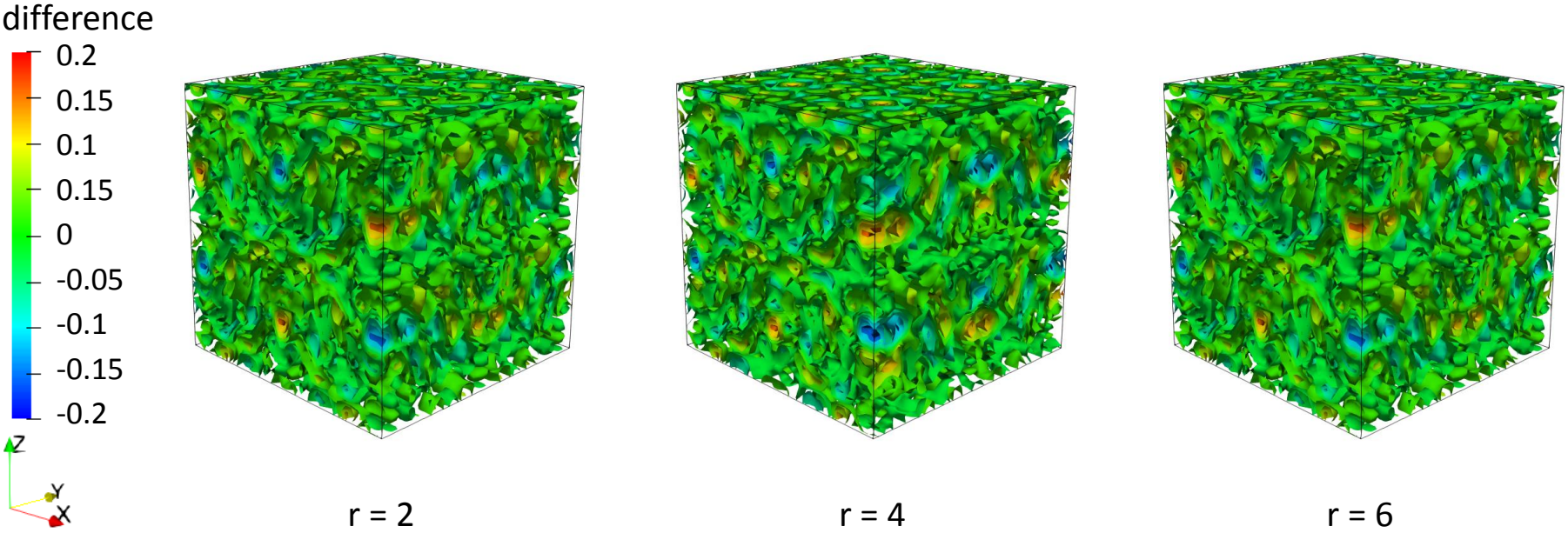
3D transient test case 2 ($m=36^3$, $n=8$)

- Comparison of original and reconstructed data at the last time step
- All reconstructed data show poor quality



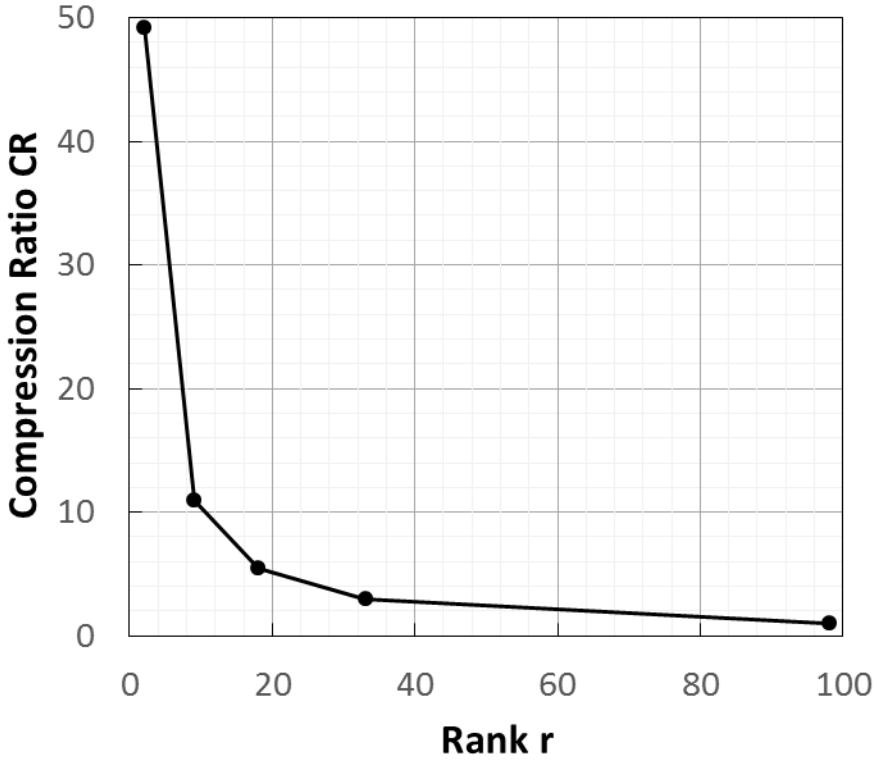
3D transient test case 2 ($m=36^3$, $n=8$)

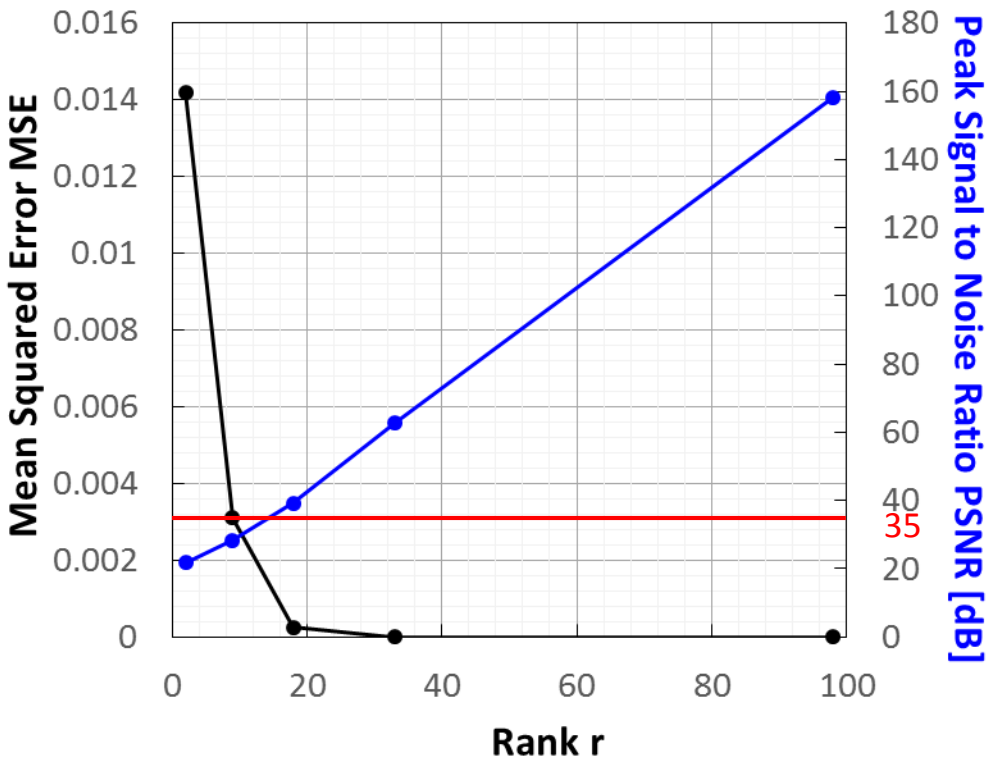
- Maximal absolute value of original data: 1.2
- Difference between original and reconstructed data at the last time step
- Three figures show similar results and difference is relative large



3D transient test case 3 ($m=36^3$, $n=99$)

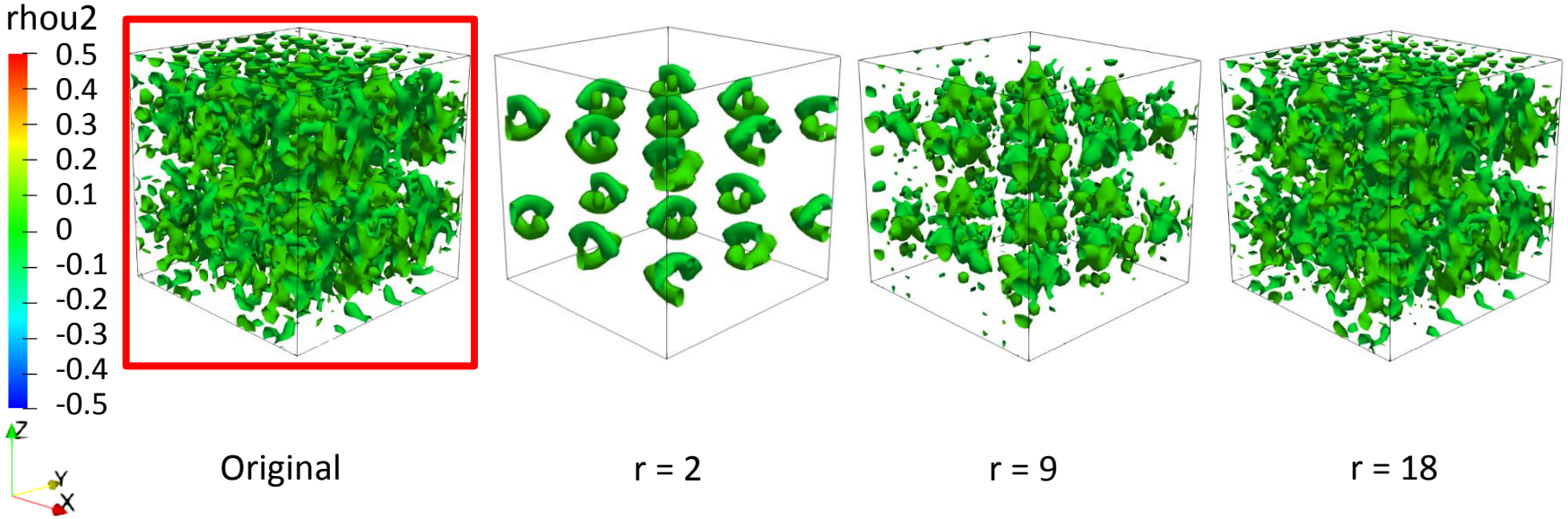
- Limit on rank number 98
- Variation of CR, MSE and PSNR according to rank value up to 98





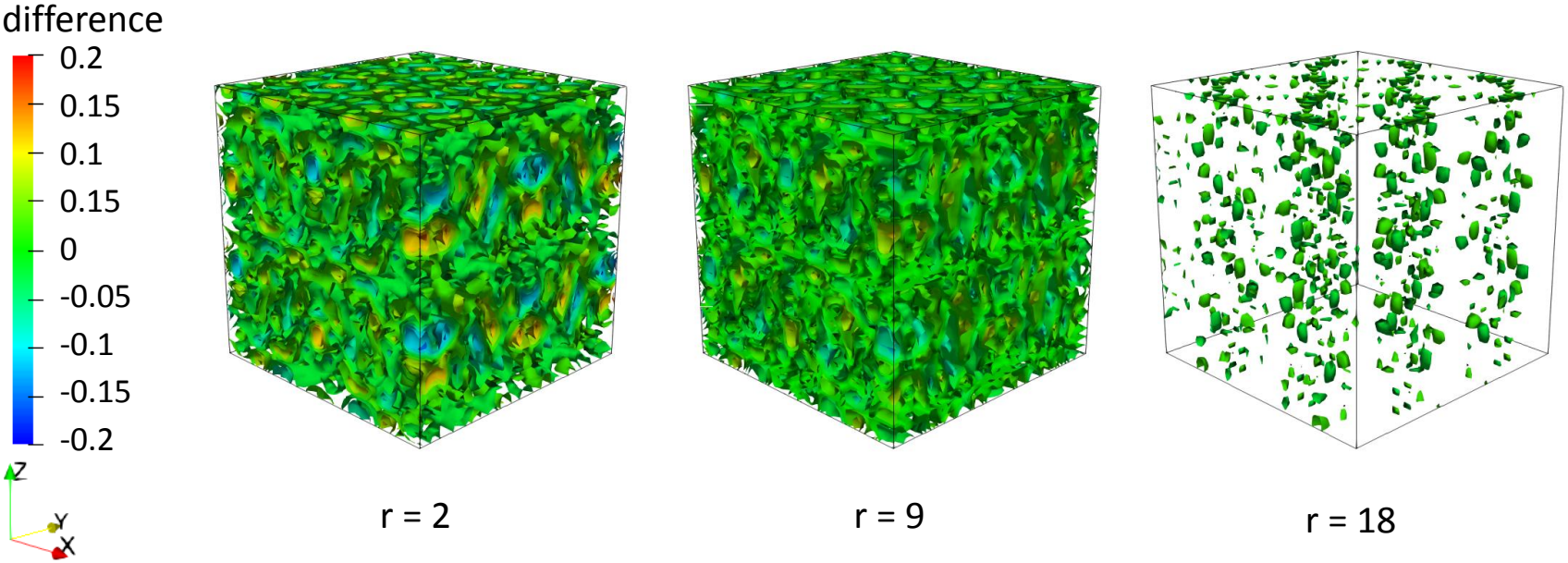
3D transient test case 3 ($m=36^3$, $n=99$)

- Comparison of original and reconstructed data at the last time step
- If the rank number is increased, reconstructed data approach the original data



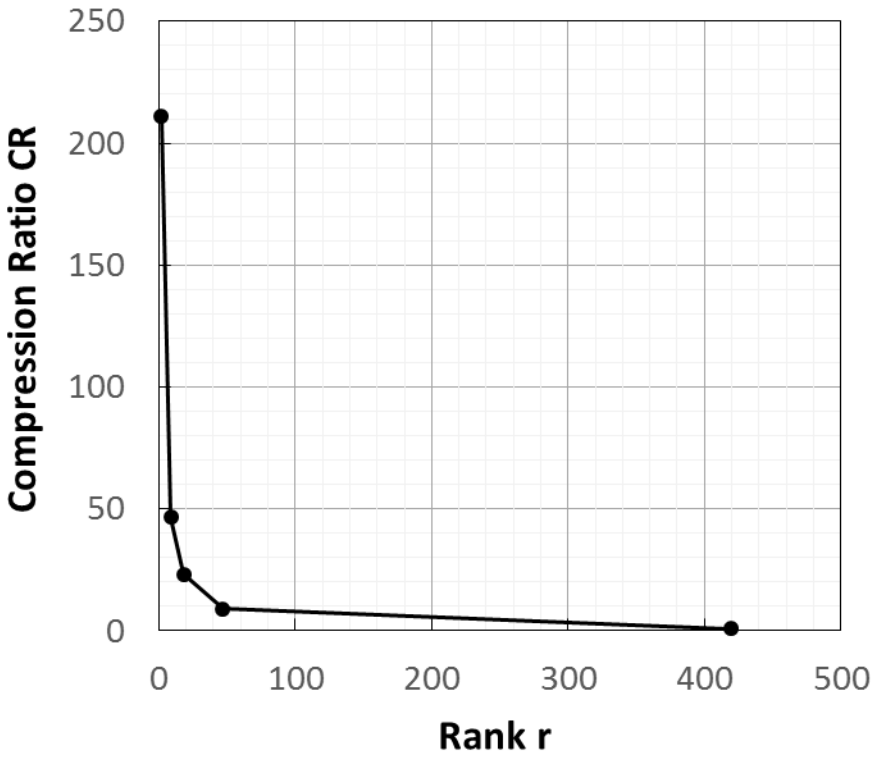
3D transient test case 3 ($m=36^3$, $n=99$)

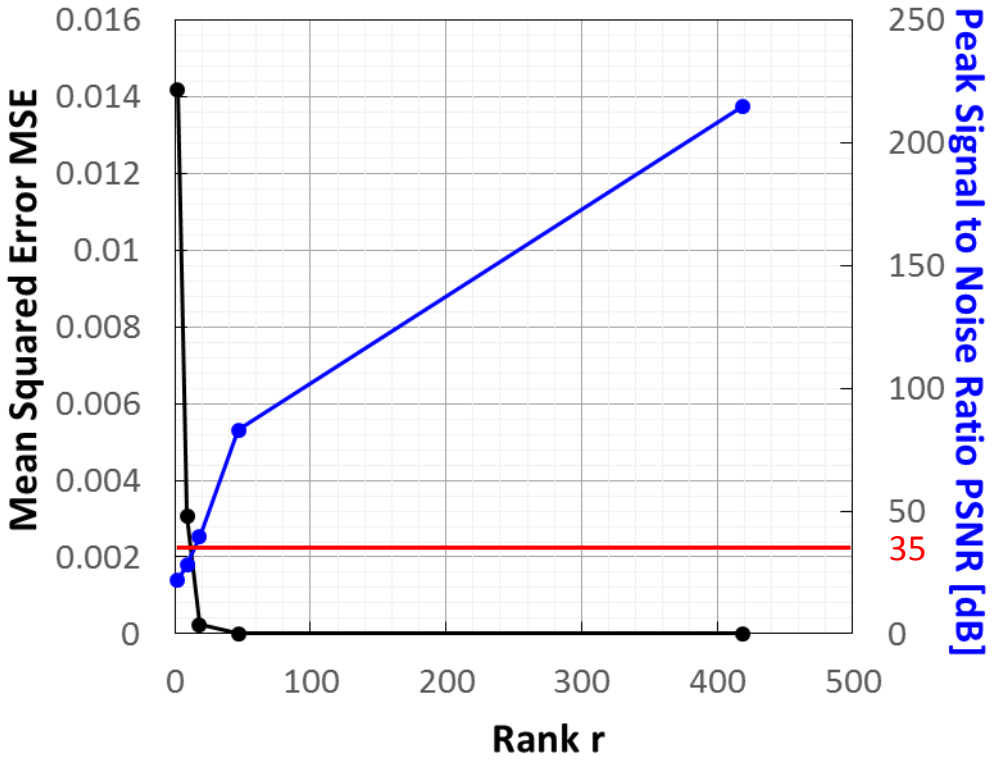
- Maximal absolute value from original data: 1.5
- Difference between original and reconstructed data at the last time step
- If rank number is increased, difference is reduced



3D transient test case 4 ($m=36^3$, $n=423$)

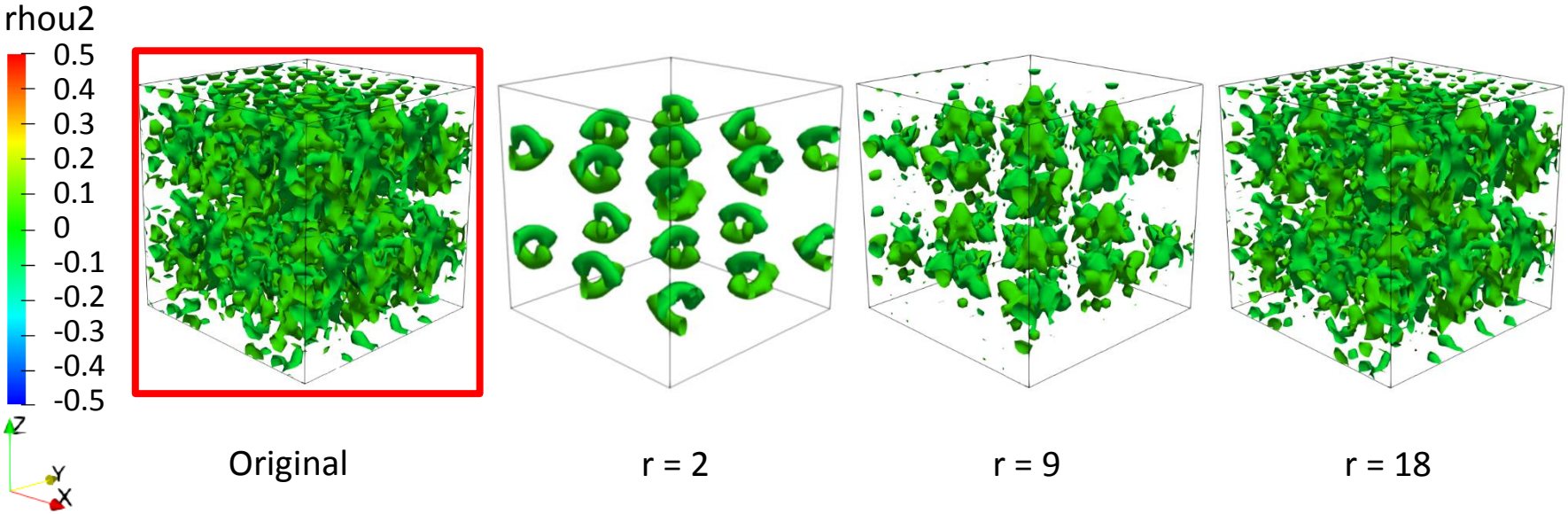
- Limit on rank number 419
- Variation of CR, MSE and PSNR according to rank value up to 419





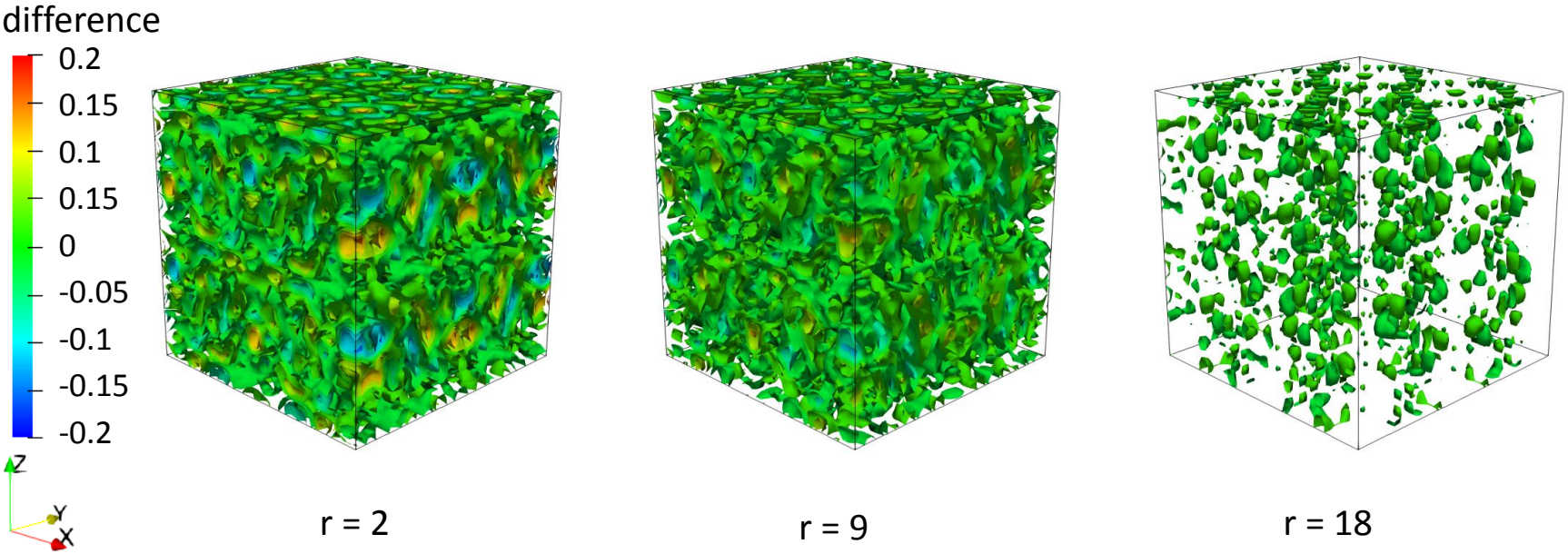
3D transient test case 4 ($m=36^3$, $n=423$)

- Comparison of original and reconstructed data at the last time step
- With rank number of 18 reconstructed data are similar as original data



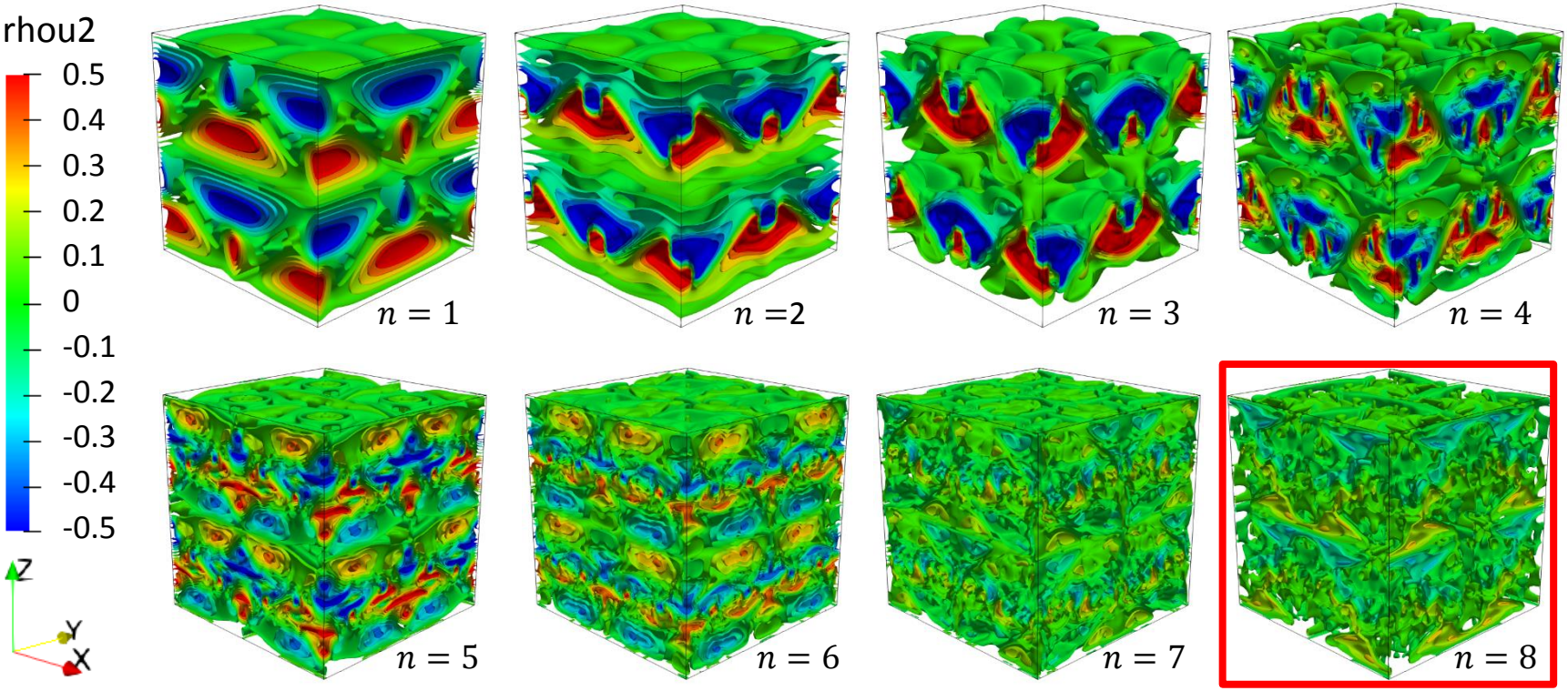
3D transient test case 4 ($m=36^3$, $n=423$)

- Maximal absolute value from original data: 1.5
- Difference between original and reconstructed data at the last time step
- Compression ratio almost 4x larger than for test case 3



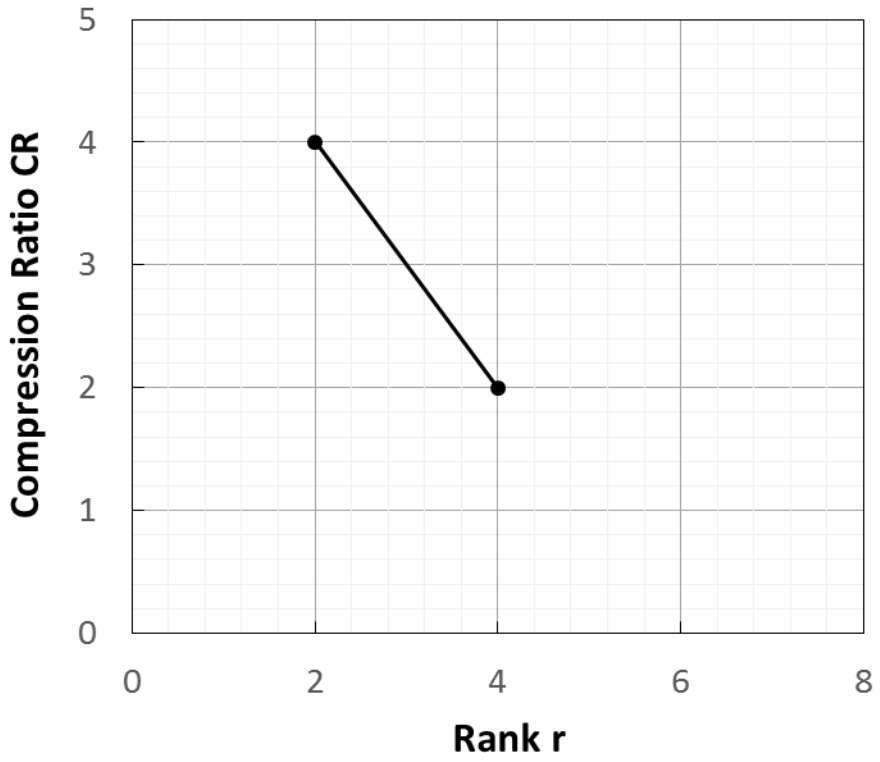
3D transient test case 5 (m=260^3, n=8)

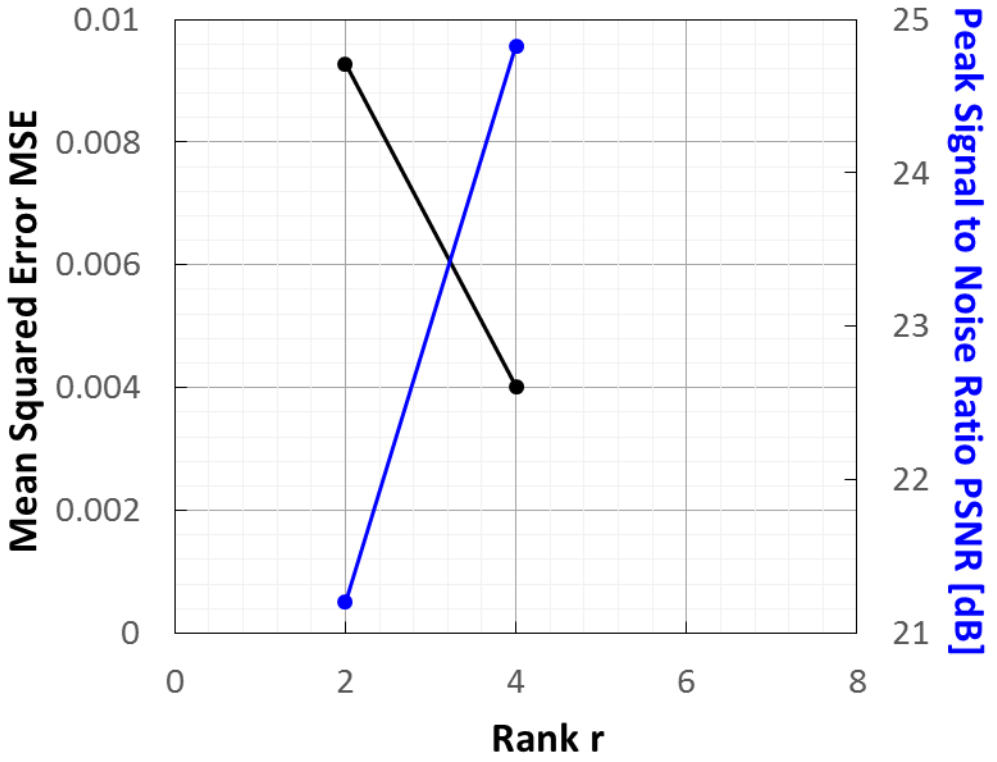
- Number of grid points: $m = 260^3 = 17576000$
- Number of time steps: $n = 8$
- Non-dimensional simulation time: t = 0~20



3D transient test case 5 ($m=260^3$, $n=8$)

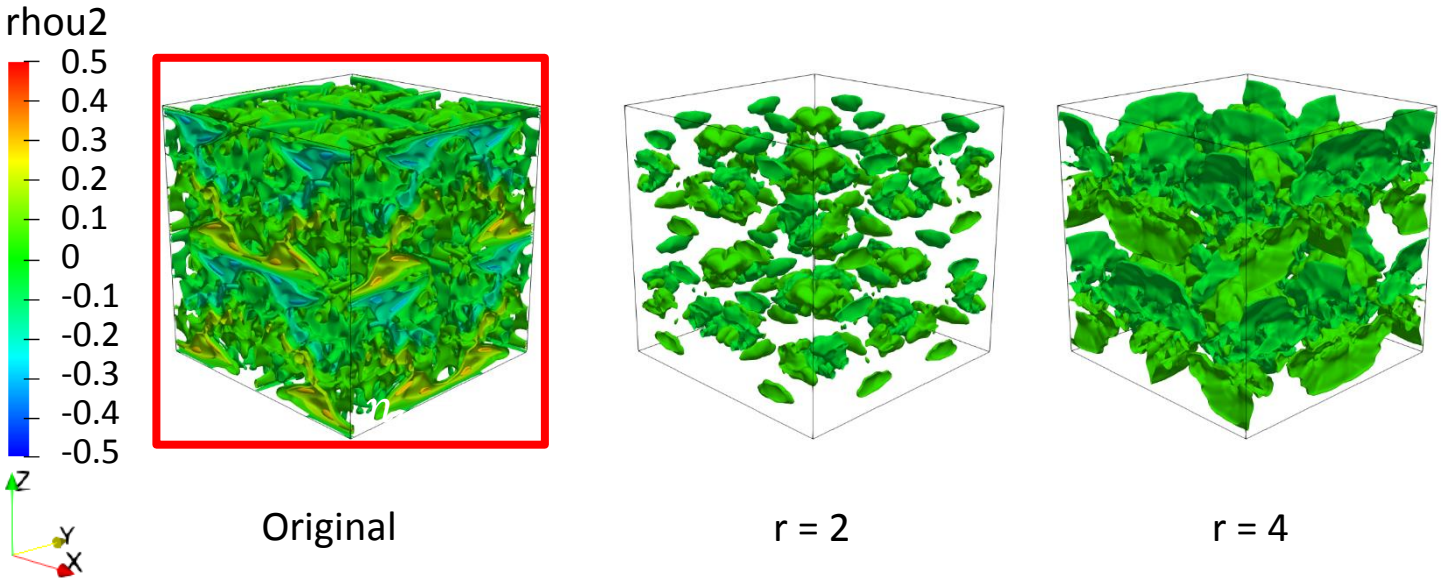
- Limit on rank number 7
- Variation of CR, MSE and PSNR according to rank value (2 and 4)





3D transient test case 5 ($m=260^3$, $n=8$)

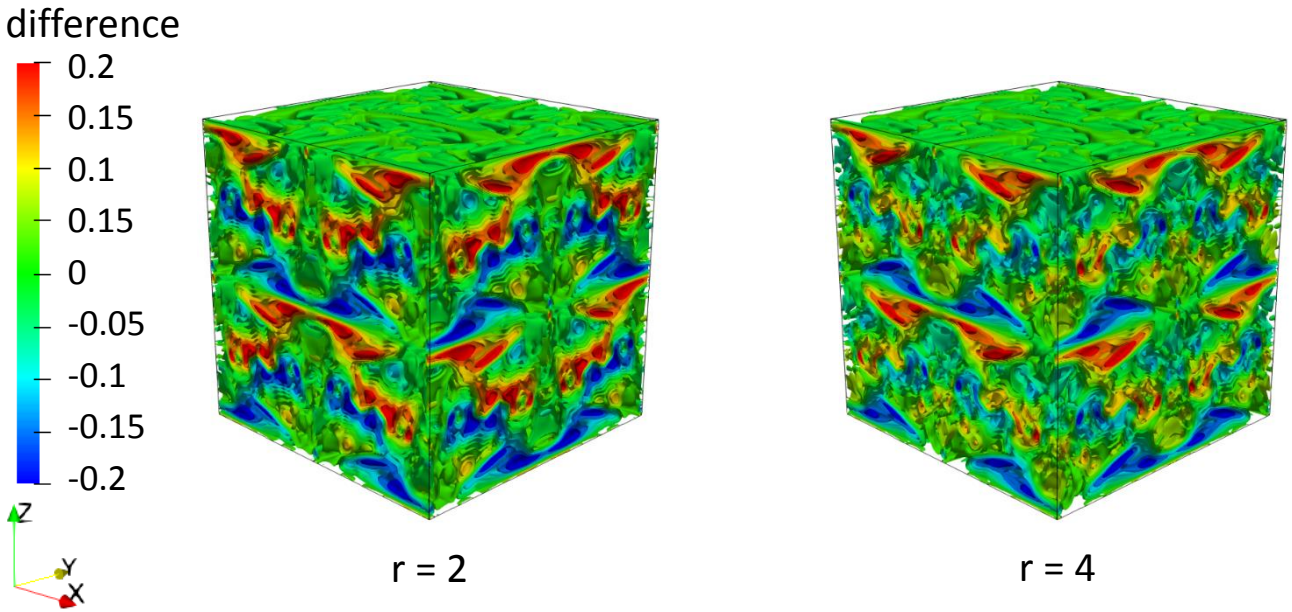
- Comparison of original and reconstructed data at the last time step
- Both reconstructed data show poor quality



Original r = 2 r = 4

3D transient test case 5 ($m=260^3$, $n=8$)

- Maximal absolute value from original data: 1.1
- Difference between original and reconstructed data at the last time step
- Difference is relative large



Summary and future work

- SVD has been implemented and parallel tested with 2D and 3D test cases
- Important parameters have been analysed for data-reduction algorithm SVD, e.g. CR, MSE and PSNR
- Results of applying SVD have been visualized and discussed
  - 2D steady-state test case: square original matrix; with small rank no. CR is large; store less data; good quality for reconstructed data
  - 3D transient test cases: same grid size and rank number; if number of time steps increased; CR increased; store less data; if PSNR value is equal to or just higher than 35 dB; good quality for results
- **Future work**
  - SVD should be further improved with very large datasets

Thanks for your attention!

Jing Zhang

jing.zhang@hlrs.de

High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany